Everything You Need To Know About Analysis

Nathanael Chwojko-Srawley

November 20, 2022

Contents

Preliminary Integration **iv**

1 Measures **1**

1.1 Sigma-Algebras and Properties . . . 3
1.2 Measure . . . 10
1.3 Outer Measure . . . 14
1.4 Borel Measure on the Real Line . . . 22
1.4.1 Lebesgue Measure . . . 27

2 Integration **35**

2.1 Measurable Functions . . . 36
2.1.1 Standard Representation . . . 44
2.2 Integration of non-negative Function . . . 47
2.3 Integration of Real and Complex Functions . . . 55
2.3.1 Relating Lebesgue and Riemann Integral . . . 63
2.4 Modes of Convergence . . . 65
2.5 Product Measure . . . 68
2.6 The n-dimensional Lebesgue Integral . . . 80

3 Signed Measure and Differentiation **83**

3.1 Signed Measures . . . 84
3.2 Lebesgue-Radon-Nikodym Theorem . . . 88
3.3 Complex Measure . . . 97
3.4 Differentiation of Euclidean Space . . . 103
3.4.1 Regular Measures . . . 110
3.5 Functions of Bounded Variation . . . 112

4 Function Analysis **130**

4.1 Normed Vector Spaces . . . 131
4.2 Linear Functionals . . . 136
4.3 Baire Category Theorem . . . 136
4.4 Topological Vector Space . . . 136
4.5 Hilbert Spaces . . . 138
4.5.1 Basic definition and properties . . . 139

5 Lebesgue Spaces **150**
5.1 Basics . . . 150
5.1.1 Relating Lebesuge spaces: Interpolation . . . 165
5.2 The dual of Lp . . . 170
5.3 Some Useful Inequalities . . . 171

6 Fourier Analysis **178**
6.1 Basic Definitions . . . 179
6.2 Convolution . . . 182
6.3 Fourier Transform . . . 189

7 Distribution Theory **197**
7.1 Intuition . . . 197
7.2 Definitions . . . 197
7.3 Compactly Supported, Tempered, and Dense Distributions . . . 202

8 More Measures and Integrals **203**
8.1 Topological Groups and Haar Measure . . . 203

The goal of this document is to encapsualte all the notes and intuition I will gain taking MAT457. The preliminary chapter will be a quick reminder on why we care to generlizatize our notion of integrability.

Assumptions in reading this text

1. basic definition and notions in topology are assumed, covering at least:
   (a) open/closed sets, closure and interior, Hausdorff space, limit definition of closure (given the space if Hausdroff, or at least T_1), continuity
   (b) Compactness, connectedness, Heine-Borel Bolazno-Weiestass in $\mathbb{R}^n$, Path-connectedness
   (c) subspace topology, product topology, Quotient topology
   (d) Metric spaces, many equivalent metric on product spaces,
2. A good understanding of Riemann integration
3. A basic understanding of point-wise convergence and uniform convergence. For point-wise convergence, we'll use $f_n \to f$ notation. For uniform convergence, we'll use $f_n \rightrightarrows f$.
4. Defining and understanding basic properties of sup, inf, lim sup, and lim inf. In particular, remember that if S is a set and $M = \sup S$, for all $\epsilon > 0$, there exists an $s \in S$ such that

$$M < s + \epsilon$$

   And similarly for the infimum, with $s < m + \epsilon$.

Since we are doing analysis, the axiom of choice is assumed. There are some fascinating results we get by assuming the axiom of choice which we will point out.

What is Analysis?

Now that I've taken the course and have done some reserach in PDE's I feel like I've gained a perspective on how to think about analysis. Fundamentally, Analysis is about *estimates* and *approximation*, especially on functions; a lot of the fundamental results in anlysis is centered around these two concepts. Since we are working over $\mathbb{R}$, we have the privelge of having an *ordering*, allowing us make precise the notion of a number being bigger than another (something we can't do over $\mathbb{C}$ or many other fields). The following bullet points elaborates on this:

1. Many times, we will want to give a numerical value to a function or a space of functions. To accomplish this, we will often define a functional (or subfounctional) a space of functions. For example, the integral $\int : C^\infty(\mathbb{R}^n) \to \mathbb{R}_{>0}$ takes in functions and assigns them numerical value giving us some idea of how fast the function grows and the total "energy", "work", or "total value" the function has. Notie that the function $\int$ as we've defiend it above is not well-defined, there are function whose integral is infinite. Conversely, we are missing many integrable function (recall if a function is merely almost everywehre continuous, it is integrable). We will refine the domain in this book.

   The integral is not the only functional. We will soon find there are in fact many functions, letting us control many properties of a function such as:

(a) the total boundedness of the function (i.e. that it nowhere goes to infinity)

(b) the "total work" of the function (i.e. a finite integral)

(c) the speed decay of a function (i.e. that it goes to zero fast enough)

(d) the "frequency" of a function (i.e. that the function doesn't start oscillating too rapidly)

(e) the regularity of a function (i.e. that the derivative or derivatives dont' explode too fast or oscilate to fast)

2. interpolation: (I'm thinking of things like $L^p + L^q$ or $L^p \cap L^q$ and how useful this is and how it shows up)

3. "perfect approximatinos" via sequences: At some point, we all must have wondered how in the world can we find the area of a curved shape(?!). We found out when learning through Riemann integrals that the key idea is that we can approximate it with rectangles, and then make the approximation "perfect". This idea of being able to find some simpler objects and use them to approximate more complicated ones is at the heart of analysis. For example, those rectangles we used in Riemann integration can be thought of as characteristic functions. Thus, something is Riemann integrable if the "upper" and "lower" characteristic functions, when taking the limit of the partitions, are equal.

   More generally, we will often be working over many different types of functions (most of this book will in fact be about how to classify different function space). One key part of working with function spaces will often be to find a subset of those functions that we can use to approximate any other function: such a set is called a *dense set*. Dense sets are sort of like a basis in vector-spaces: just like any vector in a k-vector space is a k-linear combination of basis elements, elements in most function space will be the result of a *sequence of functions* where each function in the sequence comes from the dense set.

   Two major examples of this is in the theory of integration and Fourier analysis. In integration, we will approximate any integrable function via *characteristic functions*, and in the theory of Fourier transforms, we will approximate integrable functions via its *fourier series*, where each function in the fourier series will be part of a special dense set we will define later.

4. Estimates: having given numerical values to function and having interpreted them through interpolation or a good sequence, we can now often solve problems by bounding the "badness" of the thing we are measuring. A super-simple example is that if f is bounded, then there exists a constant C such that
$$|f(x)| \leq C$$
   for all x in the codomain. Naturally, as we do more analysis, the estimates will be more sophisticated, but the idea remains the same: we will simplify our problems by bounding what we is bad and trying to "eliminate" it. One of the ways we will do this is by figuring out how common operations (such as $\frac{d}{dx}$, $\int$, $+, \cdot$, and so forth) which usually "complicate" our problem can be eliminated with some appropriate estiamtes. For example, let's say f is a bounded function, and we want to work with:
$$\left| \int_D f \right|$$
   then this can be a hard value to find. Howver, we can eliminte the complexity by doing:
$$\left| \int_D f \right| \leq \int_D |f| \leq \int_D C$$

which naturally now is much easier to work with. Of course, we might have over-simplified our problem with this estimate, which is why we will have a lot of different estimates playing different roles depending on our situation.

Preliminary Integration

In pre-university education, it is taught how to find the area of some simple shapes like rectangles, triangles, cubes, spheres, and so on. Later on, You learn how many of these shapes can be captured as the area under a particular graph (the graph usually being continuous). This means that the notion of area can be represented by functions (in particular the graph of a functions), and so it becomes a meaningful question to ask about how to find the area under the graph.

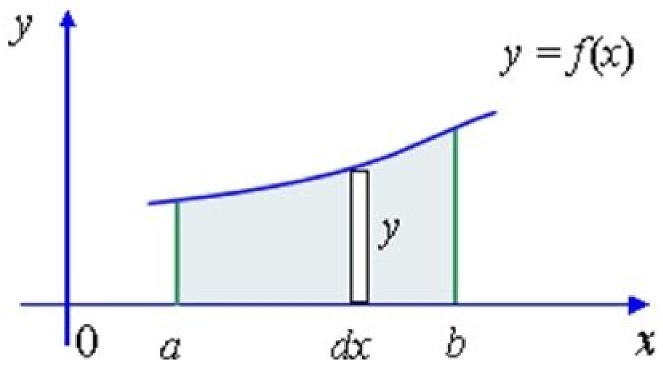


Figure 1: integrating under a graph

Let's say that f is the 1 (or more) dimensional continuous function over $\mathbb{R}$ (or $\mathbb{R}^n$), and we are trying to find the area under the graph. The first attempt at this might be the following (called Darboux sums): Given a partition the domain $P = \{x_1, x_2, ..., x_n\}$, we can make many rectangles which are all easy to measure. Then, as the partition gets finner (size(P) $\to 0$, $|P| \to \infty$), we will get a better an better approximation of the area.

If we take this approach, we will introduce some level of choice: there are many ways in which this partition can be formed, as well as a choice of the "height" of the rectangle. For example, you can choose the lowest or highest point:

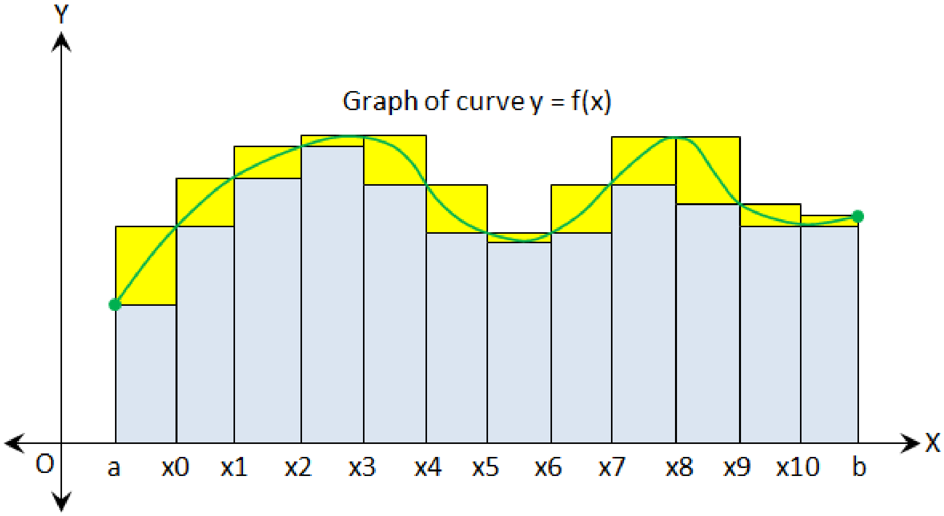


Figure 2: lower and upper sums

You can also choose any points in between, though what "in between" means starts to get hard to define in higher dimensional spaces where different orders can be put onto the "surfaces" (or hyper-surfaces) of the top of the rectangles (i.e. hyper-rectangles). However, maximal and minimal point is always well defined on these surfaces (or better, inf and sup always being well defined on these bounded surfaces), and even better, choosing the maximal point for each rectangle height upper-bounds all possible other choices of Darboux sums, and the minimal point for each rectangle's height lower-bounds all possible choices of Darboux sums. To be more precise, we'll replace the maximum and minimum condition with sup and inf, in which case the resulting sum is called the *Riemann sum*.

Intuitively, the upper and lower Riemann sums must match up and exist as $\text{size}(P) \to 0$ and $|P| \to \infty$. It would be strange if they don't match up, and so we can (as a start) define a function that is not "too weird" as one where the upper and lower Darboux sums match as $\text{size}(P) \to 0$, that is

$$\lim_{n\to\infty} \sum_{i=1}^{n-1} \left(\inf_{[x_i, x_i+1]} f(x) \right) \cdot (x_{i+1} - x_i) = \lim_{n\to\infty} \sum_{i=1}^{n-1} \left(\sup_{[x_i, x_i+1]} f(x) \right) \cdot (x_{i+1} - x_i)$$

If this is the case, f is said to be *Riemann integrable*. Fortunately, it turns out that if f is Riemann integrable on one partition, it is Riemann integrable on all partitions, and so we don't need to "remember" our initial choice of partition when we say f is Riemann integrable (we don't say "f is Riemann integrable on partition p, while f is not Riemann integrable on partition q").

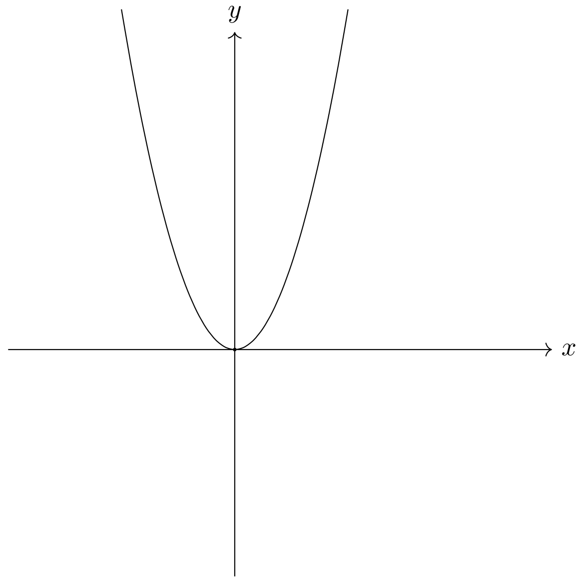
All continuous functions over a bounded domains will be integrable (so trigonometric functions, polynomials, log, $\sqrt{\quad}$ over the appropriate domain, and so forth). Since any geometric shape learnt in pre-university education can be bounded by a continuous function, we have just generalized the idea of measuring those areas[1]. In fact, being continuous *almost* encapsulates all possible functions.

[1]though we won't go into the details of finding those areas

To understand the generalization, consider

$$f_{\text{no } 1} : [-1, 1] \to \mathbb{R}, \qquad f_{\text{no } 1}(x) \begin{cases} x^3 & x \neq 1 \\ 0 & x = 1 \end{cases}$$

Notice that there is a gap in the line



however, this gap has "zero measure", we would think that the result should be the same. And in fact it is! The function $f_{\text{no } 1}$ is also Riemann integrable (as can be checked) and will have the same value as x^3 on $[-1, 1]$. In general, if finitely (or even countably) many points are ommited, then f will still have the same measure. Since $f_{\text{no } 1}$ is no longer continuous, but it as just off by some finite number of points, we say that f is continuous *almost everywhere* [2]. It should be reviewed that f is Riemann integrable if and only if f is continuous almost everywhere.

Stepping aside from measuring area under a graph for a moment, we turn to the study of another very common notion in analysis: limits. Limits are of fundamental importance in Analysis. They are the tool used to construct or define many different types of analytical objects (ex. $\mathbb{R}$, continuity, differentiability, Fourier series, sequences and series). More generally, the concept (and formalisation) of a limit is how mathematicians get to study infinites which are "well-defined". In my mind, limits are one of the distinguishing feature between algebra and analysis: many algebraic properties "break down" or are different when we try to bring in more than finitely many objects (ex. products of infinite groups cannot be direct products in **Grp** since $1 + 1 + \cdots$ is not defined (hence they are direct sums), the dimension of $\mathbb{Q}[x]$ and $\mathbb{Q}[[x]]$ as $\mathbb{Q}$-algebras differ, etc.). A good definition of limit is important to be able to define objects "at infinity", since if it is not well defined, then we can make the "result" non-unique.

Here is a silly example of that: what is $\infty - \infty$? Maybe we can say that the answer is 0? or ∞? If we are not careful with our definition, it can actually be any number. First, "notice" that $1 + 1/2 + 1/3 + \cdots \text{ “}=\text{''} \infty$, or more precisely, the series $S_n = \sum 1/n$ is unbounded from above. Since

[2]Later, almost every continuous will mean continuous up to a zero measure set once we properly define a measure

these two are "equal", we can treat them as the same. Then assuming the usual algebraic properties of distributivity work over countably many components, we hvae

$$(1+1/2+1/3+\cdots)-(1+1/2+1/3+\cdots)=(1+1/2+1/3+\cdots)-1-1/2-1/3-\cdots$$

Now, pick any $x \in \mathbb{R}$. Then this sequence can be "rearranged" to converge to that number. Start by picking 1. If $x > 1$, pick $1/2$, if it's less pick -1. Keep doing this, and this will evidentially converge to x! Thus $\infty - \infty$ can be any real number we want!

This is why the idea of a limit is important: If the limit exists, we mean that there is a unique value for which the limit will converge to. This is why limits are used to define objects with some "countably infinite" properties [3].

One important consequence of limits being unique is when we use limits to construct new functions from a limit of functions! As review, you should check the construction of a *continuous but nowhere differentiable function*. Another example is with Fourier series and Fourier transformations which are used to encode and decode signals. Another example the definition of compact exhaustion for imperfect integrals. There are many more yet to come as well!

Now, returning to integrals for a moment, we can ask ourselves if integrals commute with limits, that is

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

where the limit here is *pointwise limit*. Note that uniform limits (if it exist) will sort of commute:

$$\text{uniflim}_{n\to\infty}\int f_n = \int \lim_{n\to\infty} f_n$$

However, point-wise limit doesn't generally do

Example 0.1: Limit Doesn't Commute

Take $\chi_{\mathbb{Q}} : [0,1] \to [0,1]$ to be the indicator function on $\mathbb{Q}$, that is

$$\chi_{\mathbb{Q}}(x) = \begin{cases} 1 & x \in \mathbb{Q} \cap [0,1] \\ 0 & \text{otherwise} \end{cases}$$

Notice that $\chi_{\mathbb{Q}}$ is nowhere continuous, and hence is not Riemann integrable. We can see this more clearly by noticing that the upper Riemann sum is 1 and lower Riemann sum is 0. However

$$g_n(x) = \begin{cases} 1 & \text{if } \frac{p}{q},\ p,q \in \mathbb{N}, p \le n \\ 0 & \text{otherwise} \end{cases}$$

is integrable, and $g_n \to \chi_{\mathbb{Q}}$, that is $\lim_{n\to\infty} g_n = \chi_{\mathbb{Q}}$. Notice that each g_n are integrable, but $\chi_{\mathbb{Q}}$ is not, that is

$$\lim_{n\to\infty}\int g_n \neq \int \lim_{n\to\infty} g_n$$

Since the right hand side is not defined.

[3] Object with some "uncountably infinite" or larger properties use a generalisation of a limit called an *ultra-filter* which we don't need to study in this course

Thus, our notion of integrability does not work will with limits! Thus, since integrals gives us useful ways of analyzing our functions, and many functions are constructed as the limit of functions, we are going to spend lots of time upgrading our notion of integrability to make it commute (under appropriate circumstances) with limits.To upgrade the notion of integration, there must be some requirements we must keep in mind while finding a suitable replacement. In particular

1. The area under the graph should be the same as the Riemann Integral
2. Geometric intuition's should still be consistent, in particular
   (a) if $f \leq g$ then $\int f \leq \int g$
   (b) $\int af + bg = a\int f + b\int g$
3. A set A can be thought to be "integrable" if we define

$$\chi_A(x) \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

   and $\int \chi_A$ is defined.

This last criterion is actually very interesting as it allows us to generalize the question from finding the area's under graphs (or measuring the area under the graph) to figuring out how to measure sets in general! In fact, starting by defining how to measure sets and then defining how to integrate function turns out to be a fruitful order in which we can learn about this generalisation of integration. The reason this order is as actually an insight by Lebesgue: defining the properties of something being "measurable" will allow us to re-define the integral with more rigour. Even better, many different spaces can have different notions of a "measure" (for example, finite spaces where we measure points, or probability spaces where sets represent the probability of an event, or physics where we might want to measure the average density of a set), which the integral is the wrong concept (it's no longer strictly about "area"), but follows the same geometric intuitions as these other spaces (as we will show). Therefore, we will start by defining the spaces over which we will want to define a way to measure sets.

Chapter 1

Measures

As a start, we might want to define a reasonable definition of what it means to measure any subset of $\mathbb{R}^n$. Such a function would be of the form $\mu : P(\mathbb{R}^n) \to [0, \infty]$. If μ were to exist, we would want our common geometric intuitions to hold, that is:

1. If $E_1, E_2, ..., E_n, ...$ is a countable collection of pair-wise disjoint subsets, then
$$\mu\left(\bigcup_i E_i\right) = \sum_i \mu(E_i)$$
2. If E and F are sets such that there is a function that only translates, rotates, and reflects E to get to F (i.e. a rigid motion), then $\mu(E) = \mu(F)$
3. If C is a unit cube in $\mathbb{R}^n$, then $\mu(C) = 1$

However, these three axioms are inconsistent with one another! We have actually shown this by showing if μ is defined on $P(\mathbb{R}^n)$ (even when $n = 1$), then we can construct *Vitali sets* that will have to be both 0 and ∞ in measure – a contradiction.

Theorem 1.0.1: Vitali Sets Are Unmeasruable

Vitali sets are un-measurable

Proof **:**
Take $[0, 1]$. Partition this set as follows:

$$x \sim y \;\Leftrightarrow\; x - y \in \mathbb{Q}$$

In a sense, this is the set of cosets of $\mathbb{R}/\mathbb{Q}$ restricted to $[0, 1]$; we will rely on this intuition to label the equivalence classes. Since $\mathbb{R}/\mathbb{Q}$ is the set of irrational numbers modulo rational numbers, let N_q be an equivalence class of contains x that is equivalent to the irrational number q. Notice that there are uncountably many such equivalence relations.

Now, pick one $p \in N_q$ for each equivalence class, and let N be the collection of all these p's, so N contains one point from each equivalence class (notice that the axiom of choice must be used to create a non-empty N). We'll show that N is unmeasurable.

First, for each rational number $r \in \mathbb{Q} \cap [0, 1)$, define:

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

Then clearly, $N + r \subseteq [0, 1)$ for each $r \in \mathbb{Q} \cap [0, 1)$, and furthermore, for each $x \in [0, 1)$ there exists an r such that $x \in N + r$. To show this, let's say $y \in N$ is the element that was picked from the equivalence class of of $[x]$. Then by our choice of y, $x \in N + r$ since $r = x - y$ if $x \geq y$, or $r = x - y + 1$ if $x < y$. Next, notice that the $N + r$ also forms a partition: if $(N + r) \cap (N + s) \neq \emptyset$ then $x - r$(or $x - r + 1$) would equal $x - s$ (or $x - s + 1$), but they belond to distinct equivalence classes, and so $(N + r) \cap (N + s) = \emptyset$.

We have no reached the point where we can declare our contradiction. If μ satisfies the three wanted properties, by (1) and (2):

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

for any $r \in \mathbb{Q} \cap [0, 1)$. Since $\mathbb{Q} \cap [0, 1)$ is countable and $[0, 1)$ is the disjoint union of the $N + r$'s:

$$1 = \mu([0, 1)) = \sum_{r \in \mathbb{Q} \cap [0,1)} \mu(N + r)$$

where $1 = \mu([0, 1))$ comes from (3). Now, we have two possibilities:

1. Either $\mu(N) = 0$, in which case each $\mu(N + r) = 0$ but then $0 = 1$ – a contradiction
2. Either $\mu(N) \neq 0$ (say k, so $\mu(N + r) = k$, but then $0 = \infty$ – a contradiction

In other words, no mater what $\mu(N)$ is, we will run into a contradiction!

As you can verify, this construction required that we take advantage of the countability condition (the 1st property). You might ask if weakening it to only finitely many will fix the issue, but sort of crazily it does not! In 1924, Banach and Tarski proved the following paradox now known as the *Banach Tarski Paradox*:

> Let U and V be open subsets of $\mathbb{R}^n$, $n \geq 3$. Then there exists a $k \in \mathbb{N}$ and subsets $E_1, E_2, ..., E_k$, $F_1, F_2, ..., F_k$ such that
>
> 1. all E_j's are disjoint and their union is U
> 2. all F_j's are disjoint and their union is V
> 3. There exists a rigid motion from E_j to F_j for $j = 1, ..., k$

This is kind of insane: this means that we can take a pea-size ball, and make into the size of the earth! This means that the notion of "geometry" is not actually a consistent concept in general set theory!

Fortunately for us, The sets E_j and F_j are very strange sets; most sets that are in fact measurable. To be more precise, it is consistent to say that the axiom of choice fails and that all sets are Lebesgue measurable (which will be the measure that will replace the Riemann integrability). For the purposes of this course, that means that any set we construct without the axiom of choice (and just doing normal ZF without any sneaky logical cheating) will be measurable. For more details, see

`https://math.stackexchange.com/questions/1847052/most-functions-are-measurable`

1.1 Sigma-Algebras and Properties

Since the axiom of choice is the source of the problem of unmeasurability in the previous problems, we start by defining a space on which we avoid unmeasurable sets, namely, we want to limit ourselves to *countable* choices. The definition should be pretty general, as it should accommodate measures in more general spaces. It should also be "arithmetically expandable", as in we should be able to measure smaller parts of the space (perhaps in more convenient orientations) and add them all up. Thus, we define an *algebra*:

Definition 1.1.1: Algebra and σ-Algebra

Let X be an arbitrary set. Then let $\mathcal{A} \subseteq P(X)$ be a non-empty collection of subsets such that

1. for any finite set of sets $A_1, A_2, ..., A_n \in \mathcal{A}$, $\bigcup_i A_i \in \mathcal{A}$
2. For any $A \in \mathcal{A}$, $A^c \in \mathcal{A}$.

If it is closed under countable union, we call it a σ-algebra. We will usually denote σ-algebra on X as $\mathcal{M}(X)$.

The σ-algebra will be our natural space in which we will show in the next section we can define a "measure function" without the previous contradiction.

I like to think of the finite closure as a more general way of saying that a space is "closed under a function", the function here being $\cup$. In this way, it's sort of the most general algebraic structure. Adding the fact that it's closed under countable union brings us further way from the realm of algebra, and hence it's a σ-algebra. The term sigma is very common to mean countable, for example, a space σ-compactness if it is the countable union of compact subspaces.

Definition 1.1.2: Measurable Space

Let X be a set and $\mathcal{M} \subseteq P(X)$ be a σ-algebra. Then $(X, \mathcal{M})$ is called a *measurable space* and a set in $\mathcal{M}$ is called a *measurable set*

Proposition 1.1.1: Properties of σ-algebra

Some immediate properties of a σ-algebra:

1. It is closed under countable intersection, since $\bigcap_i X_i = (\bigcup_i X)^c$, and by difference, since $X \setminus Y = X \cap Y^c$.
2. $\emptyset$ and $X \in \mathcal{A}$, since for any $E \in \mathcal{A}$, $E \cap E^c = \emptyset \in \mathcal{A}$ and $E \cup E^c = X \in \mathcal{A}$
3. Any countable collection of subsets $\{E_i\}_{i=1}^{\infty} \subseteq \mathcal{A}$ can be turned into a countable disjoint union, where
$$F_k = E_k \setminus \left(\bigcup_{i=1}^{k-1} E_i \right) = E_k \cap \left(\bigcup_{i=1}^{k-1} E_i \right)^c$$
and by what we've just shown, every $F_k \in \mathcal{A}$, and by construction $\cup_i E_i = \cup_i F_i$. We will use this technique often very soon (note that not all sets are nonempty, that proof requires more set-theory manipulation).
4. Given any family of σ-algebras on X, their intersection is also a σ-algebra. Because of this, given $Y \subseteq X$, there is always a unique smallest σ-algebra $\mathcal{M}(Y)$ that contains Y (there is at least always one, $P(Y)$, and so this concept is well-defined). $\mathcal{M}(Y)$ is called the σ-algebra generated by Y.

A lemma we'll point out right now to simplify future proofs is the following:

Lemma 1.1.1

If $Y \subseteq \mathcal{M}(X)$, then $\mathcal{M}(Y) \subseteq \mathcal{M}(X)$

Proof **:**
Since $\mathcal{M}(X)$ is a σ-algebra that contains Y, then it will contain all countable unions and compliment of Y, and hence it will contain $\mathcal{M}(Y)$ (i.e. proposition 1.1.1[4]).

Example 1.1: σ-Algebras

1. If X is any set, then $\{\emptyset, X\}$ and $P(X)$ are trivially σ-algebras. Once we define measurable function, these will be the initial and final object in measurable spaces.
2. If X is any set, then $P(X)$ is also a σ-algebra. Usually, we will use proposition 1.1.1[5] and other tools to construct a smaller σ-algebra to not work with $P(X)$ since it's too big for most cases.
3. If X is an uncountable set, define
$$\mathcal{A} = \{E \subseteq X \mid E \text{ is countable or } E^c \text{ is countable}\}$$
this set is called the σ-algebra of countable and co-countable sets. It is useful when we want to make sure that the countable union of sets is still countable (or it's compliment is) [a]

[a] Sneakily, we use the axiom of choice for countably many choices to prove this is true

Another common σ-algebra generated on a set X is related to a topology on X, that is, if $\mathcal{T}(X)$ is a topology, of X, we will define the following σ-algebra:

Definition 1.1.3: Borel σ-Algebra

Let X be a set and $\mathcal{T}(X)$ be a topology on X. Define the *Borel σ-algebra* to be

$$\mathcal{B}_X := \mathcal{M}(\{U \subseteq X \mid U \text{ is an open set contained in } X\}$$

The elements of $\mathcal{B}_X$ are called *Borel sets*

The set $\mathcal{B}_X$ will include the open sets, closed sets, countable union and intersection of closed sets, and the countable union of countable intersections of open/closed sets, et cetera.

Since the countable intersection of open sets and countable union of closed sets are not necessarily open or closed sets respectively[1], but are members in $\mathcal{B}_X$, we will give them names: A countable intersection of open sets is called a G_δ set (G for *Gebeit* in German), and a countable union of closed sets is called a F_σ set (F for *fermé* in french). The σ is for *Summe* (sum in German) since it's a union and the δ is for *Durchschnitt* (intersection in German) since it's an intersection. We can continue this pattern and define $G_{\delta\sigma}$ to be the countable union of countable intersection of open sets, $F_{\sigma\delta}$ to be the countable intersection of countable union, and so and so on and so forth for any length of interchanging σ's and δ's. In general, this "chain" of unions and intersections does not need to terminate, and a $F_{\sigma\delta}$ (as a simple example for a chain of length 2, $\mathbb{R} \setminus \mathbb{Q}$ is a $F_{\sigma\delta}$ set, but not a F_σ set; use Baire's Categories theorem). There will be cases in which it will terminate (in fact, the majority of measure's we'll work with will have this terminate after at just F_σ or G_δ "up to a zero set", see theorem 1.4.3, and example ref:HERE for a measure that does not terminate)

One of the most common Borel sets we will be working with is $\mathcal{B}_\mathbb{R}$ with the usual euclidean topology on $\mathbb{R}$. It is useful to have an of different generating sets for $\mathcal{B}_\mathbb{R}$:

Proposition 1.1.2: Generators For $\mathcal{B}_X$

Let $\mathbb{R}$ be the real numbers and $\mathcal{B}_\mathbb{R}$ be the Borel σ-aglebra on $\mathbb{R}$ given the standard euclidean topology $\mathcal{T}(\mathbb{R})$. Then the following generate $\mathcal{B}_\mathbb{R}$:

1. The set of open intervals
2. the set of closed intervals
3. the set of half-open intervals
4. the set of open rays
5. the set of closed rays

Proof **:**
These proofs are quite easy to prove, and should be done as an exercise. A fact that you might need to recall is that every open set in $\mathbb{R}$ can be written as a countable union of open intervals (hint: $\mathbb{R}$ has a countable basis).

[1] it is possible that the countable intersection/union of open/closed sets are open/closed, and so the word "necessarily" is required

Products and σ-Algebras

Recall that the product between two sets is defined as $A \times B := \{(a,b) \mid a \in A, b \in B\}$, or more generally $\{X_\alpha\}_{\alpha \in A}$ for a collection of non-empty sets X_α[2], with $X = \prod_{\alpha \in A} X_\alpha$,. As we know from Category Theory, the product comes equipped with set functions π_α that satisfy the universal property of products (or more generally, this is a limit over the discrete category). Though we have yet to give a category where σ-algebras are the objects (we need to define what the morphisms are[3]), we can already construct our set $\prod_\alpha X_\alpha$ that will be the product of the X_α in the category of measurable spaces.

Definition 1.1.4: Product σ-Algebra

Let $\{X_\alpha\}_{\alpha \in A}$ be a collection of non-empty sets and $X = \prod_{\alpha \in A} X_\alpha$. Let

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

to be the *product σ-Algebra* on $\{M_\alpha\}_{\alpha \in A}$, i.e., it is the σ-algebra generated by all of the pre-images of the sets in M_α.

It is not difficult to check that the product σ-algebra is indeed a σ-algebra. If $|A|$ is countable we usually write $\bigotimes_{i=1}^\infty M_i$, and if $|A|$ is finite, we can write $\bigotimes_{i=1}^n M_i$ or $M_1 \bigotimes M_2 \bigotimes \cdots \bigotimes M_n$. Notice that since we allow for countable unions, it is not so easy to represent $\bigotimes_{\alpha \in A} M_\alpha$ as the product of the elements of M_α, (just like for Groups in the category of **Grp**), and hence why we do not use the $\prod$ or $\bigoplus$ notation. On the other hand, the fact that only a countable union of elements is permitted will still limit the product in the number of non-trivial components, hence justifying the $\bigotimes$ notation instead of $\prod$. In fact, it is the same as $\bigoplus$, but instead of "all but finitely many", we have "all but countably many":

Proposition 1.1.3: Countable Product σ-algebra

Let $M_n = \mathcal{M}(\mathcal{E}_n)$ be a σ-algebra for $n \in \mathbb{N}$ where $\mathcal{E}_n \subseteq P(X_n)$. If $X_n \in \mathcal{E}_n$, then

$$\bigotimes_{n=1}^\infty M_n = \mathcal{M}\left(\left\{\prod_{i=1}^n E_i \,\middle|\, E_i \in M_i\right\}\right)$$

***Proof* :**
The $\supseteq$ direction is easy (Take $\bigcap_{i=1}^n \pi_i(E_i)$ for appropriate values). For $\subseteq$, Since $X_i \in M_i$, we have that every element is inside the right hand side. Make sure you see why $X_n \in \mathcal{E}_n$ makes a difference. By lemma 1.1.1, the result follows.

We can further simplify our representation of $\bigotimes_{\alpha \in S} M_\alpha$ given a generating set for each M_α (such a set always exists, namely M_α itself is always a generating set)

[2]If they were empty, then the product would be empty
[3]this is done in chapter 2

Proposition 1.1.4: Generating set for Product σ-Algebra

Let $\{M_\alpha\}_{\alpha \in A}$ be a non-empty collection of σ-algebras, and let $\mathcal{E}_\alpha \subseteq P(X_\alpha)$ be a generating set for M_α. If $\mathcal{F} = \cup_{\alpha \in A} \mathcal{E}_\alpha$, then:

$$\bigotimes_{\alpha \in A} M_\alpha = \mathcal{M}(\{\pi_\alpha^{-1}(E_\alpha) \mid E_\alpha \in \mathcal{E}_\alpha,\ \alpha \in A\}) = \mathcal{M}(\mathcal{F})$$

Furthermore, if $|A|$ is countable and $X_\alpha \in \mathcal{E}_\alpha$ for each $\alpha \in A$, then

$$\bigotimes_{\alpha \in A} M_\alpha = \mathcal{M}\left(\left\{ \prod_{\alpha \in A} E_\alpha \,\middle|\, E_\alpha \in \mathcal{E}_\alpha,\ \alpha \in A \right\}\right)$$

***Proof* :**

Dealing with the first case, the $\supseteq$ is clear since $E_\alpha \in \mathcal{E}_\alpha$ implies $E_\alpha \in M_\alpha$, so $E_\alpha \in \bigotimes_{\alpha \in A} M_\alpha$. For the $\subseteq$ direction, If $E_\alpha \notin \mathcal{E}_\alpha$, then since $\mathcal{M}(\mathcal{E}_\alpha) = M_\alpha$, some countable unions and intersections of elements of $\mathcal{E}_\alpha$ equal E_α, proving the $\subseteq$ direction.

The countable case follows proposition 1.1.3

Borel Spaces and Products

A topological space of much study in Analysis is metric spaces. As a reminder, if $\{X_\alpha\}_{i=1}^n$ is a collection of metric spaces, then $X = \prod_{i=1}^n X_i$ has the product metric which is the generalization of the euclidean metric. As was seen in a topology class, the following metrics on a product space all define the same product topology (for notational simplicity, let's work with two metric spaces X and Y and consider $X \times Y$):

1. $d((x_1, x_2), (y_1, y_2)) = \max\{d_X(x_1, y_1), d_Y(x_2, y_2)\}$
2. $d((x_1, x_2), (y_1, y_2)) = \sqrt{d_X(x_1, y_1)^2 + d_Y(x_2, y_2)^2}$
3. $d((x_1, x_2), (y_1, y_2)) = d_X(x_1, y_1) + d_Y(x_2, y_2)$

We can ask about establishing Borel sets $\mathcal{B}_X$ on $X = \prod_{i=1}^n X_i$ with X having the product measure. It is tempting to say that this naturally splits into $\bigotimes \mathcal{B}_{X_i}$, however, this is not always the case! This is mainly due to the fact that not all topological spaces have a countable basis, and since we're allowing countable intersection of open sets, this will actually create more sets than the if we take the product of the Borel spaces:

Proposition 1.1.5: Borel sets on product Metric Space

Let $X_1, ..., X_n$ be metric spaces, and $X = \prod_{i=1}^n X_i$ be the product metric space. Then

$$\bigotimes_{i=1}^{n} \mathcal{B}_{X_i} \subseteq \mathcal{B}_X$$

If the spaces X_i are separable (i.e. has a countable dense subset, and since it's a metric space it means there is a countable basis), then

$$\bigotimes_{i=1}^{n} \mathcal{B}_{X_i} = \mathcal{B}_X$$

Proof **:**
The proof of the $\subseteq$ direction rather simple. By proposition 1.1.4, $\bigotimes_{i=1}^n \mathcal{B}_{X_i}$ is generated by $\mathcal{E} = \left\{\pi_i^{-1}(U_j) \;\middle|\; U_j \text{ is open in } X_j,\ 1 \leq j \leq n\right\}$. Since the sets of $\mathcal{E}$ are open in X, by lemma 1.1.1, $\bigotimes_{i=1}^n \mathcal{B}_{X_i} \subseteq \mathcal{B}_X$.
For the converse, since each X_j is separable, each have a countable dense subset. Since X_j is a metric space, we can form a countable basis with open balls with rational radii around each points in the countable dense subsets.

As a consequence, $\bigotimes_{i=1}^n \mathcal{B}_{\mathbb{R}} = \mathcal{B}_{\mathbb{R}^n}$, since $\mathbb{R}$ has a countable dense subset.

Elementary Sets

This is a technical result needed for later (maybe introduce it as exercises?)

Definition 1.1.5: Elementary Family

Let X be a set and $\mathcal{E} \subseteq P(X)$. Then $\mathcal{E}$ is called an *elementary family* if

1. $\emptyset \in \mathcal{E}$
2. If $E, F \in \mathcal{E}$ then $E \cap F \in \mathcal{E}$
3. If $E \in \mathcal{E}$, then E^c is a finite disjoint union of elements of $\mathcal{E}$

Proposition 1.1.6: Elementary Families and Algebra

Let $\mathcal{E}$ be an elementary family. Then the collection of finite disjoint unions of element of $\mathcal{E}$ forms an algebra

Proof **:**
(TBD) Let $U, V \in \mathcal{E}$, and consider $V^c = \sqcup_{j=1}^J F_j$ (for disjoint $F_j \in \mathcal{E}$). Then

$$U \setminus V = U \cap V^c = \sqcup_{j=1}^J (U \cap F_j)$$

notice that each $U \cap F_j \in \mathcal{E}$. So

$$U \cup V = (U \setminus V) \cup V = \sqcup_{j=1}^{J}(U \cap F_j) \sqcup V$$

where $\sqcup_{j=1}^{J}(U \cap F_j) \in A$
Continuing inductively,

(also, remember that separable means countable dense subset) Once the course starts

1. There is an interseting example ith logical statements!!

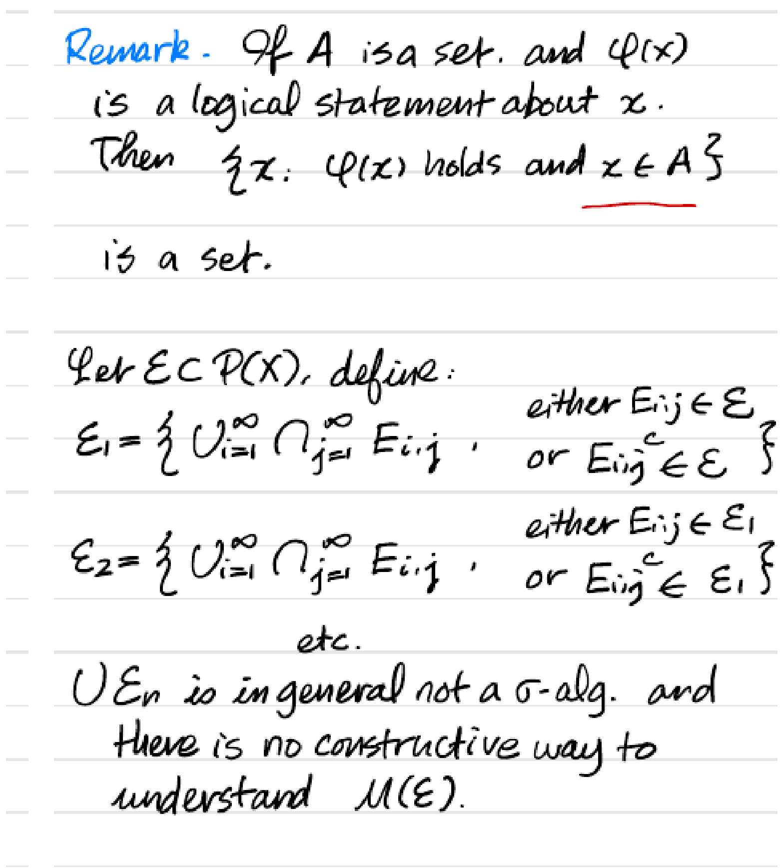


Figure 1.1: fascinating!

1.2 Measure

We can finally define the function which will give us an idea of the "size" of sets. The key idea behind the idea of notion is that the "bigger" a set is, the "larger" the the area. In other words, if μ is a measure on X, then (abusing notation a bit):

$$\mu(A \subseteq B) \;\Leftrightarrow\; \mu(A) \leq \mu(B) \tag{1.1}$$

This also means that we want the codomain of a measure to be *totally ordered*, and since we want limits to work, it also have to be *complete*. Furthermore, if we have a countable number of disjoint areas, it should possible to measure them all separately and or all together at once and get the same measure (a sort of "invariance of space"). In fact, if we just assume $\mu(\emptyset) = 0$ (i.e. "nothing" should have no area of measure), then using this "invariance of space" notion, we can deduce equation (1.1), as we'll show in the next proposition, and hence we would want the codomain of a measure to be $[0, \infty]$. We thus define a measure on these intuitions:

Definition 1.2.1: Measure

Let X be a set along with a σ-algebra $\mathcal{M}$. A function $\mu : \mathcal{M} \to [0, \infty]$ is called a *measure* on $\mathcal{M}$ (or $(X, \mathcal{M})$ or even X if the measure is clear from context) if

1. $\mu(\emptyset) = 0$
2. **(Countably Additivity)** If $\{E_i\}_{i=1}^{\infty}$ is a sequence of disjoint sets, then $\mu\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mu(E_i)$

Given that $(X, \mathcal{M})$ is a measurable space, along with a μ we can say the space has a measure:

Definition 1.2.2: Measure Space

Let $(X, \mathcal{M})$ be a measurable space. Then if μ is a measure on X, then $(X, \mathcal{M}, \mu)$ is a *measure space*

If the 2nd condition of a measure was limited to a finite sequence (or more generally, all but countably many of the sets in the sequence or nonempty), then we would say that it respects finite additivity. If μ respects finite but not countable additivity, μ is called a *finite additive measure.*

Most measure's that we will work with have some finiteness condition associated with it:

Definition 1.2.3: Finiteness Condition on μ

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

1. If $m(X) < \infty$, then μ is called a *finite measure*
2. if there exists sets $\{E_i\}_{i=1}^{\infty} \subseteq \mathcal{M}$ such that $\bigcup_i^{\infty} E_i = X$ and $\mu(E_i) < \infty$ for all E_i, then μ is called a *σ-finite measure.* More generally, if $E = \bigcup_{i=1}^{\infty} E_i$ for some sequence of E_i, then E is called σ-finite on μ (some say that E is said to be σ-finite on μ)
3. if for for all $E \in \mathcal{M}$ such that $\mu(E) = \infty$, there exists a $D \subseteq E$, $D \in \mathcal{M}$ such that $\mu(D) < \infty$, then μ is called a *semi-finite measure*

If $\mu(X) < \infty$, then this implies for all $S \in \mathcal{M}$, $\mu(S) < \infty$ (as can quickly be checked). Naturally, all finite measure's are σ-finite, and all σ-finite measure are semi-finite, but the converses are not true (as we'll explore in the following examples). Note too that most measure's we will be working over will be σ-finite (think of $\mathbb{R}$ with the intuitive notion of length of an interval).

Example 1.2: Measures

1. Let X be nonempty, $\mathcal{M} = P(X)$, and $f : X \to [0, \infty]$. Then we can define

$$\mu_f(E) = \sum_{e \in E} f(e)$$

   where the sum can be infinite (the domain of f is the positive part of the extended real numbers). Notice that f is semifinite if and only if $f(x) < \infty$ for all $x \in X$, and μ is σ-finite if and only if μ is semi-finite and $\{x \mid f(x) > 0\}$ if countable.

   If $f(x) = 1$ for all $x \in X$, then μ_f is called the *counting measure*. On the other hand, if for some $x_0 \in X$, $f(x_0) = 1$ and $f(x) = 0$ if $x \neq x_0$, then μ_f is called a *point mass* or *Dirac measure* at x_0

2. Let X be an uncountable set an $\mathcal{M}$ be the σ-algebra of countable or co-countable sets (recall example 1.1). Define

$$\mu(E) = \begin{cases} 0 & E \text{ is countable} \\ 1 & E \text{ is co-countable} \end{cases}$$

   Then μ is a measure!

3. An example of an finite additive measure that is not a measure is the following: let X be some infinite set and $\mathcal{M} = P(X)$. Define μ to be

$$\mu(E) = \begin{cases} 0 & E \text{ is finite} \\ \infty & E \text{ is infinite} \end{cases}$$

   Then μ is an finite additive measure, but not a measure

Proposition 1.2.1: Properties of Measures

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

1. **(Monotonicity)** Let $E, F \in \mathcal{M}$ and $E \subseteq F$. Then $\mu(E) \leq \mu(F)$
2. **(Subadditivity)** if $\{E_i\}_{i=1}^{\infty} \subseteq \mathcal{M}$, then $\mu\left(\cup_i \mu(E_i)\right) \leq \sum_i \mu(E_i)$
3. **(Continuity from Below)** if $\{E_i\}_{i=1}^{\infty} \subseteq \mathcal{M}$, and $E_1 \subseteq E_2 \subseteq \cdots$, then
$$\mu\left(\cup_i E_i\right) = \lim_{i \to \infty} \mu(E_i)$$
4. **Continuity from Above** if $\{E_i\}_{i=1}^{\infty} \subseteq \mathcal{M}$ and $E_1 \supseteq E_2 \supseteq$ *and* $\mu(E_1) < \infty$. Then
$$\mu\left(\cap_i E_i\right) = \lim_{i \to \infty} \mu(E_i)$$

Proof **:**
here. Do them as exercises! I also like this solution by Hytham Farah for continuity from above:

$$\begin{aligned}
\mu\Big(\bigcup_{n\geq 1} F_n\Big) &= \mu\Big(\bigcup_{n\geq 1} E_n\Big) \\
&= \mu\Big(\big(\bigcap_{n\geq 1} E_n^c\big)^c\Big) \\
&= \mu(X) - \mu\Big(\bigcap_{n\geq 1} E_n^c\Big) \\
&= \mu(X) - \lim_{n\to\infty} \mu(E_n^c) \\
&= \mu(X) - \lim_{n\to\infty} \big(\mu(X) - \mu(E_n)\big) \\
&= \lim_{n\to\infty} \mu(E_n) \\
&= \sum_{i=1}^{\infty} F_n.
\end{aligned}$$

Notice that for (d), we can instead limit it to $\mu(E_i) < \infty$ for some i, and the proof remains the same. We still require that some i exists such that $\mu(E_i) < \infty$, for it's possible that all $\mu(E_i)$ is infinite, but $\mu(\cap_i E_i) < 0$. For example, take the measurable space $(\mathbb{N}, P(\mathbb{N}))$ with the counting measure. Then the sets
$$E_i = \{n \in \mathbb{N} \mid n \geq i\}$$
then every E_i is infinite but their intersection is empty and thus has measure 0 in the counting measure.

zero measure sets

Frequently, there are sets in our measure that have a "trivial size" in the current measure. This is akin to when we integrate, the border has no effect on the value of the integral. Such sets are said to be zero sets or null set:

Definition 1.2.4: Null Set

Let $(X, \mathcal{M}, \mu)$ be a measure space and $E \in \mathcal{M}$. Then if $\mu(E) = 0$, E is said to be a *null set* or *zero set.*

By countably subadditivity, the countable union of null sets is a null set, hence null sets are "invariant" under countable union. Due to this, we can work with measurable sets in $\mathcal{M}$ *up to null sets.* We shall often state in proofs that a statement is true *almost everywhere* (often abbreviated to a.e.) if it is true on the sets except for the null sets.

If we take $E \in \mathcal{M}$ such that $\mu(E) = 0$, then by monotonicity, if $F \subseteq E$ the $\mu(F) = 0$, *provided that* $F \in \mathcal{M}$. This is not always the case:

Example 1.3: Not Complete Space

Partition $[0, 1]$ into 10 (really any number) of subintervals, $\left[\frac{i}{n}, \frac{i+1}{n}\right]$ for $0 \leq i \leq 9$, and let each of their measures be 0. Take $\mathcal{M}$ to be the σ-algebra defined on this set. Then is clearly an uncountable number of sets that do not have a measure defined on them that are subset of zero-measure sets.

However, adding all of these F's that are subsets of null sets does not break our measure. A measure who's domain includes all subsets of null sets is a *complete measure* (i.e. the σ-algebra is closed under subsets of null sets). Proving that this is still a measure is useful to eliminate future "trivial obstructions" to our proofs:

Theorem 1.2.1: Completing a Measure

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

***Proof* :**

First of all, since $\mathcal{M}$ and $\mathcal{N}$ are both closed under countable union, so $\overline{\mathcal{M}}$ is closed under countable union. To show it's closed under compliments, consider some $E \cup F \in \overline{\mathcal{M}}$ ($E \in \mathcal{M}$, $F \subseteq N \in \mathcal{N}$). Without loss of generality, assume $E \cap N = \emptyset$ (if not, take $F \setminus E$ and $N \setminus E$ and continue the proof, then this will have he same value as our original $E \cap N$). Then

$$E \cup F = (E \cup N) \cap (N^c \cup F)$$

So we get

$$(E \cup F)^c = (E \cup N)^c \cup (N \setminus F)$$

Since $E \cup N \in \mathcal{M}$, $(E \cup N)^c \in \mathcal{M}$. Since $N \setminus F \in \mathcal{N}$, Since it's closed under union, $(E \cup F)^c \in \overline{\mathcal{M}}$. Thus $\overline{\mathcal{M}}$ is a σ-algebra.

Since $\overline{\mathcal{M}}$ is a σ-algebra, we can try to define a measure, and indeed

$$\overline{\mu}(E \cup F) = \mu(E)$$

is a well-defined measure. If

$$E_1 \cup F_1 = E_2 \cup F_2$$

Then $E_1 \subseteq E_2 \cup F_2$, and so by monotonicity:

$$\mu(E_1) \leq \mu(E_2 \cup F_2) \leq (E_2 \cup N_2) = \mu(E_2) + \mu(N_2) = \mu(E_2)$$

Similarly, $\mu(E_1) \geq \mu(E_2)$. Thus $\overline{\mu}(E_1 \cup F_1) = \overline{\mu}(E_2 \cup F_2)$, and so $\overline{\mu}$ is well-defined. Since $\overline{\mu}$ has all subsets of zero-sets in its domain, it is a complete measure.

Furthermore, if $\overline{\nu}$ was another complete measure on $\mathcal{M}$, then it is easy to see that $\overline{\nu} = \overline{\mu}$, (since if it were different, then some subset of a zero-set must have non-zero measure, a contradiction to monotonicity) showing that $\overline{\mu}$ is unique.

Definition 1.2.5: Completion of Measure

Let $(X, \mathcal{M}, \mu)$ be a measurable space. Then $\overline{\mathcal{M}}$ and $\overline{\mu}$ is called the *completion of* $\mathcal{M}$ *with respect to* μ and the *completion of* μ respectively

1.3 Outer Measure

We now move on to constructing a particular type of measure which is draws inspiration from how we measured area's when we calculated Riemann Integrals. In the abstract, when we tried defining a Jordan measure (or content[4]), we approximated the inner area of a shape E in $\mathbb{R}^n$ by the limit of sums of rectangles contained in E, and the outer-area of E in $\mathbb{R}^n$ by the limit of the sums of rectangles containing E. The two numbers are called the inner and outer area of E. If the two matched, we would call the result the area of E. This intuition can be generalized to the countable case, except we can replace rectangles with some other "basic sets" that we might want that form an algebra. Furthermore, the intuition of having the "inner area" and "outer-area' is actually a bit restricting to work with in the general measure case, and so that condition will be replaced with a different measureability condition we will soon introduce.

For the sake of generality, we will start by definition a more general notion of the measure called the outer-measure, and then build-up on how form the outer-measure, we can get a measure that satisfies this generalization of Riemann integration:

[4]since "Jordan measures" only allow finite additivity, some don't like to use the word measure and opt for the word content. In this book, we would say *finite additive measure* as has already been stated earlier

Definition 1.3.1: Outer Measure

A function $\mu^* : P(X) \to [0, \infty]$ is called an *outer measure* if

1. $\mu^*(0) = \emptyset$
2. if $A \subseteq B$, then $\mu^*(A) \leq \mu^*(B)$
3. $\mu^* \left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mu^*(A_i)$ (even if the sets are disjoint!)[a]

[a]We will return to see when equality holds in definition 1.3.2

Note that we have not yet proven that an outer-measure is a measure; in particular notice that the domain is the powerset. What we will do is define a generalization of the Riemann integral that will be our "base-line" for the outer-measure:

Proposition 1.3.1: Outer Measure Construction

Let $\mathcal{E} \subseteq P(X)$ where $\emptyset, X \in \mathcal{E}$, and let $\rho : \mathcal{E} \to [0, 1]$ where $\rho(\emptyset) = 0$. For any $A \subseteq X$, define

$$\mu^*(A) = \inf \left\{ \sum_{1}^{\infty} \rho(E_j) \;\middle|\; E_j \in \mathcal{E} \text{ and } A \subseteq \bigcup_{1}^{\infty} E_j \right\}$$

Then μ^* is an outer measure

Proof **:**
First, this measure is well-defined since for any $A \subseteq X$, If we take $E_i = X$, then $A \subseteq \cup_i E_i$. Next, we verify the 3 properties

1. Clearly, $\mu^*(\emptyset) = 0$ by taking $E_i = \emptyset$ for all i
2. If $A \subseteq B$, then let $\mathcal{A}$ and $\mathcal{B}$ sets that correspond to $\mu^*(A)$ and $\mu^*(B)$. Then since all of the elements of $\mathcal{A} \supseteq \mathcal{B}$, by the property of the infimum $\inf \mathcal{A} \leq \inf \mathcal{B}$, which transferring notations gives us $\mu^*(A) \leq \mu^*(B)$
3. Let $\{E_i\}_{i=1}^{\infty} \subseteq P(X)$ be a collection of (not necessarily disjoint) subsets. Let $\epsilon > 0$. Consider $\mu^*(E_k)$ for some $E_k \in \{E_i\}_{i=1}^{\infty}$. By the property of the infimum, there must exist some $\{F_j^k\}_{j=1}^{\infty}$ such that $E_i \subseteq \cup_j F_j^k$ where $\sum_{j=1}^{\infty} \rho(F_j^k) \leq \mu^*(E_k) + \epsilon 2^{-k}$.

   Now, let $E = \cup_i E_i$. By construction, $E \subseteq \cup_{i,k=1}^{\infty} F_j^k$, and

   $$\sum_{j,k} \rho(F_j^k) \leq \mu^*(A) + \epsilon$$

   Since ϵ can be arbitrarily small, this completes the proof.

As mentioned before, an outer-measure is not necessarily a measure. What we will do to allow this to be a measure is single out sets that have the following property

Definition 1.3.2: μ^*-measureable

Let $E \subseteq X$ be a set and μ^* be an outer measure on X. Then E is called μ^**-measurable* if

$$\mu^*(S) = \mu^*(S \cap E) + \mu^*(S \cap E^c) \quad \forall S \subseteq X$$

In other word, if E can naturally "split" a set S into a part that's "inside" E or "outside" E. The set $S \subseteq X$ is called the *test set*.

By definition of the outer measure, since $S = (S \cap E) \cup (S \cap E^c)$:

$$\mu^*(S) \leq \mu^*(S \cap E) + \mu^*(S \cap E^c) \quad \forall S \subseteq X$$

Therefore, what is more important is to check the $\geq$ direction. Furthermore, if $\mu^*(A) = \infty$. Then $\geq$ holds trivially, so to show a set is μ^* measurable, it is equivalent to show that for all finite measure sets A, $\mu^*(S) \geq \mu(S \cap E) + \mu(S \cap E^c)$.

As mentioned in the definition, this gives us a sense of being able to measure the "inside" and "outside" of a set. If $E \subsetneq A$, that is the μ^*-measurable set belong in a better set, then $\mu^*(S) = \mu^*(S \cap E) + \mu^*(S \setminus E)$ which can be thought of as saying that we are measuring S from the "inside" of E and the "outisde" of E, and getting the same result. In fact, if $\mu^*(X) < \infty$, then we can define $\mu_*(E) = \mu^*(X) - \mu(E^c)$, and have that E is μ^*-measurable if and only if $\mu_*(E) = \mu^*(E)$ aligning with our intuition for inside equalling outside (see exercise ref:HERE). We do not do this method since it requires some finiteness condition on μ. We can conversely define μ_* in terms of supremums and need that E contains the collection of $\cup_i A_i$, and that the elements of $\{A_i\}_{i=1}^{\infty}$ must be disjoint (see exercise ref:HERE).

The big thing that is in the way of saying that an outer measure is a measure is that its domain is $P(X)$, which can cause problems when X is uncountable. However, all sets that satisfy μ^*-measureability turn out form a σ-algebra and make μ^* a complete measure!

Theorem 1.3.1: Carathéodory's Theorem

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

Notice that in the process of defining this measure, we started with the powerset, and then simply choose all sets that restrict (which we will show forms a σ-algebra). In this way proving a function is an outer-measure is a little less finicky since we don't need to be so careful in defining over which σ-algebra we want to define it over: the σ-algebra comes with μ^*

***Proof* :**
Let $\mathcal{M}$ be the collection of μ^*-measurable sets. Clearly, $\emptyset \in \mathcal{M}$ since for all $S \subseteq X$:

$$\mu^*(S \cap \emptyset) + \mu^*(S \cap X) = 0 + \mu^*(S)$$

implying μ^*-measureability. Next, $\mathcal{M}$ is closed under compliments, since if $E \in \mathcal{M}$, then

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

which also shows that $X \in \mathcal{M}$. Finally, we need to show that $\mathcal{M}$ is closed under countable unions. We'll start by showing finite unions, which then means it suffices to show the union of two sets is closed (due to induction). Let $A, B \in \mathcal{M}$. As we remarked, it suffices to show that for any $E \subseteq X$ such that $\mu^*(E) < \infty$ that $\mu^*(E) \geq \mu^*(E \cap (A \cup B)) + \mu^*(E \cap (A \cup B)^c)$. Since A and B are μ^* measurable, we have

$$\begin{aligned} \mu^*(E) &= \mu^*(E \cap A) + \mu^*(E \cap A^c) && E \text{ is the test set} \\ &= \mu^*(E \cap A \cap B) + \mu^*(E \cap A \cap B^c) \\ &\quad + \mu^*(E \cap A^c \cap B) + \mu^*(E \cap A^c \cap B^c) && E \cap A \text{ and } E \cap A^c \text{ are the test sets} \end{aligned}$$

Then notice we can write

$$A \cup B = (A \cap B) \cup (A \cap B^c) \cup (A^c \cap B)$$

Thus, using subadditivity, we can re-write 3 of the terms in the previous expression as:

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

Since $E \cap A^c \cap B^c = E \cap (A \cup B)^c$, we get:

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

showing that $\mathcal{M}$ is closed under finite union. Since $\mathcal{M}$ is closed under compliment and contains $\emptyset$, $\mathcal{M}$ is an algebra. Even better, If we let A and B be disjoint, and let $A \cup B$ be the test set, then since A is μ^* measurable:

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

showing that $\mu^*|_{\mathcal{M}}$ is a finitely additive measure.

Next, we must show that $\mathcal{M}$ is a σ-algebra, which as a consequence will show that $\mu^*|_{\mathcal{M}}$ is a measure. Let $\{E_i\}_{i=1}^{\infty} \subseteq \mathcal{M}$. To simplify our task, re-construct $\{E_i\}_{i=1}^{\infty}$ to be the set of disjoint sets $\{D_i\}_{i=1}^{\infty}$. Let $B_n = \bigcup_{i=1}^{n} D_i$ and $B = \bigcup_{i=1}^{\infty} D_i$. Our goal is to show that for any $E \subseteq X$ such that $\mu^*(E) < \infty$ that $\mu^*(E) \geq \mu^*(E \cap B) + \mu^*(E \cap B^c)$. We will start by considering $\mu^*(E \cap B_n)$:

$$\begin{aligned} \mu^*(E \cap B_n) &= \mu^*(E \cap B_n \cap D_n) + \mu^*(E \cap B_n \cap D_n^c) && E \cap B_n \text{ is the test set} \\ &= \mu^*(E \cap D_n) + \mu^*(E \cap B_{n-1}) \end{aligned}$$

Notice that on the right hand side we have diminished to the case of $E \cap B_{n-1}$ from $E \cap B_n$. Using induction, we can see that the expression simplifies to

$$\mu^*(E \cap B_n) = \sum_{i=1}^{n} \mu^*(E \cap D_i)$$

Continuing, since B_n is measurable

$$\mu^*(E) = \mu^*(E \cap B_n) + \mu^*(E \cap B_n^c) \overset{!}{\geq} \mu^*(E \cap B_n) + \mu^*(E \cap B^c) = \sum_{i=1}^{n} \mu^*(E \cap D_i) + \mu^*(E \cap B_n^c)$$

where $\overset{!}{\geq}$ comes from monotonicity after switching B_n^c to B^c. Since this equality holds true for all

n, we get:

$$\begin{aligned}
\mu^*(E) &\geq \sum_{i=1}^{\infty} \mu^*(E \cap D_i) + \mu^*(E \cap B^c) \\
&\geq \mu^*\left(\bigcup_i E \cap D_i\right) + \mu^*(E \cap B^c) && \text{subadditivity} \\
&= \mu^*(E \cap B) + \mu^*(E \cap B^c) && \text{by construction of } E \cap B \\
&\geq \mu^*(E) && \text{subadditivity}
\end{aligned}$$

Since we got $\mu^*(E) \geq \mu^*(E)$, we see that in fact all inequalities are equalities, getting us that

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

showing us that $B \in \mathcal{M}$. Furthermore, replacing $E = B$ in the previous proof shows us that

$$\mu^*(B) = \sum_{i=1}^{\infty} \mu^*(D_i)$$

giving us that μ^* is a measure on $\mathcal{M}$.

Finally, we must show that $\mathcal{M}$ is complete, that is, if $E \in \mathcal{M}$ such that $\mu^*(E) = 0$, then all subsets of E are in $\mathcal{M}$. Notice that since μ^* is defiend on all of $P(X)$, it is equivalent to show that if $\mu^*(A) = 0$, then $A \in \mathcal{M}$. Then for any test set E, by monotonicity:

$$\mu^*(E) \leq \mu^*(E \cap A) + \mu^*(E \cap A^c) = 0 + \mu^*(E \cap A^c) \leq \mu^*(E)$$

meaning equality holds, showing that $A \in \mathcal{M}$. Thus, $\mu^*|_{\mathcal{M}}$ is a complete measure, as we sought to show.

Carathéodory's Theorem is quite useful being able to state the existence of a measure given an outer-measure, but it's not good to construct a measure (for example, can you tell me a non-trivial element in $\mathcal{M}$ without directly appealing to the definition?). The following concept is used to more systematically construct the set $\mathcal{M}$ and the measure μ induced by μ^*:

Definition 1.3.3: Premeasure

Let $(X, \mathcal{A})$ be an algebra. Then μ_0 is called a *premeasure* if

1. $\mu_0(\emptyset) = 0$
2. **"Respects" countable union**: If $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{A}$ is a collection of disjoint such that $\cup_i A_i \in \mathcal{A}$, then $\mu_0(\cup_i A_i) = \sum_i \mu_0(A_i)$

Note that the premeasure is automatically finitely additive since we can let all but finitely many of the terms in (2) be $\emptyset$ (and hence are a stronger condition than finitely additive measure). However, it is not yet a measure since it's domain is not necessarily closed under countable union. However, if it is closed for countable union for a particular collection, that collection must satisfy the "countable" condition. Interestingly, by the definition of an algebra, this implies that the countable union can be represented as a finite union and intersection of elements from $\mathcal{A}$.

As before, we can define semifinite and σ-finite in a similar way. Using μ_0, we can define the following outer-measure:

$$\mu^* : P(X) \to [0, \infty] \quad \mu^*(E) = \inf \left\{ \sum_i^\infty \mu_0(A_i) \;\middle|\; A_i \in \mathcal{A},\ E \subseteq \cup_i A_i \right\}$$

Proposition 1.3.2: Premeasure To Measure

Let μ_0 be a premeasure on $\mathcal{A}$, and let μ^* be defined as above. Then

1. $\mu^*|_{\mathcal{A}} = \mu_0$
2. Every set in $\mathcal{A}$ is μ^*-measurable

Essentially, μ^* is an extension of μ_0 and the associated restriction of μ^* to a measure will contain $\mathcal{A}$

***Proof* :**

1. Let's show $\leq$ and $\geq$ for any $E \in \mathcal{A}$. The $\leq$ direction is obvious since $\mu^*(E) \leq \mu_0(E)$ since $E \subseteq \bigcup_{i=1}^\infty A_i$ where $A_1 = E$ and $A_i = \emptyset$ for $i > 1$, and so it is part of the set over which we take the infimum, and so the result *has* to be smaller.

   For the $\geq$ direction, let's say $E \subseteq \bigcup_{i=1}^\infty A_i$ for any $\{A_i\}_{i=1}^\infty \in \mathcal{A}$. Let $B_n = E \cap \big(A_n \setminus \cup_1^{n-1} A_i\big)$. Then the collection $\{B_i\}_{i=1}^\infty$ is a disjoint collection of members of $\mathcal{A}$ whose union is E. Thus, by the property of premeasure

   $$\mu_0(E) = \sum_1^\infty \mu_0(B_i) \overset{!}{\leq} \sum_1^\infty \mu_0(A_i)$$

   where the $\overset{!}{\leq}$ inequality comes from the fact that $\mu_0(B_i) \leq \mu_0(A_i)$. Since this is true for any arbitrary collection $\{A_i\}_{i=1}^\infty$, it follows that

   $$\mu_0(E) \leq \mu^*(E)$$

   Thus establishing equality

2. We need to show that for any $E \subseteq X$ and $A \in \mathcal{A}$, $\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c)$. As mentioned earlier, it suffice to prove $\geq$ where $\mu^*(E) < \infty$.

   Let $\epsilon > 0$. By the property of the infimum, there exists a subset $\{B_i\}_{i=1}^\infty \subseteq \mathcal{A}$ where $E \subseteq \cup_i B_i$ such that $\sum_i^\infty \mu_0(B_i) \leq \mu^*(E) + \epsilon$. By part (1), μ^* is [countably] additive on $\mathcal{A}$ (since it equals the pre-measure)

   $$\begin{aligned} \mu^*(E) + \epsilon &\geq \sum_{i=1}^\infty \mu_0(B_i \cap A) + \sum_{i=1}^\infty (B_i \cap A^c) && \text{additivity} \\ &\geq \mu^*(E \cap A) + \mu^*(E \cap A^c) && \text{monotinicity} \end{aligned}$$

   and since ϵ was arbitrary, A must be μ^*-measurable, completing the proof

We can in fact be more precise with the measures that extend a premeasure:

Theorem 1.3.2: Premeasure Extensions

Let $\mathcal{A} \subseteq P(X)$ be an algebra, μ_0 a premeasure on $\mathcal{A}$, and $\mathcal{M} = \mathcal{M}(\mathcal{A})$ (i.e. $\mathcal{M}$ is the σ-algebra generated by $\mathcal{A}$). Let μ be the extension of μ_0 from proposition 1.3.2. Then if ν is another measure on $\mathcal{M}$ that extends μ_0, then

$$\nu(E) \leq \mu(E) \qquad \forall E \in \mathcal{M}$$

with equality when $\mu(E) < \infty$ (i.e., it possible that μ is infinite while ν is finite). If μ_0 is σ-finite, then μ is the unique extension of μ_0 to a measure on $\mathcal{M}$

In other words, we can extend a pre-measure to a measure by knowing what it does on the algebra (which in most cases will be a collection of simpler sets, like intervals or rectangles), and then what happens on the rest of the measurable sets we use the infimum definition. If we define another extension of μ_0, then that measure will be *finer* than the μ given by μ^*, i.e, μ is the *coarsest* extension. If μ_0 is σ-finite, then this will in fact be the only extension of μ_0

Proof **:**
Let's say ν is an extension of μ_0 so that $(X, \mathcal{M}(\mathcal{A}), \nu)$ is a measure space. To show that $\nu(E) \leq \mu(E)$, let's say $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{A}$ where $E \subseteq \cup_i A_i$. Then since ν is a measure, it is sub-additive on non-disjoint sets, and so

$$\nu(E) \leq \sum_i \nu(A_i)$$

Since ν extends μ_0, by proposition 1.3.2 $\nu|_{\mathcal{A}} = \mu_0$, so

$$\nu(E) \leq \sum_i \nu(A_i) = \sum_i \mu_0(A_i)$$

Since $\{A_i\}_{i=1}^{\infty}$ was an arbitrary choice, this works for any collection, and so

$$\nu(E) \leq \mu(E) \quad [= \mu^*(E)]$$

We show that $\nu(E) = \mu(E)$ if $\mu(E) < \infty$. Let $A = \cup_i A_i$ for some collection $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{A}$. Then

$$\nu(A) = \lim_{n\to\infty} \nu\left(\bigcup_{i=1}^{n}(A_i)\right) = \lim_{n\to\infty} \mu\left(\bigcup_{i=1}^{n}(A_i)\right) = \mu(A)$$

If μ is not σ-finite, then there can be more than one extension:

Example 1.4: 2 extensions of pre-measure

1. Let $\mathcal{A}$ be the collection of finite unions of the sets of the form $(a, b] \cap \mathbb{Q}$ with $-\infty \leq a \leq b \leq \infty$. Then $\mathcal{A}$ is an algebra on $\mathbb{Q}$

***Proof* :**
Recall proposition 1.1.6 where we can construct an algebra using an elementary family. Thus, we will show that the collection $\mathcal{A}$ satisfies the 3 properties of an elementary family:

(a) To get the empty-set in $\mathcal{A}$, we will understand the notion of "finite union" to also include "no union". Otherwise, since $\mathbb{Q}$ is dense in $\mathbb{R}$, all intervals $(a, b]$ intersected with $\mathbb{Q}$ is nonempty (by MAT157 knowledge). Since $a < b$, the singleton interval is not admissible, and so we will have to assume this "no union" property

(b) Let $E, F \in \mathcal{A}$. We want to show that $E \cap F \in \mathcal{A}$. By definition, $E = (a, b] \cap \mathbb{Q}$ and $F = (c, d] \cap \mathbb{Q}$. Then by some basic set-theory properties, we know that the intersection of two left-open/right-closed intervals is again a left-open/right-closed interval or $\emptyset$. If their intersection is $\emptyset$, then we're done. If not, let $(e, f] = (a, b] \cap (c, d]$. Then $E \cap F = (e, f] \cap \mathbb{Q}$. Thus, $E \cap F \in \mathcal{A}$

(c) Let $E \in \mathcal{A}$, so that $E = (a, b] \cap \mathbb{Q}$. Let $i \in \mathbb{N}$ ($0 \notin \mathbb{N}$) and:

$$E_i = (a - i, a - (i + 1)] \cap \mathbb{Q} \quad F_i = (b + (i - 1), b + i] \cap \mathbb{Q}$$

then each $E_i, F_i \in \mathcal{A}$, $E_i \cap E_j = F_i \cap F_j = \emptyset$ if $i \neq j$ and $E_i \cap F_j = \emptyset$ for all $i, j \in \mathbb{N}$. Let $E = \{E_i\}_{i=1}^{\infty}$ and $F = \{F_i\}_{i=1}^{\infty}$ and let $G = E \cup F$. Then

$$\bigcup_{g \in G} G = E^c$$

since G is a countable union of disjoint elements of $\mathcal{A}$, this completes the proof.

Thus, by proposition 1.7, $\mathcal{A}$ is an algebra.

2. The σ-algebra generated by $\mathcal{A}$ is $P(\mathbb{Q})$

***Proof* :**
Let $x \in \mathbb{Q}$ Notice that

$$\{x\} = \bigcap_{k=1}^{\infty} \left(x - \frac{1}{k}, x\right] \cap \mathbb{Q}$$

hence, every singleton is inside the σ-algebra of $\mathbb{Q}$. Since every subset of $\mathbb{Q}$ is countable, every subset is the countable union of singletons. Since every singleton is inside the σ-algebra, we have that the σ-algebra is $P(\mathbb{Q})$.

3. Define μ_0 on $\mathcal{A}$ by $\mu_0(\emptyset) = 0$ and $\mu_0(A) = \infty$ for $A \neq \emptyset$. Then μ_0 is a premeasure on $\mathcal{A}$, and there is more than one measure on $P(\mathbb{Q})$ whose restriction to $\mathcal{A}$ is μ_0

Proof **:**
μ_0 is trivially a pre-measure. By definition, $\mu_0(\emptyset) = 0$. For any nonempty $A \in \mathcal{A}$, we have $\mu_0(A) = \infty$. When checking countable additivity, unless all the sets are empty, we get that both sides are infinite.

Thus, we can define μ to be the μ^* induced by μ_0 restricted to all measurable sets. By construction, all sets are measurable. If a sets is non-empty, then it will always be covered by a nonempty set, which has measure ∞. Thus, If $E \in P(X)$ for any nonempty test set $S \subseteq \mathbb{Q}$

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

and if $S = \emptyset$

$$0 = \mu^*(S) = \mu^*(S \cap E) + \mu^*(S \cap E^c) = 0$$

Thus, $\mathcal{M}_\mu = P(\mathbb{Q})$.

We can define another extension of μ_0 using the counting measure, in particular, let μ_C be the counting measure on $P(\mathbb{Q})$. Then since every $A \in \mathcal{A}$ has countable cardinality

$$\mu_C(A) = \infty$$

Thus, $\mu_C|_{\mathcal{A}} = \mu_0$. However

$$\mu_C(\{1\}) = 1 \neq \infty = \mu(\{1\})$$

Thus, we have defined two separate extensions of μ_0.

1.4 Borel Measure on the Real Line

The full title of this seciton should really be "Borel Mearues on the Real line defined through Distributions", but that title seems to cumbersome. Nevertheless, this is exactly what we will be doing in this section. We will extensively concentrating on these types of measures, since they are, in a sense, the focus of "real analysis". From a high-level persepctive, the fact that we are studying "real analysis" justifies why we will take so much time understanding this measure. More concretly, these measure are pervasive in analysis in general: they are central to defining integration, to the study of Fourier series, to studying L^p spaces, to linking differentiation to integration, to generalizing the notation of itegration of groups (in particular, it will make sense to integrate locally compact topological groups), and so much more. Another way of looking at it is that the euclidean topology is central to the study of many fields of mathematics and physics, and is the basis for our notion of manifolds (which are locally euclidean). Therefore, getting a sense of how to define notions of "size" in $\mathbb{R}$ (and then generalize to $\mathbb{R}^n$, $\mathbb{R}^\infty$, or manifolds) gives us a solid first step in gaining an intuition on measures.

One way to define a measure in $\mathbb{R}$ is by using our intuition of the fact that the length of (a, b) is $b - a$. This is a good intuition and will be the basis to the generalization of this idea we'll do in order to be able to define a larger class of measures based on this idea which will also have more uses outside of $\mathbb{R}$. Here is how we will motivate the construction of these measures: let μ be a finite measure on Borel sets (a measure define on Borel sets is called the *Borel Measure*). Define:

$$F(x) = \mu((\infty, x])$$

F is sometimes called the *distribution function* of μ. Intuitively, it shows the "accumulated size" of the space. Clearly, F is an increasing function, and is right-continuous since

$$(\infty, x] = \bigcap_k (\infty, x + 1/k]$$

or any decreasing function, $x_n \searrow x$ (i.e. limits are preserved from the right). Furthermore, if we take $a < b$, then $(-\infty, b] = (-\infty, a] \cup (a, b]$, thus

$$\mu((a, b]) = F(b) - F(a)$$

Notice how similar this is to the usual result from Riemann integration! In the following, instead of defining F in terms of μ, we will define μ from an increasing right-continuous function F by showing how F define's a pre-measure and using Carathéodory. The special case of $F(x) = x$ will give us our usual notion of length in $\mathbb{R}$.

Throughout this section, we will use sets of the form $(a, b]$ where $a < b$, along with the edge cases of (a, ∞) and $\emptyset$. Let all intervals of this type be called h-intervals (h for half open). It is clear that the union or compliment of two h-intervals are h-intervals (or two disjoint h-intervals) and that h-intervals are closed under finite operations, and so h-intervals from a algebra $\mathcal{A}$. By proposition 1.1.2, the σ-algebra by the set $\mathcal{A}$ is $\mathcal{B}_{\mathbb{R}}$. Using this, we'll first show that we can define a premeasure μ based off of an F:

Proposition 1.4.1: Distribution and Premeasure

Let $F : \mathbb{R} \to \mathbb{R}$ be increasing and right continuous. If $(a_i, b_i]$ is a set of disjoint h-intervals. If we define μ_0 to be:

$$\mu_0 \left(\bigcup_i^{\infty} (a_i, b_i] \right) = \sum_{i=1}^{\infty} F(b_i) - F(a_i)$$

and $\mu_0(\emptyset) = 0$, then μ_0 is a premeasure on the algebra $\mathcal{A}$

Proof **:**
We will first show that μ_0 is well-defined, then we simply must show that $\mu\left(\bigcup_i (a_i, b_i]\right) = \sum_i \mu_0((a_i, b_i])$.

To show it's well-defined, we must show that for any finite representation of $(a, b]$, $\mu_0((a, b]) = F(b) - F(a)$. To that end, let $\{(a_i, b_i]\}_{i=1}^n$ be disjoint where $(a, b] = \bigsqcup i = 1^n [a_i, b_i)$. Since the co-domain is commutative, we can re-arrange the order that we are taking the union (i.e., re-index the set) so that

$$a < a_1 < b_1 = a_2 < b_2 = a_3 < \cdots < b_n = b$$

Then the resulting μ_0 is simply an alternating series:

$$\sum_{j=1}^{n} F(b_j) - F(a_j) = F(b) - F(a)$$

showing that it is independent of representation of the intervals. Thus, if we have any disjoint collection $\{I_i\}_{i=1}^n$ and $\{J_j\}_{j=1}^n$ such that $\cup_i I_i = \cup_j J_j$, since we can always form alternative series:

$$\sum_i \mu_0(I_i) = \sum_{i,j} \mu_0(I_i \cap J_j) = \sum_j \mu_0(J_j)$$

and so μ_0 is well-defined. Since it is well-defined, then we also get that μ_0 is finitely additive (essentially by definition), and since $\mu_0(\emptyset) = 0$, it is at least a finite measure.

Next, we need to show that it is a premeasure. Since $\mu_0(\emptyset) = 0$, it remains to check that if $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{A}$ where $\cup_i^{\infty} A_i \in \mathcal{A}$, then

$$\mu_0\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu_0(A_i)$$

Since $\cup_i^{\infty} I_i = I \in \mathcal{A}$, there exists (at least one) subsets $\{A_i\}_{i=1}^{n} \subseteq \mathcal{A}$ such that $\cup_i^{\infty} I_i = \cup_i^n A_i$. Therefore the elements from I can be partitioned so that the union of each partition represents a single h-interval. By finite additivity of μ_0, it suffices to consider what happens at one of these partitions. So, let J be the union of one of the families so that $(a, b]J = \cup_{j=1}^{\infty} I_{i_j}$. We will show that $\mu_0(\cup_j^{\infty} J_j) = \sum_j^{\infty} \mu_0(J_j)$ by showing $\leq$ and $\geq$. For the $\geq$ direction, notice that

$$\mu_0(J) = \mu_0\left(\bigcup_{j}^{n} J\right) + \mu_0\left(J \setminus \bigcup_{j}^{n} J\right) \geq \mu_0\left(\bigcup_{j}^{n} J\right) = \sum_{j}^{n} \mu_0\left(J\right)$$

Thus, as $n \to \infty$ we have $\mu_0(J) \geq \sum_j^{\infty} \mu_0(J_j)$.

For the $\leq$ direction, we will take advantage of the right-continuity of F.

(Might not be the case TBD p.34 Folland) First, if either b is infinite, the result holds automatically. If a is infinite, then b would have to also be infinite, and so we have the zero-intervals. So, let a and b be finite.

Let $\epsilon > 0$. Then since F is right-continuous, there exists a δ such that $F(a + \delta) - F(a) < \epsilon$ or similarly $-F(a) < -F(a + \delta) + \epsilon$. Similarly, for the same epsilon, there exists a δ_j such that $F(b_j + \delta_j) - F(b_j) < \frac{\epsilon}{2^n}$. Notice that $(a_j, b_j] \subseteq (a_j, b_j + \delta_j)$, so the open intervals cover $[a, b]$, and since $[a, b]$ is compact, there exists a finite sub-cover! If $\{(a_j, b_j + \delta_j)\}_{i=1}^{N}$ is the finite subcover of $[a, b]$, if we discard all sets that are contained in a bigger set in this family, then we get that

$$b_j + \delta_j \in (a_{j+1}, \beta_{j+1} + \delta_{j+1}) \quad \text{for } j = 1, ...N - 1$$

With this, we get the following sequence of inequalities:

$$\begin{aligned}
\mu_0(J) &= F(b) - F(a) \\
&\leq F(b) - F(a + \delta) + \epsilon \\
&\leq F(b_N + \delta_N) - F(a_1) + \epsilon && \text{extrema of cover} \\
&= F(b_N + \delta_N) - F(a_N) + \sum_{j=1}^{N-1} (F(a_{j+1}) - F(a_j))\epsilon && \text{fancy 0} \\
&= F(b_N + \delta_N) - F(a_N) + \sum_{j=1}^{N-1} (F(b_{j+1} + \delta_j) - F(a_j))\epsilon && \text{expanding} \\
&< \sum_{j=1}^{N} [f(b_j - f(a_j) + \epsilon 2^{-n}] + \epsilon \\
&< \sum_{j=1}^{\infty} \mu_0(J_j) + 2\epsilon
\end{aligned}$$

and since ϵ was arbitrary, the result follows.

This result let's us prove that *any* increasing right-continuous function defines a measure!

Theorem 1.4.1: Distributions define Measure

If $F : \mathbb{R} \to \mathbb{R}$ is an increasing right-continuous function, there is a unique Borel measure μ_F on $\mathbb{R}$ such that $\mu_F((a,b]) = F(b) - F(a)$ for all $a, b \in \mathbb{R}$. This function is unique up to the following condition: if G is another such function, then $F - G$ is constant.
Conversely, if μ is a Borel measure that is bounded on all finite Borel sets, we can define

$$F(x) = \begin{cases} \mu((0,x]) & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -\mu((-x,0]) & \text{if } x < 0 \end{cases}$$

Then F is an increasing right-continuous function

***Proof* :**
First, by the previous proposition, F induces a premeasure on $\mathcal{A}$. If $F - G$ is constant, then the sums will cancel out in $F(b) - F(a)$, and so $\mu_F = \mu_G$. Similarly, if the two are equal, F and G can only differ up to a constant and that these measures are σ-finite (recall that $\mathbb{R} = \bigcup_{-\infty}^{\infty}(-i, i+1])$.

For the converse, note that if μ is monotonic, then so is F. Next, above/bellow continuity of μ will imply right-continuity of F for $x \geq 0$ and $x < 0$. Therefore, by construction, $\mu = \mu_F$ on $\mathcal{A}$, and so by proposition 1.4.1, $\mu = \mu_F$ on $\mathcal{B}_{\mathbb{R}}$ by uniqueness in theorem 1.3.2.

Note that F need not be left-continuous. Something with the point-density measure that proves the counter-case.

We have been using intervals of the form $(a, b]$ and right-continuous function. The same theory can be developed with $[a, b)$ intervals and left-continuous functions. If we restrict μ to being a finite Borel measure, then we get $\mu = \mu_F$ where $F(x) = \mu((-\infty, x])$ (i.e. it differs up to a constant of the μ_F when μ is a Borel measure). Next, by the construction of the measure from the pre-measure, the resulting measure is in fact a complete measure. Even better, taking completion of μ_F and then applying Carathéodory is the same as just taking completion of μ after applying Carathéodory to μ_F, and so taking the completion on μ_0 or μ_F does not make a difference! The complete measure is also usually denoted by μ_F and called the *Lebesgue-Stieljes measure* of F.

As a consequence of the Lebesgue-Stieljes measure μ being complete, if we have a measure on a Borel space (which we usually call a Borel measure), then then then the set of measurable sets $\mathcal{M}^*$ might be larger than the number of Borel sets (in particular in that it may contain some zero sets that are not Borel sets).

The Lebesgue-Stieljes measure has a couple of nice properties worth exploring. Let μ be a complete Lebesgue-Stieljes measure on $\mathbb{R}$ associated to F, and let $\mathcal{M}_\mu$ be the associated μ^*-measurable sets. Then for any $E \in \mathcal{M}_\mu$:

$$\mu(E) = \inf\left\{\sum_1^\infty F(b_i) - F(a_i) \;\middle|\; E \subseteq \cup_i^\infty (a_i, b_i]\right\} = \inf\left\{\sum_1^\infty \mu((a_i, b_i]) \;\middle|\; E \subseteq \cup_i^\infty (a_i, b_i]\right\}$$

Notice that the contribution of the right end points of every h intervals is clearly negligable and so we can replace it with open intervals

Lemma 1.4.1: h-Intervals to Open Intervals

Let μ be the complete Lebesgue-Stieljes measure on $\mathbb{R}$ associated to F. Then for any $E \in \mathcal{M}_\mu$:

$$\mu(E) = \inf \left\{ \sum_1^\infty \mu((a_i, b_i)) \;\middle|\; E \subseteq \cup_i^\infty (a_i, b_i) \right\}$$

***Proof* :**
This is the same epsilon trick you have seen before, right it down here as an exercise later

In fact, we can write down $\mu(E)$ in even simple terms

Theorem 1.4.2: Lebesgue-Stieljes in Terms of Open and Compact sets

Let μ be the complete Lebesgue-Stieljes measure on $\mathbb{R}$ associated to $\mathbb{F}$. Then for any $E \in \mathcal{M}_\mu$:

$$\mu(E) = \inf \{\mu(U) \mid U \supseteq E \text{ and } U \text{ is open}\}$$

$$\mu(E) = \sup \{\mu(K) \mid K \subseteq E \text{ and } U \text{ is compact}\}$$

***Proof* :**
By lemma 1.4.1, for all $\epsilon > 0$, there exists open intervals (a_i, b_i) such that $E \subseteq \cup_i^\infty (a_i, b_i)$ and $\mu(E) \leq \sum_i^\infty (a_i, b_i) + \epsilon$. If we let $U = \cup_i (a_i, b_i)$, we have that $\mu(U) \leq \mu(E) + \epsilon$. Conversely, since $E \subseteq U$, $\mu(U) \geq \mu(E)$, and so the result for the first statements follows.

For the 2nd equivalence, let's first suppose E is bounded. If E is closed, then it is compact and so E is contained in the set that we are taking the sup over, and so equality is immediate. If E was open, then we can choose an $\epsilon > 0$ such that $\overline{E} \setminus E \subseteq U$ and $\mu(U) \leq \mu(\overline{E} \setminus E) + \epsilon$. Set $K = \overline{E} \setminus U$. Then K is compact, and so

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

showing that any compact set arbitrarily approximates K.

If E is unbounded, let $E_i = E \cap (i, i+1]$ so that $E = \cup_i E_i >$ Then by what we've just shown, for all $\epsilon > 0$, there exists a $K_j \subseteq E_j$ such that $\mu(K_j) \leq \mu(E_j) - \frac{\epsilon}{2^n}$. Take $H_n = \cup_{-n}^n K_i$. Then H_n is compact, $H_n \subseteq E$, and

$$\mu(H_n) \geq \mu(\cup_{-n}^n E_i) - \epsilon$$

Since this equality holds when $n \to \infty$, the result follows

This gives us a strikingly easy representation of representing every measurable subset of $\mathbb{R}$!

Theorem 1.4.3: $E \subseteq \mathbb{R}$ Representation

Let $E \subseteq \mathbb{R}$. Then the following are equivalent

1. $E \in \mathcal{M}_\mu$
2. $E = V \setminus N_1$ where V is a G_δ set and $\mu(N_1) = 0$
3. $E = W \cup N_2$ where W is a H_σ set and $\mu(N_2) = 0$

***Proof* :**
Clearly, (2) and (3) imply (1). For (1) implies (2) or (3), I would highly recommend you try and solve this in the $\mu(E) < \infty$ case to jog your memory.

For the $\mu(E) = \infty$, TBD (exercise 25 p.37 Folland)

This means that all borel sets, or more generally sets in $\mathcal{M}_\mu$ (also contain all unions of borel sets with measure zero sets). Note that this theorem and proof would be much harrier if we did not assume the measure was complete. ‘

I will leave this section with another way of interpreting finite measurable sets that Folland left as an exercise (we'll use the notation of $A \triangle B = (A \setminus B) \cup (B \setminus A)$)

Proposition 1.4.2: Diminishing Difference

Let $E \in \mathcal{M}_\mu$ and $\mu(E) < \infty$. Then for every $\epsilon > 0$, there exists a set A which is the finite union of open intervals and

$$\mu(E \triangle A) < \epsilon$$

1.4.1 Lebesgue Measure

We will study the case where the measure μ is defined through the distribution $F(x) = x$, i.e., when F represents our usual intuition of length! This is arguably the most important measure on $\mathbb{R}$, and most certainly the most used. For that reason, it has some special symbology: we will denote the measure by m instead of μ and the set of measurable sets by $\mathcal{L}$. If we restrict m to $\mathcal{B}_\mathbb{R}$, we will also write it as m.

The first properties worth establishing is to see that measurable sets in $\mathbb{R}$ inherit some of our intuitions how a "shape" should act under translation and dilation, that it's invariant under translation (just like we assumed for the construction of Vitali sets!) and it scales linearly with dilation:

Proposition 1.4.3: Scaling and Dilation

Let $(X, \mathcal{L}, m)$ be a Lebesgue measure space. Define

$$E + s := \{e + s \mid e \in E\} \qquad rE := \{re \mid e \in E\}$$

Then $E + s, rE \in \mathcal{L}$ for all $r, s \in \mathbb{R}$, and:

$$m(E + s) = m(E) \qquad m(rE) = |r|m(E)$$

***Proof* :**
We essentially notice that open intervals are invariant under translation and operate in the form described in the proposition, then apply theorem 1.3.2 (which is a really cool application of this theorem!!)

For any $E \in \mathcal{B}_{\mathbb{R}}$ define $m_s(E) = m(E + s)$ and $m^r(E) = M(rE)$. Then m_s and m^r clearly agree with m and $|r|m$ on finite unions of intervals, and so by theorem 1.3.2 they agree on all of $\mathcal{B}_{\mathbb{R}}$. As a consequence, if $m(E) = 0$, then $m(E + s) = m(rE) = 0$. Thus, for all sets in $\mathcal{L}$ (i.e. all unions of borel sets and Lebesgue zero-sets) must be preserved and linearly scaled (respectively) under translation and dilation:

$$m(E + s) = m(E) \qquad m(rE) = |r|m(E)$$

as we sought to show

Since we are working over $\mathcal{M}(\mathcal{B}_{\mathbb{R}})$, the next interesting questions we may ask about m is how it interacts with the topological sets on $\mathbb{R}$. For starters, it is clear (or should be immediately proven) that any singleton has measure 0. Therefore, by countable additivity, every countable set has measure zero; in particular $m(\mathbb{Q}) = 0$. Notice that the rational are dense which is a "topologically large" concept, however it is measure-theoretically small. In fact, we can even have a collection of *open dense* sets, and still be measure-theoretically small. Take some enumeration $\{r_i\}_{i=1}^{\infty}$ of the rational numbers in $[0, 1]$ and let I be the indexing set for this enumeration $i \in I$. Let I_n be the interval of size $\frac{\epsilon}{2^n}$ which is centered at r_n. Then set $U = \bigcup_n I_n$. By construction, U must be dense and open in $(0, 1)$, however

$$\mu(U) \leq \sum_{i=1}^{\infty} \mu(I_n) = \sum_{n=1}^{\infty} \frac{\epsilon}{2^n} = \epsilon$$

and hence can be as measure-theoretically small as we want. From this we can also find the opposite, where we have something topologically small but measure-theoretically big, by taking $K = [0, 1] - U$. By monotonicity

$$\mu(K) \geq 1 - \epsilon$$

and yet K is nowhere dense.

There is at least one intuition that can remain: a nonempty *open* set will have non-zero measure. For more on how open/closed sets interact with the measure, see the exercises.

Note that the converse that a non-zero measure set will have an open intervals, is not actually true! For that, we can explore the concept of *cantor sets*

Cantor Set

(was gonna look through: (commented so that this complies

Definition 1.4.1: Classical Cantor Set

Let $I = [0, 1]$. From this, we'll construct the cantor set iteratively

1. Start by removing the middle third $I_1 = I - \left(\frac{1}{3}, \frac{2}{3}\right)$
2. Given I_i, remove the open middle third of each connected compoennt

Let $C = \cap_{n=0}^{\infty} I_n$ be the Cantor set. In particular, it is the *standard middle thirds Cantor set*. We'll usually refer to this set as "the" Cantor set.

Notice that C is measurable by continuity from above since $C_1 \subseteq C_2 \subseteq \cdots \subseteq C_n \subseteq$.

Proposition 1.4.4: Properties Of Cantor Sets

Give C the standard subspace metric of $[0, 1]$. Then C is:

1. nonempty and uncountable
2. compact
3. nowhere dense $(\text{int}(\overline{A}) = \emptyset$ for all $A \subseteq C$
4. $m(C) = 0$
5. totally disconnected

Proof **:**
Using some topological properties, since C is the intersection of compact spaces, it is compact.

Next, the cantor set is nonempty since $0 \in C_k$ for all k, and so $0 \in C$. To show C is uncountable, Notice that we can represent every element of the uncountable set $[0, 1]$ in it's base 3 expansion, namely for appropriate $a_k \in \{0, 1, 2\}$:

$$x = \sum_{k=1}^{\infty} a_k 3^{-k}$$

Then take:

$$f : [0, 1] \to C \qquad x = \sum_{k=1}^{\infty} a_k 3^{-k} \mapsto$$

To show total disconnectedness, let's take the same I as before. I is closed in $\mathbb{R}$, and thus in C^n. Note that I^c is the union of finite closed sets, and thus is also closed. Thus, I is clopen in C^n. Thus, $C \cap I$ is also a clopen neigborhood of x in C. Since x was arbitrary, C is totally disconnected.

Finally uncountablenss come from C being compact and complete, which makes C uncountable (exercise).

This construction does not rely on the size of the middle-third; in fact, we can make it converge to any value we want. To see this, first notice that To do this, we first prove a lemma:

Lemma 1.4.2

Suppose $\{\alpha_i\}_{i=1}^{\infty} \subseteq (0,1)$.

1. $\prod_1^\infty (1-\alpha_i) > 0$ if and only if $\sum_1^\infty \alpha_i < \infty$ (Compare $\sum_1^\infty \log(1-\alpha_i)$ to $\sum_1^\infty \alpha_i$)
2. Given $\beta \in (0,1)$, there always exists a sequence $\{\alpha_i\}_{i=1}^{\infty}$ such that $\prod_1^\infty (1-\alpha_i) = \beta$

Proof **:**
First, we see that we can convert $\prod_i^\infty (1-\alpha_i)$ to $\sum_i^\infty \log(1-\alpha_j)$ by considering

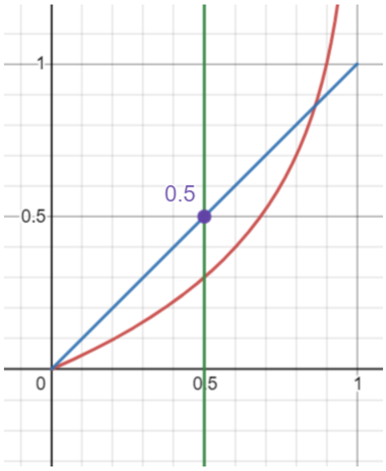
$$\begin{aligned}
\log\left(\prod_i^\infty (1-\alpha_i)\right) &= \log\left(\lim_{n\to\infty}\prod_i^n (1-\alpha_i)\right) \\
&\overset{!}{=} \lim_{n\to\infty} \log(\prod_i^n (1-\alpha_i)) \\
&= \lim_{n\to\infty} \sum_i^n \log(1-\alpha_i) \\
&= \sum_{i=1}^\infty \log(1-\alpha_i)
\end{aligned}$$

where $\overset{!}{=}$ comes from continuity of log for $1-\alpha_i > 0$. Notice that it's impossible that $1-\alpha_i \le 0$ form some α_i terms since $\alpha_i \in (0,1)$, and so this proof works. Thus, we get the equivalent condition of

$$\sum_i^\infty \log(1-\alpha_i) > -\infty$$

With this setup, we proceed with the proof

($\Leftarrow$) Let $\sum_i^\infty \alpha_i < \infty$. Then we know that $\alpha_i \to 0$. Thus, pick an N such that for all $i > N$, $\alpha_i < 0.5$. Then notice that $\log(1-\alpha_i) < \alpha_i$. This is easier to see with a visual aid:



In particular, for all $0 < x < 1/2$, we have $|\log(1-x)| - x > 0$, showing that x is greater than $\log(1-x)$ for all $x \in (0, 1/2)$.

Then from here, this is a simple application of the Weiestrass M-test. Let $K = \sum_i^N \log(1-\alpha_i)$ to eliminate the irregular part. Then since $|\log(1-\alpha_i)| < \alpha_i$, for all $i > N$ and $\sum_i^\infty \alpha_i$ converges, by the Weiestrass M-test the $\sum_{i=N+1}^\infty \log(1-\alpha_i)$ converges. Let's say it converges to L. Then

$$\sum_{i=1}^{\infty} \log(1-\alpha_i) = K + L > -\infty$$

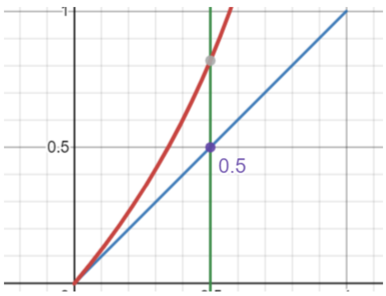
as we sought to show.

($\Rightarrow$) let $\prod(1-\alpha_i) > 0$ so equivalently $\sum_i^\infty \log(1-\alpha_i) > -\infty$ so $\sum_i^\infty \log(1-\alpha_i) = K$ for appropriate $K \in \mathbb{R}$. Then since the series converges, we have that

$$eK = e\sum_i^\infty \log(1-\alpha_i) = \sum_i^\infty e\log(1-\alpha_i)$$

i.e., we can scale the terms by e. We will show why this scaling is important in a moment; before we do we must setup a little more.

Since the series converges, it must be that $e\log(1-\alpha_i) \to 0$. So pick N such that for all $i > N$, $e\log(1-\alpha_i) < 1/2$. Then notice that $\alpha_i < e\log(1-\alpha_i)$. This is easier to see with a visual aid:



Or, to carry out some of the computations, if you raise $e\log(1-\alpha_i) = \log(1-\alpha_i)e$ to the power of e, you get

$$(e^{\log(1-\alpha_i)})^e = (1-\alpha_i)^e \geq \alpha_i^e \quad \forall \alpha_i \in (0, 1/2]$$

and so the inequality holds. Then the proof continues as for the $\Leftarrow$ direction by setting up an appropriate constant term to bound the irregularities, then using the Weiestrass M test bound every α_i by $\log(1-\alpha_i)$, and since $\sum_i^\infty \log(1-\alpha_i)$ converges, we get that $\sum_{i=N}^\infty \alpha_i$ also converge, and since it will be the sum of a constant plus the convergent value, $\sum_{i=1}^\infty \alpha_i$ converges, as we sought to show.

Now, let β be given. Let $x_i = \prod_{i=1}^n (1-\alpha_i)$. We'll construct a sequence where

$$x_1 = \beta + \frac{1-\beta}{2} \qquad x_2 = \beta + \frac{1-\beta}{4} \qquad \cdots \qquad x_n = \beta + \frac{1-\beta}{2^n} \cdots$$

to accomplish this, let

$$\alpha_i = 1 - \frac{\beta + \frac{1-\beta}{2^i}}{\beta + \frac{1-\beta}{2^{i-1}}} \tag{1.2}$$

This construction comes from solving the following equation:

$$\left(\beta + \frac{1-\beta}{2}\right) y_2 = \beta + \frac{1-\beta}{4} \;\Rightarrow\; y_2 = \frac{\beta + \frac{1-\beta}{4}}{\beta + \frac{1-\beta}{2}}$$

so $\alpha_2 = 1 - y_2$, and clearly $\alpha_2 \in (0,1)$ since the denominator is bigger than the numerator in the above calculations. Generalizing this pattern, we get the that α_i would be of the from presented in equation (1.2) and $\alpha_i \in (0,1)$. Thus, we get that

$$\lim_{n\to\infty} \prod_i^n (1-\alpha_i) = \lim_{n\to\infty} \beta + \frac{1-\beta}{2^n} = \beta$$

as we sought to show

Now, let's say we are starting with $[0,1]$ but on each iteration eliminates intervals that sum to α_i. Let K_i be the ith step in this itteration. Notice that:

$$m(K_{i+1}) = (1-\alpha_i) m(K_i)$$

Then by what we've just shown, we can make $K = \cap_i^\infty K_i$ be any value up to and including 0 and 1.

Proposition 1.4.5: All Cantor Set Homeomorphism

Let C_n be a cantor set where $m(C_n) = n$. Then $C_n \cong C_0$

***Proof* :**
section 9 of Pugh – will get back to it

You can also prove uncountability directly by making a string represent which side of the split interval you go to.

Next, you might've seen many uncountable sets which are dense in $\mathbb{R}$. Here we have an absolutely facsinating result: C is *nowhere dense in* $[0,1]$.

Definition 1.4.2: Dense, Somwehre Dense, Nowhere Dense

If $S \subseteq M$ and $\overline{S} = M$, then S is dense in M. The set S is somehwere dense if there exists an open nonemtpy set $U \subseteq M$ such that $\overline{S \cap U} \supset U$. If S is not somwerhe dense, then it is *nowhere dense*

Proposition 1.4.6: C Nowhere Dense

The Cantor set contains no interval and is nowehre dense in $\mathbb{R}$

Proof **:**
To prove it doens't contian an interval, assume it does, then chose an n alrge enough, that is $(\frac{1}{3})^n < b - a$, so that it will not be in the interval.

For anowhere dense, assuem ti's somwerhe then, then $\overline{C \cap U} \supset U \supset (a,b)$ – a contradiction.

We know present how Cantor sets are almost initial objects in compact metric space

Theorem 1.4.4: Cantor Surjection Theorem

Given a compact nonempty metric space M. there is a continuous surjection of C onto M

Proof **:**
exercise

Finally, we present how all cantor sets are homeomorphic.

Theorem 1.4.5: Moore-Kline Theorem

We say M is a Cantor space if it is compact, nonempty, perfect, and totally disconnected (like the cantor set). Every cantor space M is homeomorphic to the standard middle-thirds Cantor set C.
In other words, those properties *define* Cantor sets.

Proof **:**
Pugh. p.112

Corollary 1.4.1: Cantor And Fat Cantor Set

The fact Cantor set is homeomorphic to the standard Cantor set

***Proof* :**
Using 1.4.5, the result is immediate.

This is intersting, since it shows that meausre is *not* a topological poperty! This should makes sense with $(0, 1)$ and $\mathbb{R}$ being homeomorphic with the same measures, but different, but this is an even more extreme example.

Corollary 1.4.2: Product Of Cantor Sets

A cantor set is homeomorhic to its own cartesian product; $C \equiv C \times C$.

***Proof* :**

A fact that Folland added that I thought was cool is that since every subset of the cantor set is Lebesgue Measurable (since the cantor set has measure zero and the Lebesgue measure is complete), we have that

$$|\mathcal{L}| = |P(\mathbb{R})| > \aleph_1 \quad \text{but} \quad |\mathcal{B}_{\mathbb{R}}| = \aleph_1$$

so there are a *tone* of zero-measure sets, however, we can essentially ignore them because they are zero measure sets!!

For those who are Categorically inclined, Cantor sets are almost like the initial objects of the category of compact metric spaces. They are not initial because there does not exist a unique uniformally continuous function from the cantor set to any compact metric space, but there always will exist at least 1 such function.

Exercise 1

1. If $E \in \mathcal{L}$ and $m(E) > 0$, for any $\alpha < 1$, there is an open interval I such that $m(E \cap I) > \alpha m(I)$
2. If $E \in \mathcal{L}$, and $m(E) > 0$, the set $E - E = \{x - y \mid x, y \in E\}$ contains an interval centered at 0 (If I is as in the previous exercise with $\alpha > \frac{3}{4}$, then $E - E$ contains $(-\frac{1}{2}m(I), \frac{1}{2}m(I))$).
3. Let E be a Lebesgue measurable set (so $E \in \mathcal{L} = \overline{\mathcal{M}}(\mathcal{B}_{\mathbb{R}})$)
    1. If $E \subseteq N$ where N is a non-measurable set described in section 1.1, then $m(E) = 0$
    2. if $m(E) > 0$, then E contains a non-measurable set (it suffices to assume $E \subseteq [0, 1]$. In notation of section 1.1 $E = \cup_{r \in R} E \cap N_r$

Chapter 2

Integration

In this chapter, we now work with function's between measure spaces that preserve measure structure! Measurable functions feel like the types of functions we "usually" work with, the same way measurable sets are the sets we "usually" work with. Being continuous can be a bit strong (ex. being a monotone function doesn't imply continuous), but measureability rectifies that by in a sense expanding what are the possible "pre-images" we allow to take (ex. Monotone functions *are* always measurable, as we will demonstrate).

After defining measurable functions, we will focus on real and complex functions (i.e. functions with codomain $\mathbb{R}$ or $\mathbb{C}$). In the real case, I like to think of real functions as given every point in the domain X a value or a weight, in contrast to a measure which is a real function which simply assigns a constant value at each point (this will become more clear as we introduce the definitions). For the complex case, we simply split the problem into the real and imaginary component[1]. Then, we will define how to "measure" a real (resp. complex) measurable function, which can be interpreted physically as a way capture the amount of total "value" (or "weighted size"') that has occured. We then focus on measurable functions which have "finite measure", and will call them *integrable functions*. Similarly to how uncountable infinities has caused us problems with measure, integrable functions will allow us to have some regularity in how we can combine measurable functions without running into the problems of infinities (as we'll soon show). One result of this is that we can think of the integral as being a "measure" for a function, and since integrable function will have "finite measure", it will make it more meaningful to see which functions are "close" (in the appropraite sense of the word close) to a certain function.

After we have defined integrable functions, we will have to take a moment to compare different types of convergences, since at that stage we will have 4 different ways functions can converge. We will then move on to trying to bring the theory of the product of σ-algebras to measure and integrable functions more generally, allowing us to construct larger measures and integrable functions from smaller components. Importantly, we will show that we can find the a measure $\mu \times \nu$ by sepearting any problem to only measuring "slices" of the set with respect to μ, then "slices" of the set with respect to ν (this is Fubini's Theorem).

[1] I'm not yet sure why we would consider the complex case

2.1 Measurable Functions

Recall that if $f : X \to Y$ is a function, then $f^{-1} : P(Y) \to P(X)$ is a function where $f^{-1}(E) = \{x \in X \mid f(x) \in E\}$ which preserves unions, intersections, and compliments. Thus, if $\mathcal{N}$ is a σ-algebra on Y, $f^{-1}(\mathcal{N})$ is a σ-algebra on X. If X and Y are measure space, if $f^{-1}(\mathcal{N})$ maps into $\mathcal{M}$, we will call f *measurable*:

Definition 2.1.1: Measurable Function

Let $f : X \to Y$ be a function and $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ be measurable spaces. Then f is called a $(\mathcal{M}, \mathcal{N})$-*measurable function* (or just measurable) if for all $E \in \mathcal{N}$

$$f^{-1}(E) \in \mathcal{M}$$

Just like in topology, this could be thought of as a surjection of the measurable space on X onto the measurable space on Y. It should be clear that the composition of two measurable functions is measurable (in particular, the composition of an $(\mathcal{M}, \mathcal{N})$-measurable function with $(\mathcal{N}, \mathcal{O})$-measurable function is $(\mathcal{M}, \mathcal{O})$-measurable), and that the identity function $\mathrm{id} : X \to X$ is clearly a measurable function, and so the collection of measurable functions and measurable spaces from a category, usually denoted **Meas**.

Just like lemma 1.1.1 was useful for simplifying proofs, we proof the following:

Lemma 2.1.1: Measurable Function and Generating Sets

Let $\mathcal{N}$ be generated by $\mathcal{E}$. Then $f : X \to Y$ is a $(\mathcal{M}, \mathcal{N})$-measurable function if and only if for all $E \in \mathcal{E}$, $f^{-1}(E) \in \mathcal{M}$

Proof **:**
The $\Rightarrow$ direction is *a-fortiori* true by definition. The $\Leftarrow$ direction comes from f^{-1} preserving unions and compliments, so $\left\{E \subseteq Y \;\middle|\; f^{-1}(E) \in \mathcal{M}\right\}$ is a σ-algebra that contains $\mathcal{E}$, and so contains $\mathcal{N}$.

From this, there is a very nice corrollary when $\mathcal{M}$ is a collection of Borel sets:

Corollary 2.1.1: Continuous Function $\Rightarrow$ [Borel] Measurable Function

Let $f : X \to Y$ be a continuous function and Y have the borel σ-algebar. Then f is a measurable function.

Proof **:**
Since f is continuous, the pre-image of an open set (in particular, open intervals) is open, and hence by proposition 1.1.2 and lemma 2.1.1 it is measurable

Naturally, in analysis, the most important measurable functions will be of the form

$$f : \mathbb{R} \to \mathbb{R} \qquad f : \mathbb{R} \to \mathbb{C}$$

This also means that the domain can either have the $\mathcal{B}_{\mathbb{R}}$ σ-algebra or $\mathcal{L}$ σ-algebra.

Example 2.1: Measurable Functions
Most functions you have worked with will in fact be measurable

1. As already shown, all continuous functions are measurable
2. All monotone functions are measurable: the pre-image of closed reays will either be open or closed rays, and so by proposition 1.1.2, they are measurable.
3. All indicator functions over measurable sets are measurable (hence, bump functions are essentially measurable)
4. All Riemann integrable functions are measruable (we will show this this in section 2.3.1)
5. We will soon show that the point-wise limit of measurable functions is also measurable, hence

It is in practice rare to work with non-measurable functions unless a contrived example is used, like a measurable function purposefully built to map to non-measurable sets (like in the case of the composition of two Lebesgue measurable functions)

Convention We will *always* assume the codomain will have $\mathcal{B}_{\mathbb{R}}$ (resp. $\mathcal{B}_{\mathbb{C}}$) σ-algebara. In fact, if we have a function $f : X \to \mathbb{R}$ (resp. $f : X \to \mathbb{C}$), then we will sometimes say f is $\mathcal{M}$-measurable instead of $(\mathcal{M}, \mathcal{B}_{\mathbb{R}})$-measurable (resp. $(\mathcal{M}, \mathcal{B}_{\mathbb{R}})$-measurable). If the domain is also Borel, we will say that f is *Borel measurable* (and ommit the borel measure if it's $\mathbb{R}$ or $\mathbb{C}$). Similarly, if the domain is $\mathcal{L}$, we will call f Lebesgue measruable.

Becasue of the convention of the co-domain being Borel, it is possible that the composition of two Lebesgue measurable functions is *not* Lebesgue measurable: if $f : \mathbb{R} \to \mathbb{R}$, $g : \mathbb{R} \to \mathbb{R}$ are two Lebesgue measruable functions where $E \in \mathcal{B}_{\mathbb{R}}$, $g^{-1}(E) \in \mathcal{L}$ but $g^{-1}(E) \notin \mathcal{B}_{\mathbb{R}}$, it is possible that $f^{-1}(g^{-1}(E)) \notin \mathcal{L}$

Example 2.2: Composition Not Lebesgue Measurable
Let $f : [0, 1] \to [0, 1]$ be the cantor function form section 1.5, and let $g(x) = f(x) + x$.

1. g is a bijection from $[0, 1]$ to $[0, 2]$ and $h = g^{-1}$ is continuous from $[0, 2]$ to $[0, 1]$

   Proof **:**
   First, since f is increasing (though not injective) and x is strictly increasing, then $f + x$ is injective. Since f and x are both continuous, so is $g = f + x$. Hence the pre-image of a closed set must be closed. If we consider $g^{-1}([0, 2]) = D$, then using the fact that g is a monotonic functions, since $g(0) = 0$ and $g(1) = 2$, it must be that $D = [0, 1]$, showing that g is indeed surjective, and hence bijective.

   To show that $g^{-1} : [0, 2] \to [0, 1]$, is continuous, bijective, and the codomain is Hausdorff, then g is homeomorphic (recall that these imply that g is an open map: if $U \subseteq [0, 1]$ then $S = [0, 1] - U$ is closed, and since g is continuous $g([0, 1] - U) = [0, 2] - g(U)$ is closed, so $g(U)^c$ is closed, so $g(U)$ is open), and hence g^{-1} is continuous.

2. If C is the cantor set, then $m(g(C)) = 1$

***Proof* :**
We will take advantage of translation invariance of the Lebesgue measure. Let C be the cantor set on $[0, 1]$. Then C can be thought of as the interval $[0, 1]$ minus all of the open sets that are eliminated during construction:

$$C = [0, 1] - I_1 - I_{21} - I_{22} - \cdots - I_{ik} - \cdots$$

Let $I = \bigsqcup_{i,k} I_{ik}$ collection of these sets so that $C = [0, 1] - I$. Since $m(C) = 0$, $m(I) = 1$. Consider now $g(I)$. Since g is injective and I is the disjoint union of sets, then there exists sets $\{J_j\}_{j=1}^{\infty}$ such that $J_j = g(I_{ik})$. We now examine the measure of the sets J_j.

By definition, I_{ik} does not contain any points of the cantor set, and so for appropriate $c_{ik} \in C$,

$$g_{I_{ik}}(x) = f(c_{ik}) + x$$

that is, g on the interval I_{ik} is of the form $x + f(c_{ik})$ for constant $f(c_{ik})$! Since the Lebesgue measure is translation invariant, we get:

$$m(g(I_{ik})) = m(g(I_{ik}) - f(c))$$

and so, since the I_{ik} intervals are all disjoint and g is injective:

$$g(\bigsqcup I_{ik}) = \bigsqcup g(I_{ik})$$

and so the measure of $g(I)$ is in fact exactly the measure of I, that is, $m(g(I)) = m(J) = 1$. Thus:

$$2 = g(m([0, 2])) = m(g(C) \sqcup J) = m(C) + m(J) = m(g(C)) + 1$$

and so $m(g(C)) = 1$, as we sought to show.

3. by exercise 29, $g(C)$ contains a Lebesgue non-measurable set A. Let $B = g^{-1}(A)$. Then B is Lebesgue measurable, but not Borel

***Proof* :**
Since $A \subseteq g(C)$, $B \subseteq g^{-1}(g(C)) = C$. Since $m(C)$ is zero, by completeness, $m(A) = 0$, and so A is Lebesgue measurable.

Next, to show that B is not Borel-measurable, For the sake of contradiction let's assume it was. Then since g^{-1} is continuous, it is Borel-measurable, meaning the pre-image (with respect to g^{-1}) of a measurable set is measurable. However, $g(B) = A$, which is not measurable, and hence B is not Borel measurable.

4. There exists a Lebesgue measurable function F and a continuous function G on $\mathbb{R}$ such that $F \circ G$ is not Lebesgue measurable

Proof **:**
Let F be the indicator function on B, I_B, and $G = g^{-1}$. Then F and G are clearly Lebesgue measurable, However, G^{-1} does not map all Lebesgue measurable sets to Lebesgue measurable sets (in particular, B maps to a Borel set), and so

$$(F \circ G)^{-1}(B) = (G^{-1} \circ F^{-1})(B) = (g(I_B^{-1}(B)) = g(B) = A$$

However, A is not Lebesgue measurable set, and so $F \circ G$ is not Lebesgue measurable.

Just like for Borel sets earlier, we have the following equivalence for measurable functions with codomain $\mathbb{R}$:

Proposition 2.1.1: Borel sets and Measurable Functions

Let $(X, \mathcal{M})$ be a measurable space and $f : X \to \mathbb{R}$. Then the following are equivalent:

1. f is $\mathcal{M}$-measurable
2. $f^{-1}((a, \infty)) \in \mathcal{M}$ for all $a \in \mathbb{R}$
3. $f^{-1}([a, \infty)) \in \mathcal{M}$ for all $a \in \mathbb{R}$
4. $f^{-1}((-\infty, a)) \in \mathcal{M}$ for all $a \in \mathbb{R}$
5. $f^{-1}((-\infty, a]) \in \mathcal{M}$ for all $a \in \mathbb{R}$

Proof **:**
This follows from proposition 1.1.2 and lemma 2.1.1.

Similarly, we will soon be encountering co-domains which are product spaces (ex. $\mathbb{R}^n$ or $\mathbb{C}^n$). In particular, let's say we had some set X, and some family of measurable spaces $\{(Y_\alpha, \mathcal{M}_\alpha)\}_{\alpha \in A}$, and $f : X \to Y_\alpha$ is a map for each $\alpha \in A$. Then we can impose a unique smallest σ-algebra on X that makes each f_α a measurable function, in particular, $\mathcal{M}_X = \left\{ f_\alpha^{-1}(E_\alpha) \;\middle|\; E_\alpha \in \mathcal{M}_\alpha,\ \alpha \in A \right\}$ [2]. This σ-algebra on X is called the *σ-algebra generated by* $\{f_\alpha\}_{\alpha \in A}$.

Proposition 2.1.2: Product of Measurable Functions

Let $(X, \mathcal{M})$ be a measurable space and $(Y_\alpha, \mathcal{N}_\alpha)$ be a collection of measurable spaces, so $Y = \prod_{\alpha \in A} Y_\alpha$, $\mathcal{N} = \bigotimes_{\alpha \in A} \mathcal{N}_\alpha$, and $\pi_\alpha : Y \to Y_\alpha$ is the coordinate map. Then $f : X \to Y$ is $(\mathcal{M}, \mathcal{N})$-measurable if and only if $f_\alpha = \pi_\alpha \circ f$ is $(\mathcal{M}, \mathcal{N}_\alpha)$-measurable.

Proof **:**
Let's say f is $(\mathcal{M}, \mathcal{N})$-measurable. Then since the composition of two measurable functions is measurable, so are each f_α is measurable (since π_α is measurable by definition of $\bigotimes$).

Conversely, if every f_α is $(\mathcal{M}, \mathcal{N}_\alpha)$-measurable, then for each $E_\alpha \in \mathcal{N}_\alpha$, $f^{-1}(\pi_\alpha^{-1}(E_\alpha)) = f^{-1}(E_\alpha) \in$

[2] A similar strategy is used in the constrution of *weak* topologies. If you're more categorical, this constructions makes X satisfy the universal property of products in the category of **Meas**

$\mathcal{M}$, and since these sets generate $\mathcal{M}$, we have that f is $(\mathcal{M}, \mathcal{N})$-measurable by proposition 2.1.1

This proposition also let's us define what it means to take the "product" of two functions. As a consequence, we can give an easy criterion for when complex functions are measurable:

Corollary 2.1.2: Complex Function Measureability

Let $f : X \to \mathbb{C}$ be a complex function. Then f is $\mathcal{M}$-measurable if and only if $\Re f$ and $\Im f$ are both measurable.

Proof **:**
This is simply taking advantage of the topology of $\mathbb{C}$, namely

$$\mathcal{B}_{\mathbb{C}} = \mathcal{B}_{\mathbb{R}^2} \overset{!}{=} \mathcal{B}_{\mathbb{R}} \otimes \mathcal{B}_{\mathbb{R}}$$

where the $\overset{!}{=}$ equality comes from the fact that $\mathbb{R}$ is separable (see proposition 1.1.5). And so proposition 2.1.2 with the fact that $\Re$ and $\Im$ can be considered projections completes the proof.

Sometimes, we want to work with the extended real line (or in general some compactification of a topological space). If $\overline{\mathbb{R}} = [-\infty, \infty]$, Then let

$$\mathcal{B}_{\overline{\mathbb{R}}} = \left\{ E \subseteq \overline{\mathbb{R}} \;\middle|\; E \cap \mathbb{R} \in \mathcal{B}_{\mathbb{R}} \right\}$$

Notice that is we put the metric ρ on $\overline{\mathbb{R}}$ to be $\rho(x, y) = |\tan^{-1}(x) - \tan^{-1}(y)|$, then $\mathcal{B}_{\overline{\mathbb{R}}}$ is indeed the usual definition of a Borel σ-algebra (exercise). It should be verified that $\mathcal{B}_{\overline{\mathbb{R}}}$ can be generated by open or closed rays, and so we will say $f : X \to \overline{\mathbb{R}}$ is $\mathcal{M}$-measurable if it is $(\mathcal{M}, \mathcal{B}_{\overline{\mathbb{R}}})$-measurable.

Next, if the codomain has some notion of $+$ and $\cdot$, we establish that the basic "$\mathbb{C}$-algebra operations" work on measurable functions:

Proposition 2.1.3: Arithmetics on Measurable Functions

Let $f : X \to \mathbb{C}$ and $g : X \to \mathbb{C}$ be $\mathcal{M}$-measurable functions. Then $f + g$, fg, and zf for $z \in \mathbb{C}$ are $\mathcal{M}$-measurable functions

Note that we have limited ourselves to $\mathbb{C}$. This still works with $\overline{\mathbb{R}}$ with the appropriate care of the $\infty - \infty$ and $0 \cdot \infty$ case (exercise 2 in book)

Proof **:**
The proof for zf being measurable is similar to that of $+$ and $\cdot$, so is left as an exercise.

Since f and g, are measurable, then their "cartesian product" is measurable by proposition 2.1.2

$$F : X \to \mathbb{C} \times \mathbb{C} \qquad F(x) = (f(x), g(x))$$

In particular, F is $(\mathcal{M}, \mathcal{B}_{\mathbb{C}} \otimes \mathcal{B}_{\mathbb{C}})$-measurable. Next, consider $p : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$, $p(x, y) = x + y$ and $t : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$, $t(x, y) = xy$ (p for plus, t for times). Then since p and t are continuous, by corrollary 2.1.1 p and t are $(\mathcal{B}_{\mathbb{C}\times\mathbb{C}}, \mathcal{B}_{\mathbb{C}})$-measurable. Thus, $p \circ F$ and $t \circ F$ is $(\mathcal{M}, \mathcal{B}_{\mathbb{C}})$-measurable, and by definition

$$f + g = p \circ F \qquad fg = t \circ F$$

completing the proof

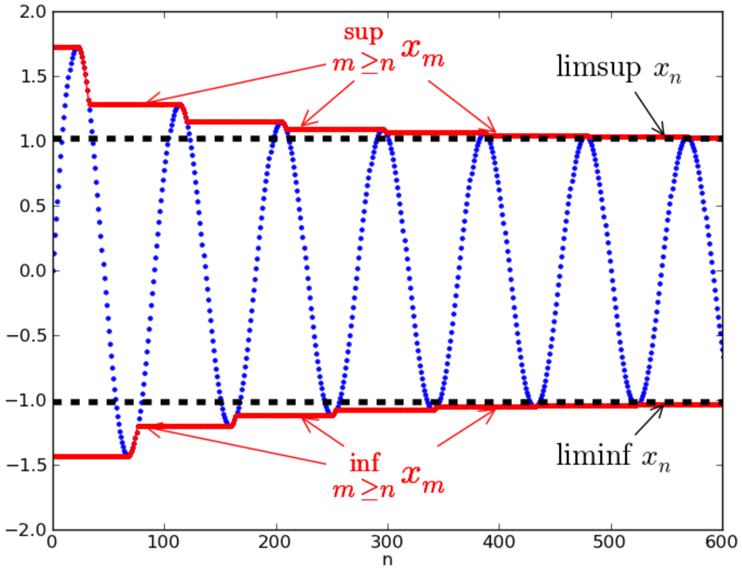
the next operations we'll show are measruable are sup, inf, lim sup and lim inf. This is perhaps one of the most important functions to be measurable since they are at the heart of many constructions in analysis. Recall that

$$\sup_i f_i(x) = \sup \{f_i(x) \mid i \in \mathbb{N}\}$$

and so this expression means we're taking the set $f_i(x)$ for each i and taking the sup of that set. In other words, for every point x we're taking the supremum of the set $\{f_1(x), f_2(x), ...\}$. Symmetrically the same for inf. For lim sup, recall that:

$$\limsup_i f_i(x) = \lim_{i\to\infty} \sup \{f_k(x) \mid k \geq i\}$$

that is, we are constantly shrinking the set over which we are doing the supremum, which in a sense means we're trying to "hug" the value from above as close as possible in the limit. Symmetrically the same for lim inf. If you're visual, I like to keep this image in mind: let's say the blue line is the value of $f_i(x_0)$ for a specified x_0. Then:



Comment here for future reference:

Proposition 2.1.4: Measurable functions and Limit Bounding

Let $f_i : X \to \overline{\mathbb{R}}$ be $\mathcal{M}$-measurable functions for each i and consider $\{f_i\}_{i=1}^{\infty}$. Then

$$g_1(x) = \sup_i f_i(x) \qquad g_2(x) = \inf_i f_i(x)$$

$$g_3(x) = \limsup_{i\to\infty} f_i(x) \qquad g_4(x) = \liminf_{i\to\infty} f_i(x)$$

The codomain $\overline{\mathbb{R}}$ was used so that the sup and inf is still considered defined if $\sup = \pm\infty$ or $\inf = \pm\infty$.

***Proof* :**

To show $\sup_i f_i$ and $\inf_i f_i$ are measurable, since the codomain domain is $\overline{\mathbb{R}}$, which is equivalent measure-theoretically to $\mathbb{R}$, we will show the pre-image of open rays is the union of measurable sets:

$$\begin{aligned} x \in g_1^{-1}((a,\infty]) &\Leftrightarrow g_1(x) > a \\ &\Leftrightarrow f_i(x) > a && \text{for some } i \\ &\Leftrightarrow x \in f_i^{-1}((a,\infty]) && \text{for some } i \\ &\Leftrightarrow x \in \bigcup_{i=1}^{n} f_i^{-1}((a,\infty]) \end{aligned}$$

and similarly for g_2, except considering $(-\infty, a]$, therefore:

$$g_1^{-1}((a,\infty]) = \bigcup_i^{\infty} f_i^{-1}((a,\infty]) \qquad g_2^{-1}(-\infty, a])) = \cup_i^{\infty} f_i^{-1}([-\infty, a))$$

since each f_i is measurable, we have a union of measurable sets, so g_1 and g_2 are measurable by proposition 1.1.5. Next, let

$$h_k(x) = \sup_{i>k} f_i(x) = \sup\{f_k(x) \mid k > i\}$$

then by what we've just established, h_k is measurable for each k. Then by of $\limsup$:

$$g_3 = \inf_k h_k$$

and so g_3 is measurable since we just established the inf of measurable functions is measurable. Similarly for g_4.

as an immediate corrollary:

Corollary 2.1.3: Measurability of Max and Min

Let f, g be measurable functions. Then so is $\max\{f, g\}$ and $\min\{f, g\}$

Corollary 2.1.4: Measurable Preserved under Pointwise limit

Let $\{f_i\}$ be a sequence of measurable function. If $\lim_{i\to\infty} f_i(x)$ exists for every x, and we define $f : X \to \overline{\mathbb{R}}$ to be

$$f(x) = \lim_{i\to\infty} f_i(x)$$

then f is measurable (i.e. point-wise convergence preserves measureability)

Recall that point-wise convergence *does not preserve continuity* (the steeple function's are the usual example). However, measureability is sufficiently general to be preserved under this weaker form of convergence! Naturally, if the sequence of functions $\{f_i\}$ are all $\mathcal{M}$-measurable and uniformally convergent, they are measurable

***Proof* :**
If f exists (i.e., every $f_n(x)$ converges pointwise to $f(x)$), then by proposition 2.1.4, $f = g_3 = g_4$, completing the proof.

Remark Note how important it is that *all* points converge! It is possible that a single point doesn't converge (ex. $\lim f_i(0)$ doesn't converge), and so the limit is *not* a measurable function, in fact it doesn't even need to be a function. Take for example the collection of functions $f_k : \mathbb{R} \to \mathbb{R}$ with the properties:

1. f_k is continuous (it can even be smooth)
2. $\int_{\mathbb{R}} f(x)dx = 1$ (In the Riemann sense), or really any constant $c \in \mathbb{R}_{>0}$
3. $\operatorname{supp} f \subseteq \left[\frac{-1}{k}, \frac{1}{k}\right]$

In other words, the collection of f_k are all smooth bump functions with volume 1, but the area in which it has support is shrinking, and so the "bump" of each f_k must be getting higher and higher (to define such functions, consider starting with $f_1 = e^{-\frac{1}{1-x^2}}$ where $x \in (-1, 1)$ and 0 otherwise). Then $f_k \to 0$ at all but 1 point, namely $f_k(0)$ does *not* converge. As a consequecne, this limit is *not* a function!

This example shows the important of somehow "bounding" your area; we will get back to this observation in section 2.2. Some might think that this example is a bit cheeky, it still feels like the function is measurable if we just circumnavigate this problem by making the domain of the function $\mathbb{R} \cup \{\infty\}$, there is the trouble of defining a measure on the 1-point compactification of $\mathbb{R}$ while extending the measure of $\mathbb{R}$[a]. If the reader wonders if there is a way of defining some generlization of a function that will accept the fact that the f_k converge to something that still somehow preserves the fact that $\int f = 1$, then we will show in in chapter 7 that this converges to something called a *distribution*[b]

[a] In particular, $\mu(\{\infty\}) = \infty$, as can be shown using continuity form above, and so somehow we have "gained" an infinite amount of area. On the other-hand, if we just ignore this and consider that ∞ is just some point like any other point, then we have "lost" area

[b] The example named here is a distribution called the *delta distribution*, or sometimes *delta function* for reasons that is covered in chapter 7

Now that we proved that these operations preserving measurability, we will introduce some new notation to decompose real and complex functions which will soon become useful when defining integrable functions.

Definition 2.1.2: Real and Complex Decomposition

Let $f : X \to \mathbb{R}$ and $g : X \to \mathbb{C}$ be real and complex measurable functions respectively. Then

$$f^+ = \max\{f, 0\} \qquad f^- = \max\{-f, 0\}$$

and

$$\operatorname{sgn}(g(x)) = \begin{cases} \frac{g(x)}{|g(x)|} & g(x) \neq 0 \\ 0 & g(x) = 0 \end{cases} \qquad |g(x)| = |z| = |x + iy| = \sqrt{x^2 + y^2}$$

which gives the decomposition of:

$$f = f^+ - f^-, \;\; g = (\operatorname{sgn} g)|g|$$

The functions f^+ and f^- are clearly well-defined and measurable. For the complex case, $|\cdot|$ is continuous everywhere and $z \mapsto \operatorname{sgn} z$ is continuous everywhere except 0. It is measurable since if $U \subseteq \mathbb{C}$ is open, then $\operatorname{sgn}^{-1}(U)$ is either V for some open set V, or $V \cup \{0\}$, meaning sgn is Borel measurable. In some textbooks, arg is used instead of sgn.

2.1.1 Standard Representation

We now how that every measurable function is in fact simply the limit of a bunch of easy to work with function, in particular, linear combination of characteristic functions! As a reminder, if $\chi_E : X \to \mathbb{R}$ (or $\chi_E : X \to \{0, 1\}$) where $E \subseteq X$, then:

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

Then χ_E is called the characteristic (or indicator) function. It is sometimes also denoted as 1_E or I_E.

Definition 2.1.3: Simple function

Let $\{E_i\} \subseteq P(\mathbb{C})$ be some finite collection of sets. Then the finite linear combination of characteristic functions on E_i multiplied by a complex constant:

$$f = \sum_{i=1}^{n} z_i \chi_{E_i}$$

is called a *simple function*

Example 2.3: Simple Functions

1. Show that $|f(X)| < \infty$ then f is in fact a simple function (hint: let $E_i = z_i$ for each $z_i \in f(X)$). In fact, each coefficient can also be distinct (let it be the respective z_i from the range)
2. Show that if f and g are simple functions, so are $f + g$ and fg

Using these, we can now show that any measurable function is simply the limit of simple functions!

Theorem 2.1.1: Simple Function Approximation

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

1. If $f : X \to [0, \infty]$ is measurable, there exists a sequence of functions $\{\phi_i\}_{i=1}^{\infty}$ such that
$$0 \leq \phi_1 \leq \phi_2 \leq \cdots \leq f$$
where $\lim_{n\to\infty} \phi_i(x) = f(x)$ (i.e. $\phi_n \to f$ pointwise). and $\phi_n \rightrightarrows f$ on a bounded domain (i.e. uniformally converges)

2. If $f : X \to \mathbb{C}$ is measurable, there exists a sequence of functions $\{\phi_i\}_{i=1}^{\infty}$ such that
$$0 \leq |\phi_1| \leq |\phi_2| \leq \cdots \leq |f|$$
where $\lim_{n\to\infty} \phi_i(x) = f(x)$ (i.e. $\phi_n \to f$ pointwise). and $\phi_n \rightrightarrows f$ on a bounded domain (i.e. uniformally converges)

***Proof* :**
Essentially, you will get characteristic functions to converge to the value like so: In particular, for $n \in \mathbb{N}_{\geq 0}$ and $0 \leq k \leq 2^{2n} - 1$, let

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

and

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

then we get that $\phi_n \leq \phi_{n+1}$ for all n and $0 \leq f - \phi_n \leq \frac{1}{2^n}$ where $f \leq 2^n$. The rest is left to check.

Generalizing to $\mathbb{C}$, decomposing to $f = g + ih = (g^+ + g^-) = i(h^+ + h^-)$, where the $+$ and $-$ functions are the positive and negative parts of f and g, then we can repeat the process for each of these and get $\psi_n^+, \psi_n^-, \zeta_n^+, \zeta_n^-$ and let $\phi_n = (\psi_n^+ - \psi_n^-) + i(\zeta_n^+ - \zeta_n^-)$.

This shows that measurable functions are a very natural type of function to work with in analysis: they are all a pointwise-limit (and on a bounded domain a uniform limit) of simple functions. The particular sequence of simpel functions will be called the *standard representation of* f. Using this result, we can define a new definition of equality up to zero-measure sets:

Proposition 2.1.5: Measurability up to Zero Set

Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be two measure with ν being a complete measure, and let $f : X \to Y$ be $(\mathcal{M}, \mathcal{N})$-measurable. Then:

1. If $f = g$ up to a zero measure set with respect to μ (shorthanded to μ-a.e.), then g is measurable
2. If f_n is measurable for each $n \in \mathbb{N}$ and $f_n \to f$ μ-a.e, then f is measurable

Proof **:**

1. Let $N = \{x \in X \mid f(x) \neq g(x)\}$. Let $B \in \mathcal{N}$ and consider $g^{-1}(B)$. Then by definition:

$$\begin{aligned} g^{-1}(B) &= \{x \in M \mid g(x) \in B\} \\ &= \{x \in \mathcal{M} \mid g(x) \in B,\ g(x) = f(x)\} \cup \{x \in \mathcal{M} \mid g \in B,\ g(x) \neq f(x)\} \end{aligned}$$

   the latter set is a subset of N, and so is measurable by the completeness of f. The former set is of the form:

$$\begin{aligned} \{x \in \mathcal{M} \mid g(x) \in B,\ g(x) = f(x)\} &= \{x \in \mathcal{M} \mid f(x) \in B,\ g(x) = f(x)\} \\ &= f^{-1}(B) \cap \{x \mid f(x) = g(x)\} \\ &= f^{-1}(B) \cap (X - \{x \mid f(x) \neq g(x)\}) \\ &= f^{-1}(B) \cap (X - N) \end{aligned}$$

   and since f measurable the pre-image of a measurable set is too, and since N is measuable, $X - N$ is too, so we have that $g^{-1}(B)$ is measurable.

2. Let's say N is the set on which $f_n \not\to f$. Define

$$g_n = f_n \chi_{X-N} + f \chi_N$$

   Then notice that $g_n = f_n$ μ-a.e., and so by part 1 are measurable, and by construction $g_n \to f$ everywehre, and so f is measurable!

However, if μ is not complete, it turns out to make little difference since we can complete the measure, produce the result, and restrict back to the original measure without changing the result:

Proposition 2.1.6: Extending To Complete Measueres

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

Proof **:**
If f is a simple function, then we see that the completion will at most remove finitely many points when we restrict every χ_E $(E \in \overline{\mathcal{M}})$ to g, and so $f = g$ $\overline{\mu}$-a.e.

Now let f be a $\overline{\mathcal{M}}$-measurable function, and by theorem 2.1.1 choose a sequence of $\overline{\mathcal{M}}$-measurable functions $\{\phi_i\}_{i=1}^{\infty}$ that converge pointwise to f. The idea is that each $\phi_i = \psi_i$ $\overline{\mu}$-a.e., that is, there exists a E_i where $\phi_i(x) \neq \psi_i(x)$ for all $x \in E_i$ but $\overline{\mu}(E_i) = 0$, and the countable union of these will still be measure 0.

In particular Let $E = \bigcup_i^\infty E_i$. By definition of completion, there must exist some N such that $\mu(N) = 0$ and $E \subseteq N$. Let $g = \lim_{\chi_{X-N}} \psi_n$. Then g is measurable by corollary 2.1.4, and clearly $f = g$ on N^c, completing the proof.

2.2 Integration of non-negative Function

We are now in a position to define how to integrate functions (integrable functions coming soon)! Throughout this section, let $(X, \mathcal{M}, \mu)$ be a measure space

Definition 2.2.1: L^+

Let
$$L^+ = \{f : X \to [0, \infty] \mid f \text{ is a } \mathcal{M}\text{-measurable function}\}$$

Sometimes, L^+ is written as $L^+(X)$, $L^+(\mu)$, or $L^+(\mu, X)$ to emphasize which space of functions we are dealing with. If it is unambiguous, we use L^+.

Definition 2.2.2: Integral of Simple Function

Let $f = \sum_1^n a_i \chi_{E_i}$ be a simple function. Then define the *integral of f with respect to μ* to be
$$\int f d\mu = \sum_i^n a_i \mu(E_i)$$

In other words, we simply took the measure of the domain of χ_{E_i} for each i. As usual, we take the convention that $0 \cdot \infty = 0$. We will also allow $\int f = \infty$ to be a valid result. If there is no confusion, we will write $\int f$ instead of $\int f d\mu$. It should also be emphasized that it is the measure of the *domain* and the coefficient from the *codomain* that we are taking.

If we want to integrate over just some $A \subseteq X$ where $A \in \mathcal{M}$, then $\phi|_A$ is also a simple function (simply take $\chi_{E \cap A}$ for each characteristic function). We usually denote this by
$$\int_A f d\mu = \int f \chi_A \, d\mu$$

Because of this notation, we can write $\int f d\mu$ as
$$\int f d\mu = \int_X f d\mu$$

Proposition 2.2.1: properties of Integrals of Simple Functions

Let ϕ and ψ be simple functions in L^+. Then

1. if $c \geq 0$, $\int c\phi d\mu = c \int \phi d\mu$
2. $\int (\phi + \psi) d\mu = \int \phi d\mu + \int \psi d\mu$
3. if $\phi \leq \psi$ then $\int \phi \leq \int \psi$
4. The map $A \mapsto \int_A 1 d\mu$ is a measure on $\mathcal{M}$ (more generally, $A \mapsto \int_A \phi d\mu$ is a meaure on $\mathcal{M}$ for a simple function ϕ)

***Proof* :**
Do these as midterm exercises

With this definition, we can define integrable function more generally:

Definition 2.2.3: Integral of Measurable Function

Let $f \in L^+$. Then define

$$\int f d\mu := \sup \left\{ \int \phi d\mu \,\middle|\, \phi \text{ is a simple function, } 0 \leq \phi \leq f \right\}$$

It's important to see why the integral of a simple function will give the same result as the "simple integral" (i.e. integral over simple functions from earlier), in particular since the set over which we are containing the supremum will contain the simple function in question, and so the maximum will be achieved by itself (and proposition2.2.1?). Furthermore, results 1 and 3 from proposition 2.2.1 holds for the definition of integration just given. The other results, in particular linearity, will be established soon. We can also see how the definition of the integral solidifies the idea that the integral is a sort of "weighted measure": the integral of a simple function is just a sum of measure of sets weighted by some constant, and we take the supremum of all such sets. We will in fact take a closer look at defining measure in terms of integration of functions in chapter 3.

Given the supremum definition of the integral, it could be rather hard to find the integral of a function in L^+. We essentially never check this, and treat the following theorem as the de-facto way of finding the integral:

Theorem 2.2.1: Monotone Convergence Theorem (MCT)

If $\{f_n\}_{i=1}^{\infty} \subseteq L^+$ is a sequence such that $f_i \leq f_{i+1}$ for all $i \in \mathbb{N}$ and $f = \lim_{n\to\infty} f_n$ ($= \sup_n f_n$), then

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

or, since $f = \lim_{n\to\infty} f_n$:

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

i.e. the limit commutes

It might be tempting to then say that $\int$ is "continuous" in the appropriate domain since it commutes with limits. However, the f_i must be *increasging* for the monotone convergence theorem to apply, and hence it is better to say that $\int$ is continuous from bellow. We will later show that the limit need not always commute (and hence it will be innapropriate to say that the integral is continuous on the space of measurable functions).

Remark Many times in Analysis, when the limit can commute with something (ex. integral, another limit, functions), then the proof that it commutes is usually given a name (ex. sequential continuity, MCT, DCT, Fatou; lemma, Moore-Osgood theorem, etc.) see:

wikipedia: Interchanging Limits

Proof **:**

We'll show $\leq$ and $\geq$. For $\leq$, notice that $\{\int f_n\}$ is an increasing sequence, and so converges (possibly to ∞). Furthermore, $f_n \leq f$ for all n, and so by proposition 2.2.1 updated for general integral functions, we have:

$$\int f_n \leq \int f$$

and so

$$\lim_{n\to\infty} \int f_n \leq f$$

Conversely, we'll show that if we "shrink" the left hand side by a factor of $\alpha \in (0,1)$ so that we get $\geq$, then since this will be true for all α, we will get the result we want.

Fix some $\alpha \in (0,1)$, and choose a simple function where $0 \leq \phi \leq f$ (such a function exists by theorem 2.1.1). Let

$$E_n = \left\{x \in \overline{\mathbb{R}} \;\middle|\; f_n(x) \geq \alpha\phi(x)\right\}$$

Then $\{E_n\}_{n=1}^{\infty}$ is an increasing sequence (via inclusion) of measurable sets who's union is X (since $\lim_{n\to\infty} f_n = f$ and the f_n's are increasing). Then

$$\int f_n \geq \int_{E_n} f_n \geq \alpha \int_{E_n} \phi$$

Since the map $E_n \mapsto \int_{E_n} \phi$ is measurable by proposition 2.2.1 and the E_n's are continuous from above, we have

$$\lim_{n\to\infty} \int_{E_n} \phi = \int \phi$$

and so

$$\lim_{n\to\infty} \int f_n \geq \alpha \int \phi$$

Since this is true for all $\alpha < 1$, it is true for $\alpha = 1$, and so $\int f_n \geq \int \phi$. Taking the supremum over all ϕ, we get that

$$\lim_{n\to\infty} \int f_n \geq \int f$$

And since we showed the $\leq$ direction, we in fact have:

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

as we sought to show

Thus, to find the value of $\int f$, it suffices to find some sequence of simple functions $\{\phi_n\}$ that converges to f, which always exists since every measurable function has a standard representation (theorem 2.1.1).

An immediate consequence of the Monotone Convergence Theorem is that linearity also applies to integrals:

Proposition 2.2.2: Linearity of Integrals

Let $\{f_i\}_{i=1}^N \subseteq L^+$ where $N \in \mathbb{N} \cup \{\infty\}$ and let $f = \sum_i^N f_i$. Then

$$\int f = \sum_i^N \int f_n$$

Proof **:**

Let's first consider the case where $n = 2$ so that we have f_1 and f_2. Let $\{\phi_i\}$ and $\{\psi_i\}$ be a sequence for the standard representation of f_1 and f_2. Then clearly $\{\phi_i + \psi_i\}$ is a sequence for a standard representation of $f_1 + f_2$. Then by the Monotone Convergence Theorem and linearity of integrals with simple functions:

$$\begin{aligned}
\int f_1 + f_2 &= \lim_{i\to\infty} \int \phi_i + \psi_i \\
&\overset{\text{MCT}}{=} \int \lim_{i\to\infty} \phi_i + \psi_i \\
&= \int \lim_{i\to\infty} \phi_i + \lim_{i\to\infty} \psi_i \\
&= \int \lim_{i\to\infty} \phi_i + \lim_{i\to\infty} \psi_i \\
&= \int \lim_{i\to\infty} \phi_i + \int \lim_{i\to\infty} \psi_i \\
&\overset{\text{MCT}}{=} \lim_{i\to\infty} \int \phi_i + \int \lim_{i\to\infty} \psi_i \\
&= \int f_1 + \int f_2
\end{aligned}$$

By induction, it holds that

$$\int \sum_i^n f_i = \sum_i^n \int f_i$$

Letting $n \to \infty$, we can once again apply the Monotone Convergence Theorem and get

$$\begin{aligned}
\int \sum_i^\infty f_i &= \int \lim_{n\to\infty} \sum_i^n f_i \\
&\overset{\text{MCT}}{=} \lim_{n\to\infty} \int \sum_i^n f_i \\
&= \lim_{n\to\infty} \sum_i^n \int f_i \\
&= \sum_i^\infty \int f_n
\end{aligned}$$

completing the proof

The next result is a generalization from the same result with Riemann integrals, except we replace the condition of "finitely many points" with "null-set amount of points". It follows the trend of integration "forgetting" up to a zero-measure amount of points:

Proposition 2.2.3: Zero Integral

Let $f \in L^+$. Then $\int f = 0$ if and only if $f = 0$ a.e.

Proof **:**
Starting with simple functions, this is evident in both directions. If ϕ was our simple function and $\int \phi = 0$, then $\sum_i^n a_i \mu(E_i) = 0$, so either each $a_i = 0$ or each $\mu(E_i) = 0$ which immediately implies ϕ is zero a.e. (not necessarily everywehre since it's possible that $E_i \neq \emptyset$ but $\mu(E_i) = 0$). Conversely, if each $\mu(E_i) = 0$ a.e., then clearly $\int \phi = 0$.

Moving onto to a measurable function f, the $\Leftarrow$ comes almost immediately. If $f = 0$ a.e., then for every simple function $\phi \leq f$, then $\phi = 0$ a.e., which we have established means $\int \phi = 0$. Then by definition of $\int f$:

$$\int f = \sup_{\phi \leq f} \int \phi = 0$$

For the $\Rightarrow$ direction, assume that $\int f = 0$, and for the sake of contradiction let's say $f \neq 0$ a.e. The idea is that we can then construct a simple function ϕ where $\phi \leq f$ whose integral is greater than zero, but $\int f = \sup_{\phi \leq f} \int \phi$. To that end take:

$$\bigcup_i^\infty E_n = \bigcup_i^\infty \left\{ x \;\middle|\; f(x) > \frac{1}{n} \right\} = \{x \mid f(x) > 0\} = E$$

Since $f \neq 0$ a.e., it must be that $\mu(E) > 0$, and so by construction for some n, $\mu(E_n) > 0$. But then

$$f \geq n^{-1} \chi_{E_n} \qquad \Leftrightarrow \qquad \int f \geq n^{-1} \mu(E_n) > 0$$

so the supremum of values in f contain simple functions of nonzero measure, showing that $\int f > 0$ – a contradiction to our original assumption.

This theorem also tells us that if $f = g$ a.e., then $\int f = \int g$, since $f - g = 0$ a.e., so $\int f - g = 0$, so $\int f = \int g$. Using this, we can upgrade the Monotone Conversely Theorem by needing the convergence to happen up to a measure zero set:

Corollary 2.2.1: MCT a.e.

Let $\{f_n\}_{i=1}^\infty \subseteq L^+$ and $f \in L^+$. Then if $f_1 \leq f_2 \leq \cdots \leq f$ and $f_n \to f$ a.e., then

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

Proof **:**
Let $f_n(x) \uparrow f(x)$ for all $x \in E$ such that $\mu(E^c) = 0$. Then $f - f\chi_E = 0$ a.e. which implies $f_n - f_n \chi_E = 0$ a.e., so by the Monotone Conversely Theorem

$$\int f = \int f\chi_E \overset{\text{MCT}}{=} \lim_{n \to \infty} \int f_n \chi_E = \lim_{n \to \infty} \int f_n$$

as we sought to show.

We next address the fact that the sequence $\{f_i\}$ must be increasing (at least a.e.) for the MCT to work (recall we relied on the sequence being increasing in the construction of one of the direcitons of the inequality). If it were not, then it need not be the case that the limit passes through the integral!

Example 2.4: not increasing $\not\Rightarrow$ Limit commutes with $\int$

Let $X = \mathbb{R}$ and μ be the Lebesgue measure. Consider the two following examples:

$$\{n\chi_{(0,1/n)}\}_{n=1}^{\infty}$$

then it is clear that $n\chi_{(0,1/n)} \to 0$. However, notice that $\int n\chi_{(0,1/n)} = 1$ for each element in the sequence. Thus:

$$0 = \int \lim_{n\to\infty} n\chi_{(0,1/n)} \neq \lim_{n\to\infty} \int n\chi_{(0,1/n)} = 1$$

Even if the function is bounded, the problem remains: take $\chi_{(n,n+1)}$ which also converges pointwise to 0 but the sequence of the integrals of the functions is a sequence of constants 1 and so converges to 1 which is clearly not 0.

This is interesting if we interpret it from a sort of "physics" perspective: somehow, we have "lost our mass" in the process of taking our limit. There is in fact essentially 3 ways we can "lose mass', the first two begin the counter-examples above, the last one being a function that oscillates faster and faster (this last one requires a function that is also defined in the negative, and so we will get back to this

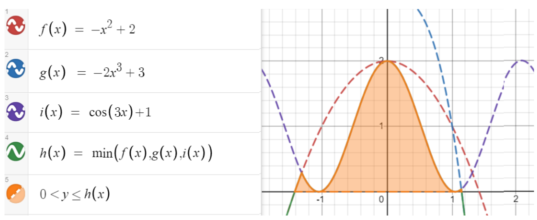
Notice that in both examples, there is some "escaping to infinity" going on. Perhaps if we have some overarching "bound" to the sequence, then we can pass the limit through, and that will indeed be the case. We address this when talking about the Dominated Convergence Theorem in the next section. However, not all hope is lost for the general case; there is weakening to only the $\liminf$ and the equality to $\leq$ which works in the general case:

Theorem 2.2.2: Fatou's Lemma

If $\{f_n\} \subseteq L^+$ is *any* sequence, then

$$\int \liminf_n f_n \leq \liminf_n \int f_n$$

To visualize this, consider:



The integral of the orange region weill be less than or equal to the $\liminf$ of those functions over that region. In this case, the integral would be equal, but in the case where the mass somehow "escapes", we will have an inequality.

***Proof* :**
This comes directly from properties of the lim inf. Let $k \geq 1$. Then

$$\inf_{i \geq k} f_i \leq f_j \quad \forall\, j \geq k$$

Then by theorem 2.2.1

$$\int \inf_{i \geq k} f_i \leq \int f_j \quad \forall j \geq k$$

and since this holds for all $j \geq k$, we get:

$$\int \inf_{i \geq k} f_i \leq \inf_{j \geq k} \int f_j$$

Since the lim inf defines an increasing sequence of function, by letting $k \to \infty$ and applying the Monotone Conversely Theorem:

$$\int \liminf f_n = \lim_{k \to \infty} \int \inf_{n \geq k} f_n \leq \liminf \int f_n$$

as we sought to show

Like before, the result holds if we loosen the condition so that $f_n \to f$ a.e.

Corollary 2.2.2: Fatou's Lemma a.e.

let $\{f_n\} \subseteq L^+$ be any sequence and $f \in L^+$. If $f_n \to f$ a.e. Then

$$\int f \leq \liminf_n \int f_n$$

***Proof* :**
Using proposition 2.2.1, make $f_n \to f$ everywhere arguing that it is equivalent to do so, and then the rest is Fatou's Lemma.

Finally, we show that if $\int f < \infty$, we can guarantee that f cannot "explode" too much, and that even better, the set over which the value of f is non-zero is σ-finite, i.e., it is not "pathological"

Proposition 2.2.4

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

***Proof* :**
was left as an exercise to the reader in the book

2.3 Integration of Real and Complex Functions

We'll again let $(X, \mathcal{M}, \mu)$ be the fixed measure space for this section, except the codomain now will be either $\mathbb{R}$ or $\mathbb{C}$. For the case of $\mathbb{R}$, we can quickly extend our results from last section by splitting f into f^+ and f^- (which are both measurable if f is by corrollary 2.1.3), and if at least one of f^+ or f^- is finite, then

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

Notice that if they are both infinity, we have $\infty - \infty$ which is in general not well-defined. Thus, we will limit our attention to functions where $\int |f| < \infty$ since $|f| = f^+ + f^-$:

For the case of $\mathbb{C}$, using the triangle inequality we get that:

$$|f| = |\Re f + \Im f| \leq |\Re f| + |\Im f| \leq 2|f|$$

and so both $\Re f$ and $\Im f$ must be finite to get a well-defined result. In that case, we define

$$\int f := \int \Re f + i \int \Im f$$

We give a name for compltex functions (the real functions being a spcial case) that satisfy this finiteness condition:

Definition 2.3.1: Integrable Function

Let $f : X \to \mathbb{C}$ be measurable. We say f is *integrable* if

$$\int |f| < \infty$$

Let L^1 represent the set of all integrable functions.

Like with L^+, we sometimes write $L^1(\mu)$, $L^1(X)$, or $L^1(X, \mu)$. If the codomain is $\mathbb{R}$, then we get that $\int |f^+| < \infty$ and $\int |f^-| < \infty$, as we claimed for our extension of the integral to $\mathbb{R}$. Furthermore, for any $E \in \mathcal{M}$, we will say that f is *integrable on* E if $\int_E |f| < \infty$.

The set of L^1 functions are well-behaved under addition and scalar multiplication in the sense that they form a vector-space:

Proposition 2.3.1: Vector Space on L^1

Let L^1 be the set of integrable functions. Then with $+$ and scalar multiplication, L^1 forms a vector-space (over $\mathbb{R}$ or $\mathbb{C}$ respective to the codomain). Furthermore, $\int$ is a linear functional on L^1 (i.e. $\int : L^1 \to \mathbb{R}$ is a linear functional)

Proof **:**
the proof that $cf + dg$ is integrable comes from the triangle inequality ($|cf+dg| \leq |c||f|+|d||g| < \infty$). The fact that $cf \in L^1$ for any $c \in \mathbb{R}$ and $f \in L^1$ follows similarly.

$\int$ being a functional, we need to show linearity and $\int cf = c\int f$. For linearity, assume $h = f + g$

is integrable. Then $h^+ - h^- = f^+ - f^- + g^+ - g^-$, or, re-arranging, we get:

$$h^+ + f^- + g^- = h^- + f^+ + g^+$$

These are all functions in L^+, and so by linearity for of integrals for L^1 functions:

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

re-arraning, we get:

$$\begin{aligned} &= \int f + g \\ &= \int h = \int h^+ - \int h^- = \int f^+ - \int f^- + \int g^+ - \int g^- \\ &= \int f + \int g \end{aligned}$$

showing linearity. Through a similar means, we can also show $\int cf = c \int f$, completing the proof.

Remark Since it is a vector-space, it has a basis and a dimension. What dimension is this vector-space? Since it is a real vector-space, it is isomorphic to $\mathbb{R}^N$ for appropriate cardinal N (N may be finite, countable, uncountable, etc). If $N \geq \aleph_0$, that means that it is rather difficult to define a measure on the space L^1[a]. Trying to find ways to study "measure" notions infinite dimensional vector-spaces (ex. boundedness, σ-finiteness, etc.) is accomplished by defining various different concepts like *seminorms*, *norms*, or *inner products*, which we shall do in chapter 4.

[a] For an illustration of the oddities of infinite dimensional vector-spaces, consider the [hyper]-volume of a unit ball S^n as $n \to \infty$; you should find that it goes to 0!

Next, we would like to establish some properties of integrable functions. One result is a including a useful inequality including absolute values. Another is how the results still work as long as things happen "a.e.". Finally, the set of all "interesting" points (points where $f(x) \neq 0$) are still σ-finite:

Proposition 2.3.2: Properties of Integrable Functions

Let $f, g \in L^1$. Then:

1. $\left|\int f\right| \leq \int |f|$
2. $\{x \mid f(x) \neq 0\}$ is σ-finite
3. for all $E \in \mathcal{M}$, $\int_E f = \int_E g$ if and only if $\int |f - g| = 0$ if and only if $f = g$ a.e.

***Proof* :**

1. The result is essentially immediate in the $\mathbb{R}$ case. In the $\mathbb{C}$ case, let α where $|\alpha| = 1$ be such that it "rotates" the integral so that it's in the real case, and then apply a similar strategy

to that of the $\mathbb{R}$ case.

2. Notice that $\{x \mid f(x) \neq 0\} = \{x \mid f^+(x) > 0\} \cup \{x \mid f^-(x) < 0\}$, which we've shown in proposition 2.2.4 to be σ-finite

3. The fact that $\int |f-g| = 0$ if and only if $f = g$ a.e. comes from proposition 2.2.3, (since $|f-g| \in L^+$) so we will prove $\int_E f = \int_E g$ if and only if $\int |f-g| = 0$. Let's say $\int |f-g| = 0$. Then to show that $\int_E f = \int_E g$, it is equivalent to show that

$$\left| \int_E f - \int_E g \right| = 0$$

by the epsilon principle (i.e. $a = b$ if and only if $|a-b| < \epsilon$ for all ϵ). Using proposition 2.3.2, we get that:

$$\left| \int_E f - \int_E g \right| \leq \int \chi_E |f-g| \leq \int |f-g| = 0$$

completing the $\Rightarrow$ direction.

For the other direction, let's argue contra-positively and assume $\int |f-g| \neq 0$ a.e. Let

$$u = \Re(f-g) \qquad v = \Im(f-g)$$

so that u^+, u^-, v^+, v^- are the corresponding functions. Since $f \neq g$ a.e., at least one of these must have a set with positive measure. Without loss of generality, let's say $E = \{x \mid u^+(x) > 0\}$ and $\mu(E) > 0$. Then

$$\Re\left(\int_E f - g\right) = \Re\left(\int_E f - \int_e g\right) = \int_E u^+ + \int_E u^- = \mu(E) + 0 > 0$$

since $u^- = 0$ on E. Thus, the measure will be positive.

As a consequence of this, we usually think of integrable functions *up to a zero set.* If we wanted to, we can limit our scope to of integration to functions f that are measurable on E and E^c has zero-measure to functions wehre $f = 0$ on E^c. This also means that if we are working with an $\overline{\mathbb{R}}$-value function that is finite a.e., then it is equivalent to treat it as a $\mathbb{R}$-value function.

Proposition 2.3.2 also allows us to re-define L^1 in terms of functions that are equivalent up to a measure zero set. That is, we form an equivalence relation on L^1 if and only if $f = g$ a.e., and label the collection of equivalence classes as L^1. We abuse notation and write "$f \in L^1$" to mean we are selecting a function that is integrable (or an a.e.-defined integrable function). This is because we will usually be working with functions instead of the equivalence classes, and so this is a common abuse of notation.

The reason to introduce this new definition of L^1 is so that we can define a new metric on L^1:

Definition 2.3.2: L^1 metric

Let L^1 be the set of equivalence classes as defined above. Then

$$\rho(f,g) = \int |f-g|$$

is a metric on L^1

Symmetry and triangle inequality come immediately without even needing to use the fact that these are equivalence classes. Where the new construction of L^1 becomes important is to say $\int |f-g| = 0$ if and only if $f = g$. Since this is true a.e., then we need these two to be in an equivalence class. Since we defined a metric, we can define a notion of convergence, in particular, f_n converges to f in L^1 if and only if $\int |f - f_n| \to 0$, meaning $f_n \to f$ almost everywhere (however, L^1 convergence does not imply a.e. point-wise convergence, and a.e. point-wise convergence does not imply L^1 convergence, as we'll show in section 2.4).

We will soon explore more properties of this metric space, in particular, we would want this metric space to be complete so that the limit of a sequence of L^1 function's is L^1 and find some nice dense set of L^1 (so far, we only have the [point-wise] limit of measurable function is measurable, but no such result yet for integrable functions). To do so, we need more tools to analyze swapping limits and integrals. The next result we'll cover allows us to generalize the Monotone Conversely Theorem to the general case of $f_n \to f$ a.e., with the extra condition that each $|f_n|$ is bounded by a single $|g|$ with finite area (in this way, we are getting rid of the "escaping to infinity" cases in example 2.4)

Theorem 2.3.1: Dominated Convergence Theorem (DCT)

Let $\{f_n\}_{i=1}^{\infty} \subseteq L^1$ be a sequence of integrable functions such that

1. $f_n \to f$ a.e.
2. there exists a nonnegative $g \in L^1$ such that for all $n \in \mathbb{N}$, $|f_n| \leq g$ a.e.

Then

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

or equivalently:

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

***Proof* :**
By proposition 2.1.5 and 2.1.6 , f is measurable (up to a null set, in which case we can appropriately redefine f on that null-set). Since $|f_n| < g$ for all $n \in \mathbb{N}$, $f < g$, and so $f \in L^1$.

Since all L^1 functions are broken down to real parts (complex functions into 2 real parts), it suffices to show the result for real value f_n and f. By assumption, we have $g + f_n \geq 0$ and $g - f_n \geq 0$. Thus, by Fatou's Lemma:

$$\int g + \int f \leq \liminf \int (g + f_n) = \int g + \liminf \int f_n$$

$$\int g - \int f \leq \liminf \int (g - f_n) = \int g - \limsup \int f_n$$

Therefore, $\limsup \int f_n \leq \int f \leq \liminf \int f_n$, and so the two value are actually equal and so

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

as we sought to show

Note that functions can still converge even if the hypothesis of the DCT fails:

Example 2.5: Converges Without Dct

Construct a sequence of functions $\{f_n\}$ such that all the following hold:

1. $f_n : [0,1] \to \mathbb{R}$ is continuous
2. $f_n \geq 0$
3. $\lim_{n\to\infty} f_n(x) = 0$ for all $x \in [0,1]$
4. $\lim_{n\to\infty} \int_0^1 f_n(x)\, dm = 0$
5. $\sup_n f_n(x)$ is *not* integrable on $[0,1]$ with respect to the Lebesgue measure

(that is, just because the hypothesis of the dominated convergence theorem fails does not mean the conclusion does not hold: it is not an if and only if)

Proof. The proof essentially is a formalizing the following images:

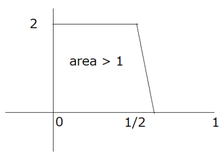


Figure 2.1: f_1

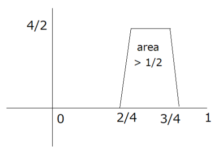


Figure 2.2: f_2

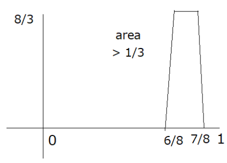


Figure 2.3: f_3

We'll first define the indicator functions, then make these functions continuous. Let

$$g_1(x) = \begin{cases} 2^1/1 & x \in [0, 1/2) \\ 0 & x \in [1 - 1/2, 1] \end{cases} \quad g_2(x) = \begin{cases} 2^2/2 & x \in [1/2, 1/4) \\ 0 & x \in [1 - 1/4, 1] \end{cases} \quad g_3(x) = \begin{cases} 2^3/3 & x \in [3/4, 7/8) \\ 0 & x \in [1 - 7/4, 1] \end{cases}$$

and so on, defining:

$$g_n(x) = \begin{cases} 2^n/n & x \in [2^{n-1} - 2/2^{n-1}, 2^{n-1} - 1/2^{n-1}) \\ 0 & x \in [1 - 2^{n-1} - 1/2^{n-1}, 1] \end{cases}$$

Next, we can make each g_n continuous by simply adding a line to omit the jump:

$$f_1(x) = \begin{cases} 2^1/1 & x \in [0, 1/2) \\ -2^{10\cdot 1}\left(x - \left(\frac{1}{2} - 1/2^{10\cdot 1}\right)\right) & x \in (1/2, 1/2 + 1/2^{10\cdot 1}] \\ 0 & x \in [1 - (1/2 + 1/2^{10\cdot 1}), 1] \end{cases}$$

and defining each f_n similarly. Then it is clear by our construction that $f_n \to 0$: since the trapezoids are moving away, they will never cross our values again once they pass them (this is the opposite of the proof for question 2 where they constantly loop back, since the proof is similar the details were ommited for this proof). Furthermore,

$$\int f_n = 1/n + 1/2^{10n}$$

Thus, $\int f_n \xrightarrow{n\to\infty} 0$, showing that the integral also converges to 0. However,

$$\int \sup_n f_n \geq \int \sup_n g_n = \sum_{k=1}^{\infty} 1/k = \infty$$

showing that the $\sup_n f_n$ is *not* integrable, as we sought to show.

□

Therefore, the DCT is a sufficient, but not necessary condition for convergence. Sufficient and necessary conditions are a bit harder to come by, and we will not need to characterize limits commuting with integrals in these other ways, but for those who are curious:

`https://journalofinequalitiesandapplications.springeropen.com/articles/10.1186/s13660-020-02502-w`

This theorem will be key in finding the integral of many function constructed by limits, and will be used to prove completeness of L^1 (after defining cauchy sequences in L^1 in section ref:HERE) As more immediate result is that the Dominated convergence theorem allow's to define a form of countable linearity for L^1 functions:

Theorem 2.3.2: Linearity for Integrable Functions

Let $\{f_i\}_{i=1}^{\infty} \subseteq L^1$ be a sequence of functions, and $f = \sum_{i=1}^{\infty} f_i$. If $\sum_{i=1}^{\infty} |f_i| < \infty$, then $\sum_{i=1}^{\infty} f_i$ converges a.e. to a function in L^1, and

$$\int \sum_{i=1}^{\infty} f_i = \sum_{i=1}^{\infty} \int f_i$$

***Proof* :**

First, since $\sum_{i=1}^{\infty} |f_i| < \infty$, by proposition 2.2.2,

$$\int \sum_{i=1}^{\infty} |f_i| = \sum_{i=1}^{\infty} \int |f_i|$$

Thus, $g = |\sum_{i=1}^{\infty} f_i|$ is a function in L^1. The function g is finite a.e. by proposition 2.3.2[2], so by the dominated convergence theorem, any partial sum of functions will commute, and so taking the limit, we get:

$$\int \sum_{i=1}^{\infty} f_i = \sum_{i=1}^{\infty} \int f_i$$

as we sought to show

The dominated convergence theorem also gives us a dense set in L^1 to work with, namely the simple functions:

Theorem 2.3.3: Simple Functions Dense in L^1

Let $f \in L^1$. Then for all $\epsilon > 0$, there exists an integrable simple function $\phi = \sum a_i \chi_{E_i}$ such that

$$\int |f - \phi| d\mu < \epsilon$$

Furthermore, if μ is a Lebesgue-Stieljes measure on $\mathbb{R}$, the sets E_i can be taken to be finite union of open intervals, and better, there is a continuous function g that vanishes outside a bounded interval (i.e. $g = 0$ outside the bounded interval) such

$$\int |f - g| d\mu < \epsilon$$

***Proof* :**

p. 55 in the book. Also proven in fuller generality that simple functions are dense in L^p spaces, which includes L^1.

For the purposes of Ordinary Differential Equations (ODE's), we want to establish is a quick result to relate differentiable functions and integral of functions:

Theorem 2.3.4: Relating Derivative and Integral

Let $(X, \mathcal{M}, \mu)$ be a measure space, and suppose $f : X \times [a, b] \to \mathbb{C}$ for $-\infty < a < b < \infty$ and that $f(\cdot, t) : X \to \mathbb{C}$ is integrable for every $t \in [a, b]$, so we can define $F(t) = \int_X f(x, t)\, d\mu$.

1. Suppose that there exists a $g \in L^1$ such that $|f(x, t)| \leq g(x)$ for all x, t. Then if $\lim_{t\to t_0} f(x, t) = f(x, t_0)$ for every x, $\lim_{t\to t_0} F(t) = F(t_0)$; that is, if $f(x, \cdot)$ is continuous at every x, so is F.

2. Suppose $\frac{\partial f}{\partial t}$ exists and there is a $g \in L^1(\mu)$ such that $\left|\frac{\partial f}{\partial t}(x, t)\right| \leq g(x)$ for all x, t. Then F is differentiable and
$$F'(x) = \int \frac{\partial f}{\partial t}(x, t)\, d\mu(x)$$

The way I like to think about is that there is a system X that at different times t acts/behaves differently. This difference in how the system acts at any given time t is captured by $f(x, t)$. The function F capture the "total value" of this action/behavior (note that f is non-negative too, so all behavior has a "non-negative" output).

Now, if the behavior varies continuously, then the total output varies continuously, and if the behavior is smooth smooth enough so that it is differentiable (at any slice), then F is also differentiable, and it is in fact even easy to compute the value of F'.

Proof **:**

1. Since $|f(x,t)|$ is dominated by g, simply apply the Dominated Convergence Theorem:

$$\lim_{t\to t_0} F(t) = \lim_{t\to t_0} \int f(x,t)\,d\mu(x) = \int \lim_{t\to t_0} f(x,t)\,d\mu(x) = \int f(x,t_0)\,d\mu(x) = F(t_0)$$

where $\{t_n\}$ is any sequence converging to t (since any $f(x,t)$ satisfies the bound, no more restrictions are needed on the choice of sequence)

2. Since $\frac{\partial f}{\partial t}$ exists, the limit of the derivative exists, and so define

$$\frac{\partial f}{\partial t} = \lim_{t_n\to t_0} h_n(x) \qquad h_n(x) = \frac{f(x,t_n) - f(x,t_0)}{t_n - t_0}$$

$\{t_n\}$ is any sequence converging to t. Since h_n difference and division of measurable functions, by proposition 2.1.3 it is measurable, and by the Dominated Convergence Theorem $\frac{\partial f}{\partial t}$ is measurable. Furthermore, by the Mean Value Theorem:

$$|h_n(x)| \le \sup_{t\in[a,b]} \left|\frac{\partial f}{\partial t}(x,t)\right| \le g(x)$$

and so by the Dominated Convergence Theorem again:

$$\begin{aligned} F'(t_0) &= \lim_{t_n\to t}\left(\frac{F(t_n) - F(t_0)}{t_n - t_0}\right) \\ &= \lim_{t_n\to t}\left(\int h_n(x)\,d\mu(x)\right) \\ &\overset{\text{D.C.T.}}{=} \int \lim_{t_n\to t} h_n(x)\,d\mu(x) \\ &= \int \frac{\partial f}{\partial t}\,d\mu(x) \end{aligned}$$

completing the proof

The final fact we shall present is a continuation of proposition 2.2.4 in the sense that we aim to further establish connections between some subsets of the domain and the resulting integral value in the co-domain, and see that if $\int$ is limited to some finite area of integration E ($\int_E < \infty$), does it actually respect continuity? In fact, it does! Even better, it is in a sense uniformally continuous with respect to the measure of the set that is being integrated over.

Proposition 2.3.3: $\int$ Uniform Continuity

Let $(X, \mathcal{M}, \mu)$ be a measure space and let $f \in L^1$. Then for all $\epsilon > 0$, there exists a $\delta > 0$ such that

$$\mu(E) < \delta \;\Rightarrow\; \int_E |f| < \epsilon$$

Proof **:**
Without loss of generality, let $f \geq 0$. Let $\epsilon > 0$, and choose ϕ such that $\int f < \int \phi + \epsilon/2$, i.e. $\int f - \int \phi < \epsilon/2$. Next, since ϕ is a simple function, we have that $\sup \phi = M < \infty$ (being the finite sum of constant terms). Now, set $\delta = \frac{\epsilon}{2M}$ so that $\mu(E) < \delta$.

$$\int_E f - \phi \leq \int f - \phi$$

then re-arranging we get:

$$\int_E f \leq \int f - \phi + \int_E \phi < \frac{\epsilon}{2} + \frac{M\epsilon}{2M} = \epsilon$$

as we sought to show.

As a consequence of this, if we have a space of integrable functions $\mathcal{F}$ with the sup norm (convergence in the sup norm is equivalent to convergence), then for any $\mu(E) < \infty$, $\int_E$ is in fact *continuous*, that is, if $f_n \rightrightarrows f$, then:

$$\int_E \lim f_n = \lim \int_E f_n$$

let's fix some E where $\int_E 1 = M < \infty$. Take $f_n \rightrightarrows f$ so that $|f(x) - f_n(x)| < \epsilon$ for all $x \in E$ when $n > N$. Then:

$$\begin{aligned}
\left| \int_E f(x)dx - \int_E f_n(x) \right| &= \left| \int_E f(x) - f_n(x) \right| \\
&\leq \int_E |f(x) - f_n(x)| \\
&\leq \int_E \epsilon \\
&= \epsilon M
\end{aligned}$$

It quickly follows form here that the limit does indeed commute.

2.3.1 Relating Lebesgue and Riemann Integral

Let's consider $f \in L^1(m)$ to be a Lebesgue integrable function on $\mathbb{R}$. As a preliminary to this book, Riemann integration of real-valued functions have been studied. So far, we have done nothing to show that the definitions of the integral will produce the same result. They surely must, since the Riemann integral correctly captures the area under surfaces. In this section, we show how the

Lebesgue integral is in fact an extension of the Riemann integral, meaning if we restrict to the set of all Riemann integrable functions (which will all be Lebesgue integral), then the Lebesgue and Riemann integral give the same result.

As a quick reminder of Riemann integration, choosing to follow Darboux' characterization, let $[a, b]$ be a closed interval (or more generally a compact subset of $\mathbb{R}^n$) Then for any finite partition $P = \{t_0, ..., t_n\}$ where $a = t_0 < t_1 < \cdots < t_n = b$, define:

$$U_P(f) = \sum_{i=1}^{n} M_i(t_i - t_{i-1}) \qquad L_P(f) = \sum_{i=1}^{n} m_i(t_i - t_{i-1})$$

where M_i and m_i are the supremum and infimum of f on $[t_{i-1}, t_i]$. Then take:

$$\overline{I}_a^b(f) = \inf_P U_P(f) \qquad \underline{I}_a^b = \sup_P L_p(f)$$

over all possible partitions P. If $\overline{I}_a^b(f) = \underline{I}_a^b$, then their common value is denoted as $\int_a^b f(x)\,dx$ and f is called Riemann integrable on $[a, b]$.

We'll now compare this to the Lebesgue integral:

Theorem 2.3.5: Lebesgue-Vitali Theorem

Let f be a bounded real-valued function on $[a, b]$.

1. If f is Riemann integrable, then f is Lebesgue measurable (and hence integrable on $[a, b]$ since f is bounded) and
$$\int_a^b f(x)\,dx = \int_{[a,b]} f\,dm$$
2. f is Riemann integrable if and only if $\{x \in [a, b] \mid f \text{ is discontinuous at } x\}$ has Lebesgue measure 0.

Note that $\chi_{\mathbb{Q}}$ on $[0, 1]$ is discontinuous everywhere, with discontinuities having Lebesgue measure 1, was shown to not be Riemann integrable (this was in fact the example of a non Riemann-integrable function presented in the preliminary chapter). However, it is certainly Lebesgue integrable!

***Proof* :**

1. Let f be Riemann integrable. Let M_i and m_i are defined as the supremum and infimum as defined above, and for each partition P, let:
$$G_P = \sum_{i=1}^{n} M_i \chi_{(t_{i-1}, t_i]} \qquad g_P = \sum_{i=1}^{n} m_i \chi_{(t_{i-1}, t_i]}$$
so $U_p(f) = \int G_P\,dm$ and $L_p(f) = \int g_P\,dm$. Now, there exists a sequence of partitions P_n such that $\max_i(t_i - t_{i-1})$ tends to zero and P_{n+1} is a refinement of P_n (meaning g_{P_n} increase in value while G_{P_n} decreases) such that $U_{P_n}(f)$ and $L_{P_n}(f)$ converge to $\int_a^b f(x)\,dx$. Let

$G = \lim G_{P_n}$ and $g = \lim g_{P_n}$. Since $g \le f \le G$, by the dominated convergence theorem:

$$\int G\, dm = \int g\, dm = \int_a^b f(x)\, dx$$

Therefore, $\int G - g = 0$ and so $G = g$ a.e. by proposition 2.2.3, and so $G = f$ a.e.

Finally, since G is measurable (being the point-wise limit of measurable functions)and m is complete, f is measurable and

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

2. This was exercise 23 in the textbook

The big advantage we have with this is that we can now use the computational techniques that were developed for Riemann integrals to compute particular Lebesgue integrals! In most cases in analysis, functions are locally Riemann integrable, and so we might what does the generality give us. The main thing it gives us is *completeness.* So though computing Lebesgue integrals might no be so easy (or if they are not an elementary Riemann integrable function), they are easier to work with in a theoretical framework, and help us define concepts like L^p space (which are important for Fourier Analysis).

(Note that in the infinite case, the Riemann and Lebesgue integral need not match. For example, $\sin(x)/x$ is Riemann integrable and produce a finite value, but it is not Lebesgue integral, producing an infinite value))

(A quick word on the Gamma function)

2.4 Modes of Convergence

So far, we have accumulated the following forms of convergence of functions:

1. uniform convergence $f_n \rightrightarrows f$
2. point-wise convergence $f_n \to f$
3. point-wise convergence a.e. $f_n \to f$ a.e.
4. L^1 convergence $f_n \to f$ if and only if $\int |f_n - f| \to 0$

The first 3 are weaker than then the next: uniform convergence $\Rightarrow$ point-wise convergence $\Rightarrow$a.e. convergence, but the converse is false in each case. Furthermore, L^1 convergence implies none of those 3, and none of those 3 imply L^1 convergence. The examples to keep in mind to compare are:

1. $f_n = n^{-1}\chi_{[0,n]}$
2. $f_n = \chi_{[n,n+1]}$
3. $f_n = n\chi_{[0,1/n]}$

4. $f_1 = \chi_{[0,1]}$, $f_2 = \chi_{[0,1/2]}$, $f_3 = \chi_{[1/2,1]}$, $f_4 = \chi_{[0,1/4]}$, $f_5 = \chi_{[1/4,2/4]}$, and so on

We now introduce another way of thinking about convergence: instead of saying that the $\int f_n$ approaches that of f (i.e., evaluate the integral and look at the limiting behavior of the integral of the function), we will want there to exist an N such that $n \geq N$, the measure of the difference between f_n and f are epsilon away from the function we are converging to f. This is not the same thing as L^1 convergence, since this (1), (3) and (4) all converge to 0 in this type of convergence. Before any further analysis let's define our terms:

Definition 2.4.1: Cauchy in Measure

Let $(X, \mathcal{M}, \mu)$ be a measure space and $\{f_n\}$ a sequence of measurable functions. Then $\{f_n\}$ said to be *cauchy in measure* if:

$$\forall \epsilon > 0,\ \exists N \text{ such that } \forall n, m \geq N,\ \mu(\{x \mid |f_m(x) - f_n(x)| \geq \epsilon\}) < \epsilon$$

Equivalently, we can write the definition as $\forall \epsilon > 0$,

$$\mu(\{x \mid |f_n(x), f_m(x)| \geq \epsilon\}) \to 0 \text{ as } n, m \to \infty$$

Definition 2.4.2: Convergence in Measure

Let $(X, \mathcal{M}, \mu)$ be a measure space and $\{f_n\}$ a sequence of measurable functions. Then $\{f_n\}$ said to be *convergent in measure* or *converge in measure* if:

$$\forall \epsilon > 0,\ \lim_{n\to\infty} \mu(\{x \mid |f(x) - f_n(x)| \geq \epsilon\}) = 0$$

With these definition, notice that (2) is in fact *not* Cauchy in measure: the difference between the two functions.

Proposition 2.4.1: L^1 convergence Then Measure Convergence

Let $(X, \mathcal{M}, \mu)$ be a measure space and let $f_n \to f$ in L^1. Then $f_n \to f$ in measure

***Proof* :**
Let ϵ be given, and define

$$E_{\epsilon,n} = \mu(\{x \mid |f_n(x) - f(x)| \geq \epsilon\})$$

Then:

$$\int |f - f_n| \geq \int_{E_{\epsilon,n}} |f - f_n| \geq \epsilon\mu(E_{\epsilon,n})$$

Thus:

$$\mu(E_{n,\epsilon}) \leq \epsilon^{-1} \int |f - f_n| \to 0$$

showing that it converges in measure.

Note that convergence in measure does not imply convergence in L^1 as we saw in example (1) and (3).

With this established, we can go onto estalbish why we care about convergence in measure: it gives us another way of finding point-wise convergent functions by establishing a condition weaker than L^1 convergence:

Theorem 2.4.1: Cauchy in Measure Then Pointwise a.e.

Let $(X, \mathcal{M}, \mu)$ be a measure space and $\{f_n\}$ Cauchy in measure. Then there exists a f such that $f_n \to f$ in measure, and furthermore there exists a subsequence $\{f_{n_j}\}$ such that $f_{n_j} \to f$ a.e.
If $f_n \to g$ in measure, then $f = g$ a.e.

***Proof* :**

We can choose a subsequence $\{g_j\} = \{f_{n_j}\}$ of $\{f_n\}$ such that if $E_j = \{x : |g_j(x) - g_{j+1}(x)| \geq 2^{-j}\}$, then $\mu(E_j) \leq 2^{-j}$. If $F_k = \bigcup_{j=k}^{\infty} E_j$, then $\mu(F_k) \leq \sum_k^{\infty} 2^{-j} = 2^{1-k}$, and if $x \notin F_k$, for $i \geq j \geq k$ we have

$$|g_j(x) - g_i(x)| \leq \sum_{l=j}^{i-1} |g_{l+1}(x) - g_l(x)| \leq \sum_{l=j}^{i-1} 2^{-l} \leq 2^{1-j}. \tag{2.31}$$

Thus $\{g_j\}$ is pointwise Cauchy on F_k^c. Let $F = \bigcap_1^{\infty} F_k = \limsup E_j$. Then $\mu(F) = 0$, and if we set $f(x) = \lim g_j(x)$ for $x \notin F$ and $f(x) = 0$ for $x \in F$, then f is measurable (see Exercises 3 and 5) and $g_j \to f$ a.e. Also, (2.31) shows that $|g_j(x) - f(x)| \leq 2^{1-j}$ for $x \notin F_k$ and $j \geq k$. Since $\mu(F_k) \to 0$ as $k \to \infty$, it follows that $g_j \to f$ in measure. But then $f_n \to f$ in measure, because

$$\{x : |f_n(x) - f(x)| \geq \epsilon\} \subset \{x : |f_n(x) - g_j(x)| \geq \tfrac{1}{2}\epsilon\} \cup \{x : |g_j(x) - f(x)| \geq \tfrac{1}{2}\epsilon\},$$

and the sets on the right both have small measure when n and j are large. Likewise, if $f_n \to g$ in measure,

$$\{x : |f(x) - g(x)| \geq \epsilon\} \subset \{x : |f(x) - f_n(x)| \geq \tfrac{1}{2}\epsilon\} \cup \{x : |f_n(x) - g(x)| \geq \tfrac{1}{2}\epsilon\}$$

for all n, hence $\mu(\{x : |f(x) - g(x)| \geq \epsilon\}) = 0$ for all ϵ. Letting ϵ tend to zero through some sequence of values, we conclude that $f = g$ a.e.

Corollary 2.4.1: L^1 convergence and Finding Subsequence

Let $f_n \to f$ in L^1. Then there exists a subsequence $\{f_{n_j}\}$ such that $f_{n_j} \to f$ a.e.

Proof **:**
By proposition 2.4.1, $f_n \to f$ in measure, and by theorem 2.4.1, there exists a subsequence that converges to f a.e.

As we saw, $f_n \to f$ a.e. does not imply $f_n \to f$ in measure, as we saw in example (2). This is due to the "escape to infinity" that is happening, which is why we care about the dominated convergence theorem to help us stop such "sneaky infinity problem". However, if μ was a finite measure, then $f_n \to f$ a.e. does imply $f_n \to f$ in measure. In fact, since the area is bounded, we can some even stronger convergent conditions:

Theorem 2.4.2: Ergoff's Theorem

Let $(X, \mathcal{M}, \mu)$ be a measure space and suppose that $\mu(X) < \infty$. If $\{f_i\}$ are measurable complex-valued functions such that $f_n \to f$ a.e., then for every $\epsilon > 0$, there exists a $E \subseteq X$ such that $\mu(E) < \epsilon$ and $f_n \to f$ uniformally on $\mu(E^c)$

Proof **:**

Without loss of generality we may assume that $f_n \to f$ everywhere on X. For $k, n \in \mathbb{N}$ let

$$E_n(k) = \bigcup_{m=n}^{\infty} \{x : |f_m(x) - f(x)| \geq k^{-1}\}.$$

Then, for fixed k, $E_n(k)$ decreases as n increases, and $\bigcap_{n=1}^{\infty} E_n(k) = \varnothing$, so since $\mu(X) < \infty$ we conclude that $\mu(E_n(k)) \to 0$ as $n \to \infty$. Given $\epsilon > 0$ and $k \in \mathbb{N}$, choose n_k so large that $\mu(E_{n_k}(k)) < \epsilon 2^{-k}$ and let $E = \bigcup_{k=1}^{\infty} E_{n_k}(k)$. Then $\mu(E) < \epsilon$, and we have $|f_n(x) - f(x)| < k^{-1}$ for $n > n_k$ and $x \notin E$. Thus $f_n \to f$ uniformly on E^c.

The type of convergence in this theorem is sometimes called *almost uniform convergence*. Almost uniform convergence implies convergene a.e. and convergence in measure.

2.5 Product Measure

As we defined in definition 1.1.4, If $\mathcal{M}$ and $\mathcal{N}$ are σ-algebras, then we can define a new σ-algebra $\mathcal{M} \otimes \mathcal{N}$. If $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are measure spaces, we will show that we can define a new measure on $\mathcal{M} \otimes \mathcal{N}$ using μ and ν. We will use the tools we have done in constructing measure to construct this measure, and show that that there is a natural measure $\mu \times \nu$ with respect to the outer-measure given our space is σ-finite (this is just the uniqueness clause of theorem 1.3.2).

First, we'll give a name to sets in $\mathcal{M} \otimes \mathcal{N}$ that will be easier to measure (these will be our generating sets)

Definition 2.5.1: [Measurable] Rectangle

Let $A \in \mathcal{M}$ and $B \in \mathcal{N}$. Then $A \times B \in \mathcal{M} \otimes \mathcal{N}$ is called a *measurable rectangle* or just *rectangle* for short

Notice that if $A \times B$ and $C \times D$ are rectangles, they form an elementary family:

$$(A \times B) \cup (C \times D) = (A \cap C) \times (B \cap D) \qquad (A \times B)^c = (X \times B^c) \cup (A^c \times B)$$

and so by proposition 1.1.6, the collection of finite disjoint unions of rectangles forms an algebra $\mathcal{A}$, and they generate $\mathcal{M} \otimes \mathcal{N}$.

Furthermore, not all sets in $\mathcal{M} \otimes \mathcal{N}$ are measurable rectangles: if we take $(\mathbb{R}, \mathcal{L}(\mathbb{R}), m)$ and consider $\mathcal{L}(\mathbb{R}) \otimes \mathcal{L}(\mathbb{R})$, then the set of points $\Delta = \{(x,x) \mid x \in \mathbb{R}\}$ is in $\mathcal{L}(\mathbb{R}) \otimes \mathcal{L}(\mathbb{R})$ (being the countable intersection of ever smaller squares on the line), however it is very far from being the product of two sets in $\mathcal{L}(\mathbb{R})$.

We'll now define how to integrate simple functions on $X \times Y$ with the associated measure spaces $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$: Let $A \times B$ be a rectangle that is the (finite or countable) disjoint union of rectangle $A_i \times B_i$. Then notice that:

$$\chi_{A\times B}(x,y) = \chi_A(x)\chi_B(y)$$

and so:

$$\chi_{A\times B}(x,y) = \sum_i \chi_{A_i\times B_i}(x,y) = \sum_i \chi_{A_i}(x)\chi_{B_i}(y)$$

If we integrate with respect to x, treating $\chi_{B_i}(y)$ as some constant, then by theorem 2.2.2

$$\int \chi_A(x)\chi_B(y)d\mu = \mu(A)\chi_B(y)d\mu = \mu(A)\chi_B(y) = \sum_i \mu(A_i)\chi_{B_i}(y)$$

It might be confusing which variable is associated to which measure, in which case it is common to write:

$$\int \chi_A(x)\chi_B(y)d\mu(x)$$

to emphasize the function with the x variable is the one being measured by μ. Integrating with respect to y yields:

$$\int \mu(A)\chi_B(y)d\nu = \mu(A)\nu(B) = \sum_i \mu(A_i)\nu(B_i)$$

Thus, we can define $\pi : \mathcal{A} \to \mathbb{R}$ where if $E \in \mathcal{A}$ is a finite disjoint union of rectangles $A_i \times B_i$, then:

$$\pi(E) = \sum_i \mu(A_i)\nu(B_i)$$

with the usual convention of $0 \times \infty = 0$ (i.e., if one of the dimensions is "0", then we are making the value 0). Notice that π is independent of representation, since any finite disjoint union of rectangles have a common refinement. Thus π is a pre-measure, and so by theorm 1.3.2, π generates an outer measure on $X \times Y$ whose restriction to $\mathcal{M} \otimes \mathcal{N}$ is a measure where restricted to $\mathcal{A}$ is π (or conversely, whose measure extends π). Such a measure is given a name:

Definition 2.5.2: Product Measure

Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{M}, \nu)$ be measure spaces, and define π as in the previous construction. Then the induced outer-measure is called the *product measure* and is usually denoted $\mu \times \nu$.

Notice further that if μ and ν are σ-finite, so $\mu(X) = \sum_i^\infty \mu(A_i)$ and $\nu(Y) = \sum_i^\infty \nu(B_i)$, then $(\mu \times \nu)(X \times Y) = \sum_{i,j} \mu(A_i)\nu(B_j)$, and since all terms are finite, $\mu \times \nu$ is also σ-finite, in which case by the same theorem as invoked in defining $\mu \times \nu$, the product measure is the unique measure such that $\mu \times \nu(A \times B) = \mu(A)\nu(B)$.

Using induction, we can extend this construction to any finite product of measures, that is, for measure spaces $(X_i, \mathcal{M}_i, \mu_i)$, we can define $\mu_1 \times \cdots \mu_n$ on $\mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \cdots \otimes \mathcal{M}_n$ such that

$$\mu_1 \times \mu_2 \times \cdots \mu_n(A_1 \times A_2 \times \cdots \times A_n) = \prod_i^n \mu_i(A_n)$$

and if every μ_i is σ-finite, then the result product measure is the unique extension from the premeasure on the rectangles of $\prod_i^n \mathcal{M}_i$.

Next, (should I prove associativity, it is exercise 45)

With what we've established so far, we can already measure sets using the outer-measure definition. However, this is very tedious, and ultimately comes down to some measure properties of the respective component measure. There is an easier way to use these component measure to get the resulting measure: notice in our construction of π that we integrated with respect to one variable first, and then the next. We will ostensibly generalizing this trick so that we can measure spaces, or more specifically integrate functions on product spaces, by fixing all variables except one, integrating on it, and then moving the result out (since it's a constant) and integrate the next variable, until we integrate over all variables.

Since we can generalise our results by induction, let's take $n = 2$ and consider the measure space $(X \times Y, \mathcal{M} \otimes \mathcal{N}, \mu \times \nu)$. The following definition tries to capture the idea of fixing everything except one variable:

Definition 2.5.3: n-section

Let $(X \times Y, \mathcal{M} \otimes \mathcal{N}, \mu \times \nu)$ be a product measure space, and consider $E \subseteq X \times Y$. Then for any $x \in X$, $y \in Y$, define the *x-section* E_x and *y-section* E^y to be:

$$E_x = \{y \in Y \mid (x, y) \in E\} \qquad E^y = \{x \in X \mid (x, y) \in E\}$$

If f is a function on $X \times Y$, we define the *x-section* f_x and *y-section* f^y to be:

$$f_x(y) = f(x, y) \qquad f^y(x) = f(x, y)$$

The most simple example of the section of a function would be:

$$(\chi_E)_x = \chi_{E_x} \qquad (\chi_E)^y = \chi_{E^y}$$

Proposition 2.5.1: Measure of Section

1. Let $E \in \mathcal{M} \otimes \mathcal{N}$. Then $E_x \in \mathcal{N}$ and $E^y \in \mathcal{M}$ for all $x \in X$ and $y \in Y$ (i.e., all sets in a produce σ-algebra can be broken down into x-sections and y-section)

2. If f is $\mathcal{M} \otimes \mathcal{N}$-measurable, then f_x is $\mathcal{N}$-measurable and f^y is $\mathcal{M}$-measurable

***Proof* :**

1. The proof of this fact comes down to taking advantage of the fact that the pre-measure of rectangles generates $\mathcal{M} \otimes \mathcal{N}$. Let $\mathcal{R}$ be the collection of all subsets of $E \subseteq X \times Y$ such that $E_x \in \mathcal{N}$ and $E^y \in \mathcal{M}$ for all $x \in X$ and $y \in Y$. This clearly contains the empty set, and the set must contain all rectangles, since if $A \neq 0$ and $x \in A$, then $(A \times B)_x = B$ (if $x \notin A$, then $(A \times B)_x = \emptyset$). Now, since

$$\left(\bigcup_i^\infty E_i\right)_x = \bigcup_i^\infty (E_i)_x \qquad (E^c)_x = (E_x)^c$$

by some basic set manipulation, we see that $\mathcal{R}$ is a σ-algebra. Since it contains the generating set of $\mathcal{M} \otimes \mathcal{N}$, we have that $\mathcal{R} \supseteq \mathcal{M} \otimes \mathcal{N}$. Thus, all the sets in $\mathcal{M} \otimes \mathcal{N}$ can be decompose in this way (note that we have not proved equality: we have not proved that if every slice is measurable then E is measurable)

2. By some set-theory manipulation, and part (1):

$$f_x^{-1}(B) = (f^{-1}(B))_x \qquad f^y{-}1(B) = (f^{-1}(B))^y$$

and so f_x and f^y must be $\mathcal{N}$ and $\mathcal{M}$ measurable (respectively)

With this, we are almost ready to start proving we can integrate $A \times B \in \mathcal{M} \otimes \mathcal{N}$ with respect to 1 measure at a time. We first need to prove a technical result for the next theorem which gives (again) another way of constructing σ-algebra. This one is particularly useful, since it will let us prove certain sets are σ-algebras by using the MCT and DCT:

Definition 2.5.4: Monotone Class

Let X be a set and $P(X)$ the powerset of X. Then the subset $\mathcal{C} \subseteq P(X)$ is called a *monotone class* if

1. it is cosed under countable unions of increasing sets, that is, if $E_1 \subseteq E_2 \subseteq \cdots$ where $\{E_i\} \subseteq \mathcal{C}$, then $\cup_i E_i \in \mathcal{C}$

2. it is cosed under intersection unions of decreasing sets, that is, if $E_1 \supseteq E_2 \supseteq \cdots$ where $\{E_i\} \subseteq \mathcal{C}$, then $\cap_i E_i \in \mathcal{C}$

Clearly, every σ-algebra is a monotone class (we've even used this property in a couple of proofs). Furthermore, the intersection of any family of monotone class is again a monotone class, and so we can define the unique smallest monotone class that contains the set $\mathcal{E} \subseteq P(X)$, where $\mathcal{E}$ is called the monotone class generated set, and the monotone class is called the monotone class generated by $\mathcal{E}$, and is dentoed $\mathcal{C}(\mathcal{E})$

Lemma 2.5.1: Monotone Class Lemma

Let X be a set and $\mathcal{A} \subseteq X$ be an algebra on some subsets of X. Then the monotone class $\mathcal{C}$ generated by $\mathcal{A}$ is equal to the σ-algebra $\mathcal{M}$ generated by $\mathcal{A}$, that is:

$$\mathcal{C}(\mathcal{A}) = \mathcal{M}(\mathcal{A})$$

***Proof* :**

Since $\mathcal{M}$ is itself a monotone class, we automatically have $\mathcal{E} \subseteq \mathcal{M}$, so we'll prove $\mathcal{E} \supseteq \mathcal{M}$.

This is just a lot of set-theory manipulation, I'll leave it for another time

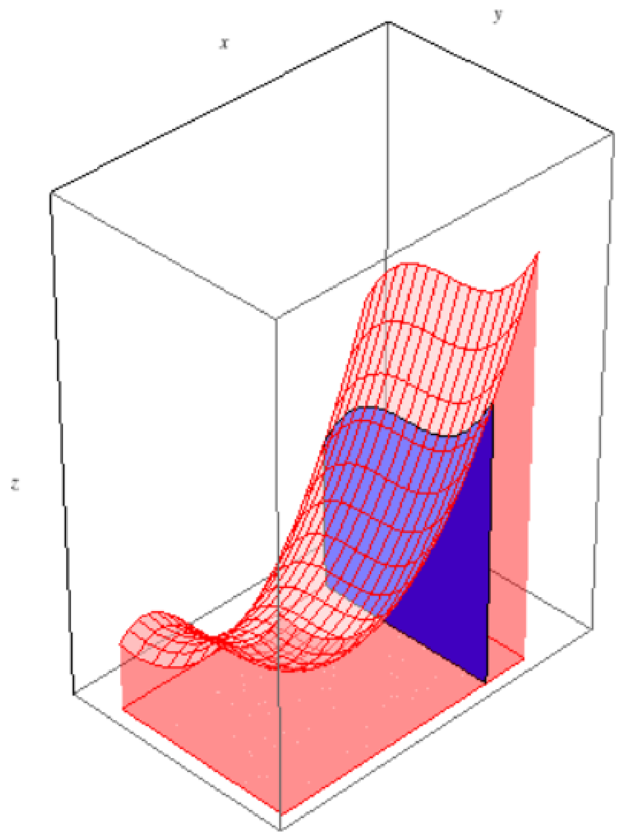
We now are able to start proving the main results:

Theorem 2.5.1: Fubini-Tonelli for Characteristic Functions

Let $(X, \mathcal{M}, \mu)$ and $(N, \mathcal{N}, \nu)$ be measure spaces with μ, ν being σ-finite. Then if $E \in \mathcal{M} \otimes \mathcal{N}$, the functions $x \mapsto \nu(E_x)$ and $y \mapsto \mu(E^y)$ are measurable on X and Y respectively, and

$$(\mu \times \nu)(E) = \int \nu(E_x) d\mu = \int \mu(E^y) d\nu$$

If you are visual, you can think of $\mu(E_x)$ as measuring the area under the slice (inside the appropriate measure so that it doesn't have measure 0), and see that are saying there exists a function that maps x to every slice of the graph:



In particular, it maps x the measure of that slice, so we get a new function which at every slice captures the effect of the X component. After that, we integrate the function to get the effect from the Y component.

Furthermore, it will turn out that we can measure the slices with respect to X, or with respect to Y, and get the same result, hence the second statement of this theorem! After proving the theorem, we'll go over why the assumption in the statement of the theorem are important.

Proof **:**
We'll first prove it for when μ and ν are finite, and then generalize for when they are σ-finite. The way we'll do it is by starting with a dense subset of $\mathcal{M} \otimes \mathcal{N}$ for which we know the conclusion holds, and then generalize the result using the MCT and DCT. In particular, we will be taking advantage that rectangles already satisfy the property of the theorem, and since rectangles are a generating set for $\mathcal{M} \otimes \mathcal{N}$, we will be able to extend this result any measurable set.

Let $\mathcal{C}$ be the subset of $\mathcal{M} \otimes \mathcal{N}$ for which the result of the theorem is true, that is, the two maps are measurable and we can integrate one way or another. Then if $A \in \mathcal{M}$ and $B \in \mathcal{N}$, we have that $A \times B \in \mathcal{C}$. To see this, first compute the values $\nu(E_x)$ and $\mu(E^y)$:

$$\nu(E_x) = \chi_A(x)\nu(B) \qquad \mu(E^y) = \mu(A)\chi_B(y)$$

Then the functions is measurable since it's a slice (theorem 2.5.1), and:

$$\int \nu(E_x)d\mu = \int \chi_A(x)\nu(B)d\mu = \mu(A)\nu(B) = \int \mu(A)\chi_B(y)d\nu = \int \mu(E^y)d\nu$$

Showing the two integrals are equivalent, and since $(\mu \times \nu)(A \times B) = \mu(A)\nu(B)$, all equality hold, and so $E = A \times B \in \mathcal{C}$. By additivity of the integral, it follows that all finite disjoint unions of rectangles are also in $\mathcal{C}$. Thus, by lemma 2.5.1, it suffices to show that $\mathcal{C}$ is a monotone class, for then it is a σ-algebra containing the generating set of $\mathcal{M} \otimes \mathcal{N}$, meaning $\mathcal{C} \supseteq \mathcal{M} \otimes \mathcal{N}$, and so all measurable sets can be broken down in this way.

First, let $\{E_n\}_{n=1}^\infty \subseteq \mathcal{C}$ be an increasing subset and let $E = \bigcup_n^\infty E_n$. We want to show that $E \in \mathcal{C}$. Since $E_n \in \mathcal{C}$ for all $n \in \mathbb{N}$, the sequence of y-section functions f_n, $f_n(y) = \mu((E_n)^y)$ are measurable and increase pointwise to $f(y) = \mu(E^y)$. Since the limit of measurable functions is measurable (corrollary 2.1.4), f is measurable. Thus, we can find the value of $\int \mu(E^y)d\nu$ by applying the Monotone Convergence Theorem:

$$\begin{aligned}
\int \mu(E^y)d\nu &= \int \lim_{n\to\infty} \mu((E_n)^y)d\nu \\
&= \lim_{n\to\infty} \int \mu((E_n)^y)d\nu && \text{MCT} \\
&= \lim_{n\to\infty} (\mu \times \nu)(E_n) && E_n \in \mathcal{C} \\
&= (\mu \times \nu)(E) && \cup_i E_i = E
\end{aligned}$$

Doing the same trick with respect to x, we can see that:

$$\int \nu(E_x)d\mu = (\mu \times \nu)(E)$$

which combined together shows us that:

$$\int \mu(E^y)d\nu = \int \nu(E_x)d\mu = (\mu \times \nu)(E)$$

and so $E \in \mathcal{C}$.

Next, let $\{E_n\}_{n=1}^\infty \subseteq \mathcal{C}$ be a decreasing subset and let $\bigcap_n E_n = E$. Then since $E_1 \in \mathcal{C}$, the function f_1, $f_1(y) = \mu((E_1)^y)$ is measurable. In fact, it's in L^1, since:

$$\mu((E_1)^y) \leq \mu(X) < \infty \qquad \nu(Y) < \infty$$

and so the resulting integral must have finite measure. Furthermore, f_1 will dominate the sequence of functions f_n, and so we can apply the Dominated Convergence Theorem to do the same limit trick we have just done and conclude that $E \in \mathcal{C}$. Therefore, $\mathcal{C}$ is a monotone class, hence a σ-algebra containing the generating set of $\mathcal{M} \otimes \mathcal{N}$, and the result is true for all measurable sets.

Now, suppose that μ and ν are σ-finite. Then we can write $X \times Y$ to be the union of an increasing sequence of rectangles $\{X_i \times Y_i\}$ of finite measure. Thus, for any $E \in \mathcal{M} \otimes \mathcal{N}$, take $E \cap (X_i \times Y_i)$, then this set now has finite measure, and so what we have just proved applies giving us:

$$(\mu \times \nu(E \cap (X_i \times Y_i)) = \int \chi_{X_i}(x)\nu(E_x \cap Y_i)d\mu = \int \mu(E^y \cap X_i)\chi_{Y_i}(y)d\nu$$

and so, a final application of the Monotone Convergence Theorem gives us our desired result

The hypothesis that $\mu \times \nu$ is σ-finite is necessary:

Example 2.6: Necessity of σ-finite

To compute the iterated integrals, we start by finding the x and y section. Using the definition of the diagonal we get that: $(\chi_D)_x = \chi_{\{x\}}$ and $(\chi_D)_y = \chi_{\{y\}}$. Thus, we get:

$$\int \left[\int \chi_{\{y\}} d\mu(x) \right] d\nu(y) = \int 0\nu(y) = 0$$

and

$$\int \left[\int \chi_{\{x\}} d\nu(y) \right] d\mu(x) = \int 1 d\mu(x) = 1$$

showing that the iterated iterated iterated integrals are not equal. For the integral $\int \chi_D \, d(\mu \times \nu)$, since χ_D is a simple function, we have shown in class that the integral is the measure of D with respect to the product measure $\mu \times \nu$ (namely, since χ_D will be in the supremum set over which we are computing the integral), which is:

$$\int \chi_D d(\mu \times \nu) = \mu \times \nu(D) = \inf \left\{ \sum_i^\infty \mu(A_i)\nu(B_i) \;\middle|\; A_i \times B_i \in \mathcal{A},\ D \subseteq \bigcup_i A_i \times B_i \right\}$$

where $\mathcal{A}$ is the collection of all rectangles of $\mathcal{M} \otimes \mathcal{N}$. We'll show $\mu \times \nu(D) = \infty$.

Take any countable cover of D by rectangle $\mathcal{C}$ (so $\{A_i \times B_i\}_{i=1}^\infty = \mathcal{C}$ for appropriate $A_i \times B_i$). Without loss of generality, we can make these covers disjoint; we can refine the rectangles. If we can show that there must be a non-zero measure set A_i and an infinite measure set B_i such that $A_i \times B_i \in \mathcal{C}$ then we are done. For the sake of contradiction, let's say no such set exists in any countable cover $\mathcal{C}$. We'll show that $\mathcal{C}$ doesn't cover D.

Split $\mathcal{C}$ into $F \subseteq \mathcal{C}$ and $I \subseteq \mathcal{C}$ where F is the collection of all sets where A_i has ***finite*** non-zero measure (the sets have finite measure since if $A_i \subseteq X$, $\mu(A_i) \leq \mu(X) = 1$), and I is the set where B_i has ***infinite*** measure. By our contradiction assumption, these set's are disjoint, since the first component of the sets in I must have zero measure.

The set F can only cover countably many points of D: Since each B_i has finite measure, any set $A_i \times B_i \in F$ can cover at most finitely many points of D (given appropriate A_i), and so $\bigcup_{A_i \times B_i \in F} A_i \times B_i$ can at most cover countably many points of D (since the 2nd component of D is covered by elements from B_i given appropriate A_i). Therefore, an uncountable number of points are not being covered by F.

The set I misses at least uncountably many points. Since each component A_i has zero-measure and the collection of x components of D (i.e. the set $[0,1]$) has measure one, the union of the sets A_i cannot cover all of the x-components (given appropriate B_i components) or else $0 = \mu(\bigcup_i A_i) \geq \mu([0,1]) = 1$, a contradiction. If the B_i component's don't line up correctly, then this set cover's even fewer points of D. Therefore, I can at most cover a measure zero amount of x-components points from D. Since any set two sets containing each other with a different measure (ex. $\alpha \subseteq \beta$, $0 = \mu(\alpha) \leq (\beta) = 1$) must have a difference with at least uncountable cardinality[a] ($|B \setminus A| \geq \aleph_1$), the set I must miss at least uncountably many points.

However, the cover F can only cover at most countably many points. Therefore, the cover $F \cup I = \mathcal{C}$ cannot cover all of D, a contradiction to the assumption that $\mathcal{C}$ is a cover for D. Hence, there is at least one set $A_i \times B_i$ with $\mu(A_i) > 0$ and $\mu(B_i) = \infty$ for any cover of D, meaning all covers of

D must have a set with infinite measure, and so $\infty \leq \mu \times \nu(D)$, and hence

$$\int \chi_D \, d(\mu \times \nu) = \infty$$

and so the two iterated integrals, as well as the integral of χ_D has unequal measures, as we sought to show.

[a] If two sets containing each other had differed by countably many point, then they would have the same measure

Notice that the functions within the integral of the previous theorem are in fact the values of characteristic functions! Hence by the linearity of the integral, we have the result for simple functions. Thus, with the statement proven for simple functions on $\mathcal{M} \otimes \mathcal{N}$, we move on to proving it for integrable functions in general:

Theorem 2.5.2: Fubini-Tonelli Theorem

Let $(X, \mathcal{M}, \mu)$ and $(N, \mathcal{N}, \nu)$ be σ-finite measures. Then:

1. (Tonelli) If $f \in L^+(X \times Y)$, then the functions $g(x) = \int f_x d\nu$ and $h(y) \int f^y d\mu$ are in $L^+(X)$ and $L^+(Y)$ respectively, and

$$\int \left[\int f(x,y) d\nu(y) \right] d\mu(x) = \int f(x,y) d(\mu \times \nu) = \int \left[\int f(x,y) d\mu(x) \right] \nu(y)$$

2. (Fubinni) If $f \in L^1(\mu \times \nu)$, then $f_x \in L^1(\nu)$ for a.e. $x \in X$ and $f^y \in L^1(\mu)$ for a.e. $y \in Y$. Furthermore, the a.e. defined function $g(x) = \int f_x d\nu$ and $h(x) = \int f^y d\mu$

The resulting integral is called the *iterated integral*

***Proof* :**

1. This essentially comes down to using MCT and the standard representation of measurable functions. By theorem 2.5.1, we have the simple functions satisfy the theorem. To extend it to integrable functions, take $f \in L^+(X \times Y)$, and let $\{f_n\}$ be some simple representation for it, that is, a sequence of simple functions converging upwards to f. From $\{f_n\}$, we have a corresponding $\{g_n\}$ and $\{h_n\}$ converging upwards to g and h: $g_n = \int (f_n)_x$ and $h_n = \int (f_n)^y$. By theorem 2.5.1, these functions are measurable, and so h and g are measurable. With this,

we can put together all our information to form the following chain of equalities:

$$\begin{aligned}
\int \left[\int f_x \, d\nu \right] d\mu = \int g d\mu = \int \lim g_n \, d\mu & \\
= \lim \int g_n \, d\mu & \\
= \lim \int f_n \, d(\mu \times \nu) & \qquad \text{theorem } 2.5.1 \\
= \int f \, d(\mu \times \nu) & \qquad \text{MCT} \\
= \lim \int f_n \, d(\mu \times \nu) & \\
= \lim \int h_n \, d\nu & \qquad \text{theorem } 2.5.1 \\
\int \left[\int f^y \, d\mu \right] d\nu = \int h d\nu = \int \lim h_n \, d\nu &
\end{aligned}$$

Thus establishing the equality of Tonelli's Theorem. Furthermore, since $\int f \, d(\mu \times \nu) < \infty$, then $g < \infty$ a.e. and $h < \infty$ a.e., so $\int f_x < \infty$ and $\int f^y < \infty$, and therefore $f_x \in L^1(\nu)$ and $f^y \in L^1(\mu)$, which some of the preliminary results for Fubini's Theorem.

2. Let $f \in L^1(\mu \times \nu)$, so $f^+ \in L^+(\mu \times \nu)$ and $f^- \in L^+(\mu \times \nu)$ (or $\Re f$ and $\operatorname{sgn} f$). Thus, the result applies to each component function, and so by linearity applies to f

Often, the parenthesis are dropped, and we combine the integrals into a double integral:

$$\int \left[\int f(x,y) \, d\mu(x) \right] d\nu(y) = \int \int f(x,y) \, d\mu(x) \, d\nu(y) = \int \int f \, d\mu \, d\nu$$

Also, you might wonder if instead of assuming $f \in L^+(X \times Y)$ or $f \in L^1(X \times Y)$, is it possible to instead assume that each $f_x \in L^+(\nu)$, $f^y \in L^+(\mu)$ (L^1 respectively) and that the iterated integrals $\int \int f \, d\mu \, d\nu$ and $\int \int f \, d\nu \, d\mu$ existed, then $f \in L^+(X \times Y)$ (L^1 respectively), that is, is the converse true? The answer is actually no:

Example 2.7: converse not true

Let $X = Y = \mathbb{N}$, $\mathcal{M} = \mathcal{N} = P(\mathbb{N})$, $\mu = \nu$ be the counting measures. Define:

$$f(m,n) = \begin{cases} 1 & m = n \\ -1 & m = n+1 \\ 0 & \text{otherwise} \end{cases}$$

Then $\int |f| d(\mu \times \nu) = \infty$, and $\int \int f d\mu d\nu$, $\int \int f d\nu d\mu$ exist and are unequal.

This is not quite a converse because the two itterated integrals are unequal, and I don't know of an example where they are equal, but $\int f d(\mu \times \nu)$ is not equal to it (i.e. infinity).

One practical application of Fubini-Tonelli's Theorem is to solve integrals by integrating over the simpler variable first. Usually, one first uses Tonelli's theorem to evaluate the integral $\int f \, d(\mu \times \nu)$ as an iterated integral $\int\int f\,d\mu\,d\nu$ and show that the result is finite, at which point Fubini's Theorem can be invoked to get $\int\int f\,d\mu\,d\nu = \int\int f\,d\nu\,d\mu$.

To close off the discussion on product measures, notice that if μ and ν are complete measure, $\mu \times \nu$ is almost never a complete measure. Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be measure spaces, take $A \in \mathcal{M}$ is a non-empty set with $\mu(A) = 0$, and suppose $\mathcal{N} \neq P(Y)$ (For example, choose $(\mathbb{R}, \mathcal{L}, m)$ for both measure spaces). If $E \in P(Y) \setminus \mathcal{N}$, then $A \times E \notin \mathcal{M} \times \mathcal{N}$ since not all slices of $A \times E$ (in particular, the slice E) are in the respective measure spaces (proposition 2.5.1). However, $A \times E \subseteq A \times Y$, and $(\mu \times \nu)(A \times Y) = 0$ (since the measure is finite, any out-measure is equal to the usual outer-measure on these sets, which is $\mu(A)\nu(Y) = 0 \cdot n = 0$, since we take on the convention that $0 \cdot \infty = 0$). Thus, $\mu \times \nu$ is not complete. If we want, we can apply theorem 1.2.1 to complete the measure, however we can no longer directly apply Fubini-Tonelli's Theorem. (Put Why Here!). We thus have to slightly amend Fubini-Tonelli's theorem for the completion:

Theorem 2.5.3: Fubini-Tonelli's Theorem for Complete Measures

Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be complete σ-finite measures and let $(X \times Y, \mathcal{L}, \lambda)$ be the completion of $(X \times Y, \mathcal{M} \otimes \mathcal{N}, \mu \times \nu)$. If f is $\mathcal{L}$-measurable is either:

1. non-negative, that is $f \geq 0$
2. $f \in L^1(\lambda)$

Then f_x is $\mathcal{N}$-measurable for a.e. x and f^y is $\mathcal{M}$-measurable for a.e. y, and if (2) applies as well, then f_x and f^y are integrable for a.e. x and y. Furthermore, $x \mapsto \int f_x\,d\nu$ and $y \mapsto \int f^y\,d\mu$ are measurable (and in the case of (2), also integrable) and

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

***Proof* :**
We'll do the following:

1. If $E \in \mathcal{M} \otimes \mathcal{N}$ and $\mu \times \nu(E) = 0$, then $\nu(E_x) = \mu(E^y) = 0$ for a.e. x and y
2. If f is $\mathcal{L}$-measurable and $f = 0$ λ-a.e., then f_x and f^y are integrable for a.e. x and y, and $\int f_x d\nu = \int f^y d\mu = 0$ for a.e. x and y (here, the completeness of μ and ν is needed)

lemma 1:

Proof. Note that $\chi_E \in L^+(X \times Y)$ and $(X, \mathcal{M}, \mu)$ $(Y, \mathcal{N}, \nu)$ are σ-finite, and so Tonelli's Theorem applies, meaning

$$\int \chi_E d(\mu \times \nu) = \int \left[\int (\chi_E)^y d\mu(x)\right] d\nu(y) = \int \left[\int (\chi_E)_x d\nu(y)\right] d\mu(x)$$

First, since χ_E is an indicator function, we see that $(\chi_E)_x = \chi_{E_x}$, since $\chi_{E_x}(y) = \chi_E(x,y)$.

Similarly, $(\chi_E)^y = \chi_{E^y}$. Therefore, we have:

$$\begin{aligned} 0 &= \int \chi_E d(\mu \times \nu) \\ &= \int \left[\int (\chi_E)^y d\mu(x) \right] d\nu(y) \\ &= \int \left[\int (\chi_E)_x d\nu(y) \right] d\mu(x) \end{aligned}$$

showing that the function $x \mapsto \int \chi_{E_x} d\nu$ and $y \mapsto \int \chi_{E^y} d\mu$ is 0 a.e., which implies $\nu(E_x) = 0$ and $\mu(E^y) = 0$ a.e. ☐

lemma 2: Multiple parts of this proof take advantage of the fact that if μ is a measure and $\mu(A) = 0$, then $\int_A f d\mu = 0$ (since all simple functions ϕ approximating f with the property of $\phi \leq f$ will have their integral be of the form $\sum_i^n c_i \chi_{A_i} = \sum_i^n c_i \cdot 0 = 0$ for $A_i \subseteq A$, and so the supremum of the integral of such simple functions must also be 0)

Proof. Let f be $\mathcal{L}$-measurable and $f = 0$ λ-a.e. We want to show that f_x and f^y is integrable for a.e. x and y and $\int f_x d\nu = \int f^y d\mu = 0$ for a.e. x and y. What we'll do is take advantage of Fubini-Tonelli and theorem 2.12 (which states that there exists a $\mathcal{M} \otimes \mathcal{N}$-measurable function g such that $f = g$ λ-a.e.)

First, since $f = g$ λ-a.e., $f - g = 0$ λ-a.e., meaning $(f - g)_x = f_x - g_x = 0$ for a.e. x (assuming completeness) and $f^y - g^y = 0$ for a.e. y (assuming completeness), and so $f_x = g_x$ and $f^y = g^y$ for a.e. x and y. This will be useful to us later.

Next, Consider $\int f d\lambda$. We want to take advantage of Fubini's theorem, which is true for g, and so we must re-write this expression in such a way that we get it equal to g. Let's say $S = \{(x, y) \in X \times Y \mid f(x, y) \neq g(x, y)\}$. We know that $\lambda(S) = 0$, and so:

$$0 = \int f\, d\lambda = \int_S f\, d\lambda + \int_{S^c} f\, d\lambda = \int_{S^c} f\, d\lambda \tag{2.1}$$

Since $S^c \in \mathcal{L}$, we have that $S^c = A \sqcup N$ where $A \in \mathcal{M} \otimes \mathcal{N}$ and $N \subseteq B$ where $\mu \times \nu(B) = 0$. Thus, we get:

$$\int_{S^c} f d\lambda = \int_A f d\lambda + \int_N f d\lambda = \int_A f d\lambda$$

since $f = g$ on S^c, in particular $f = g$ on $A \in \mathcal{M} \otimes \mathcal{N}$, and λ is the extension of $\mu \times \nu$, we can re-write this as:

$$\int_A f d\lambda = \int_A g d(\mu \times \nu)$$

and so Fubini-Tonelli applies, giving us the iterated integrals are equal to 0 (carrying the 0 from equation (2.1)):

$$\int \left[\int_{A_x} g_x d\nu(y) \right] d\mu(x) = 0$$

and similarly for the y-section. Since $f = g$ on S^c, by what we've shown earlier, $f_x = g_x$ on A_x, (and similarly $f^y = g^y$ on A^y) and so:

$$\int_{A_x} f_x d\nu(y) = 0 \qquad \text{for a.e. } x$$

and similarly for the y-sections, which brings us close to our goal (it is possible that $A_x \subsetneq X$ or $A^y \subsetneq Y$). We now have to add back in points which we eliminated since the integral was zero on S and N:

$$\int_S f d\lambda = 0 \qquad \int_N f d\lambda = 0$$

Since they are measure zero sets in $\mathcal{L}$, there exists a $C, D \in \mathcal{M}\otimes\mathcal{N}$ such that $\mu\times\nu(C) = \mu\times\nu(D) = 0$ and $S \subseteq C$, $N \subseteq D$. Let $E = C \cup D$ so that $S \cup N \subseteq E$ and $\mu \times \nu(E) = 0$. Thus, we have:

$$\int_E f d\lambda = 0$$

since $\mu \times \nu(E) = 0$ and $E \in \mathcal{M} \otimes \mathcal{N}$, by the first lemma, $\mu(E^y) = \nu(E_x) = 0$ for a.e. x and y. But then, for a.e. x and y, $\int_{E_x} f_x d\nu(y) = \int_{E^y} f^y d\mu(x) = 0$. Since $X \times Y = S^c \cup S = A \cup E$ (by how we've broken the set down), then $X = A_x \cup E_x$ and $Y = A^y \cup E^y$, and so

$$0 \leq \int f_x d\nu(y) \leq \int_{A_x} f_x d\nu(y) + \int_{E_x} f_x d\nu(y) = 0$$

for a.e. x, implying $\int f_x d\nu(y) = 0$ for a.e. x (and similarly for y). Hence, they are integrable for a.e. x and y, completing the proof. $\square$

With this, we can prove Fubini-Tonelli's Theorem for complete measures. First, let's build-up some result and notation. Let $f \in L^+(\lambda)$ (resp. $f \in L^1(\lambda)$). Then by theorem 2.12, there exists a g such that $f = g$ λ-a.e., meaning that $f - g = 0$ λ-a.e. Thus, by the 2nd lemma, $\int (f-g)_x = 0$ and $\int (f-g)^y = 0$ are integrable for a.e. x and y. Therefore, $\int f_x - g_x = 0$ and $\int f^y - g^y = 0$ for a.e. x and y, and so $\int f_x = \int g_x$ and $\int f^y = \int g^y$ for a.e. x and y.

We can now start concluding the proof. Since μ and ν are complete and $f_x = g_x$ ($f^y = g^y$) for a.e. x and y, by proposition 2.11, f_x is $\mathcal{N}$-measurable and f^y is $\mathcal{M}$-measurable for a.e. x and y, completing the 1st assertion of the theorem. In the $f \in L^1(\lambda)$ case, the proof we've just done in the previous paragraph shows that each f_x and f^y are integrable for a.e. x and y.

Next, since $G_x : x \mapsto \int g_x d\nu$ and $G^y : y \mapsto \int g^y d\mu$ is measurable (resp. integrable), by theorem 2.11, $F_x : x \mapsto \int f_x d\nu$ and $F^y : y \mapsto \int f^y d\mu$ is measurable (resp. integrable). Finally, since $F_x = G_x$ for a.e. x and $F^y = G^y$ for a.e. y, their integrals match, and since g is $\mathcal{M}\otimes\mathcal{N}$ measurable, Tonelli's (resp. Fubini's) theorem applies, and so by one more application of theorem 2.11 :

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

Completing the proof

2.6 The n-dimensional Lebesgue Integral

In this section, we focus on $\mathbb{R}^n = \mathbb{R}\times\mathbb{R}\times\cdots\times\mathbb{R}$ with the completed product measure $m\times m\times\cdots\times m = m^n$ on $\mathcal{L}\otimes\mathcal{L}\otimes\cdots\otimes\mathcal{L} = \mathcal{L}^n$. The domain $\mathcal{L}^n$ of m^n is called the set of *Lebesgue measurable sets in* $\mathbb{R}^n$: $(\mathbb{R}^n, \mathcal{L}^n, m^n)$ (sometimes, we will restrict m^n to have domain $(\mathcal{B}_{\mathbb{R}})^n = \mathcal{B}_{\mathbb{R}^n}$). If it is unambiguous, we will usually write $(\mathbb{R}^n, \mathcal{L}^n, m)$, dropping the power on the m, and for the $n = 1$ case, we will often write $\int f(x)\, dx := \int f\, dm$.

The following 3 theorems are simply generalizations of previous properties of the Lebesgue measure. In the following, let $E = \prod_{i=1}^{n} E_i$ where $E_i \subseteq \mathbb{R}$. We will call each E_i the *sides* of E

Theorem 2.6.1: Simplifying outer-measure

Suppose $E \in \mathcal{L}^n$

1. $m(E) = \inf\{m(U) \mid E \supseteq U,\ U \text{ is open}\} = \sup\{m(K) \mid K \subseteq E,\ K \text{ is compact}\}$
2. $E = A_1 \cup N_1 = A_2 \setminus N_2$ where $A_1 \in F_\sigma$, $A_2 \in G_\delta$, and $m(N_1) = m(N_2) = 0$
3. If $m(E) < \infty$, for all $\epsilon > 0$, there is a finite disjoint collection of rectangles $\{R_i\}_{i=1}^n$ whose sides are intervals such that $m(E \Delta \cup_i^n R_i) < \epsilon$

Proof **:**
p. 70 in folland

Theorem 2.6.2: Completeness of $L^1(m)$

Let $f \in L^1(m)$ and let $\epsilon > 0$. Then there exists a simple function $\phi = \sum_{i=1}^n a_i \chi_R$ where each R_i is a product of intervals such that

$$\int |f - \phi| < \epsilon$$

and there is a continuous function that vanishes outside a bounded set such that

$$\int |f - g| < \epsilon$$

Proof **:**
Just like in theorem 2.3.3, approximate f by a simple function, then apply proposition 2.6.1(3) to approximate ϕ appropriately. Finally, approximate ϕ by continuous functions by applying the natural generalization of the argument in theorem 2.3.3

Theorem 2.6.3: Translation Invariance in $\mathbb{R}^n$

Let $(\mathbb{R}^n, \mathcal{L}^n, m)$ be a measure space. Then m is translation invariant, that is, for any $a \in \mathbb{R}^n$, if we define $\tau_a : \mathbb{R}^n \to \mathbb{R}^n$, $\tau_a(x) = a + x$, then

1. If $E \in \mathcal{L}^n$, then $\tau_a(E) \in \mathcal{L}^n$ and $m(E) = m(\tau_a(E))$
2. If $f : \mathbb{R}^n \to \mathbb{C}$ is Lebesgue measurable, then so is $f \circ \tau_a$. Furthermore, if either $f \geq 0$ or $f \in L^1(m)$, then $\int (f \circ \tau_a)\, dm = \int f\, dm$

Notice that we did not mention scaling. This is because scaling is not abstracted to linear transformations which requires more work to show how the scaling of a linear transformation affects the value, and so we will prove that later in this section.

***Proof* :**
Folland p.71

(here, some on cube approximation and Jordan content)

I'll for now simply state the effect of linear transformations and come back to this:

Theorem 2.6.4: Linear Transformation on Integrability

Suppose $T \in GL_n(\mathbb{R})$. Then:

1. If f is a Lebesgue measurable function on $\mathbb{R}^n$, so is $f \circ T$. If $f \geq 0$ or $f \in L^1(m)$, then:
$$\int f\, dm = |\det(T)| \int f \circ T\, dm$$
2. If $E \in \mathcal{L}^n$, then so is $T(E)$, and $m(T(E)) = |\det(T)| m(E)$

***Proof* :**
here

There is also the statement on how diffeomorphisms affect the integral.

Chapter 3

Signed Measure and Differentiation

In this chapter, we will take a close look at how to define measure's with respect to integrals. In particular, given the appropriate integrable function f, we will want to define a measure μ such that

$$\mu(A) = \int_A f$$

After defining some terms in section 3.1, we will focus in section 3.2 and 3.3 on how to find a function f given a measure μ so that $\mu(A) = \int_A f$. The last 2 sections then focus on the special case where the function f that we found can actually be the *derivative*. In the simple case where $X = \mathbb{R}$, then we can define:

$$F(x) = \int_{(-\infty,x]} f dx = \int_{-\infty}^{x} f dx$$

and ask ourselves what conditions does f or F requires so that $f = F'$ (or, $f = F'$ a.e.). This line of reasoning will ultimately lead to a generalisation of the Fundamental Theorem of Calculus, which will tell us when can we recover the function f given we know $\int_A f$ for all A.

3.1 Signed Measures

Definition 3.1.1: Signed Measure

Let $(X, \mathcal{M})$ be a measurable space. A *signed measure* on $(X, \mathcal{M})$ is a function: $\nu : \mathcal{M} \to [-\infty, \infty]$ such that:

1. $\nu(\emptyset) = 0$
2. ν has at most one value of $\pm\infty$
3. If $\{E_i\}_{i=1}^{\infty} \subseteq \mathcal{M}$ is a disjoint collection, then

$$\nu\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \nu(E_i)$$

   where if $\nu\left(\bigcup_{i=1}^{\infty} E_i\right) < \infty$, then $\sum_{i=1}^{\infty} \nu(E_i)$ converges absolutely.

Example 3.1: Signed Measures

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

1. Clearly, every measure on $\mathcal{M}$ is a signed measure. To distinguish between the two, we will often call measures *positive measures.*
2. Let μ_1 and μ_2 be two measure on $\mathcal{M}$ with at least one being finite. Then $\nu = \mu_1 - \mu_2$ is a signed measure.
3. If f is a $\mathcal{M}$-measurable function, then if at least one of $\int f^+$ or $\int f^-$ is finite,

$$\nu(E) = \int_E f$$

   is a signed measure. In this case, we call f the *extended μ-integrable function* (in contrast to $f \in L^1$ where both terms need to be finite)

The fact that it converges absolutely is to make sure that the order in which we sum the elements does not matter: If it did not converge absolutely, then it is possible to make the sum approach any value we want. Since $\mathcal{M}$ doesn't necessarily have an inheret order (in fact it almost never does), then having this dependence on the order of summing the elements is inconsistent with our notion of area (this is also why we only allow one of the direction to be infinite). It might seem that we must thus be careful on how we define our signed measure to make sure this condition holds, however we will soon show that every signed measrue can be decomposed into 2 measures. If the measure of all sets with positive measure and negative measure are both infinite ($\pm\infty$), then we have contradicted the definition of a signed measure. As just mentioned, all signed measures can be decomposed into two positive measures. We will spend this section proving that fact. In the next section, we will show that under appropriate conditions and choices of signed measure ν and positive meaure μ, we can find a measruable function f such that $\nu(E) = \int_E f d\mu$. This will essentially be the generalization of the derivative!

We start by building up to decomposing a signed measure into two positive meaures:

Proposition 3.1.1: Continuity from Above/Bellow for Signed Meaures

Let ν be a signed measure on $(X, \mathcal{M})$.

1. If $\{E_i\}$ is an increasing sequence in $\mathcal{M}$, then $\nu\left(\bigcup_i E_i\right) = \lim_{i\to\infty} \nu(E_i)$
2. If $\{E_i\}$ is an decreasing sequence in $\mathcal{M}$ where $\nu(E_1)$ is finite, then $\nu\left(\bigcap_i E_i\right) = \lim_{i\to\infty} \nu(E_i)$

Proof **:**
Very similar to proposition 1.2.1

For the next proposition, we introduce a quick new definition: a set $P \in \mathcal{M}$ is called *positive* (resp. *negative*) if for all $S \subseteq P$ such that $S \in \mathcal{M}$, then $\nu(S) \geq 0$ (resp. $\nu(S) \leq 0$)

Lemma 3.1.1

Any countable union of positive sets is a positive set

Proof **:**
Use proposition 3.1.1

Theorem 3.1.1: Hahn Decomposition Theorem

If ν is a signed measure on $(X, \mathcal{M})$, there exists a positive set P and a negative set N for ν $P \sqcup N = X$ and $P \cap N = \emptyset$. If P' and N' is another such pair, then $P \Delta P$ and $N \Delta N'$ is a null set for ν

Proof **:**
Let ν be a signed measure, so only ∞ or $-\infty$ can be achieved. If ∞ is achieved, take $-\nu$ so that ∞ is not achieved. Now, let m be the supremum of $\nu(E)$ where E ranges over all positive sets. Thus, there is a sequence of positive sets $\{P_i\}$ such that $\nu(P_i) \to m$. Let $P = \bigcup_i P_i$. Then by the previous lemma and proposition, P is a positive set and $\nu(P) = m$. Since ν does not achieve ∞, $m < \infty$. I claim that $N = X \setminus P$ is negative (thus completing the first assertion of the proof).

First, N cannot contain non-null positive sets. If $E \subseteq N$ was a positive set (i.e. $m(E) > 0$), then $E \cap P = \emptyset$ and $E \cup P$ is a positive set, so $m(E \cup P) = m(E) + m(P) > m$, which is a contradiction to m being the supremum.

Now, let's say there exists a $A \subseteq N$ such that $\nu(A) > 0$. We will reach a contradiction in using the following fact: since A cannot be positive, then there must exist a $B \subseteq A$ such that $\nu(B) < 0$. Then let $C = A \setminus B$ so that $\nu(C) > \nu(A)$ and $C \subseteq A$.

With this, we proceed in forming a contradiction: Define the following sequence $\{A_i\}$ and $\{n_i\}$ as follows:

1. Let n_1 be the smallest positive integer such that there exists a $A_1 \subseteq N$ where $\nu(A_1) > n_1^{-1}$ (this set exists since there is a $A \subseteq N$ where $\nu(A) > 0$

2. Let n_k be the smallest positive integer such that there exists a $A_k \subseteq A_{k-1}$ such that $\nu(A_{k-1}) > \nu(A_k) + n_k^{-1}$ (since A_{k-1} is positive, we use the earlier remark fact to find such a set)

From this sequence of positive decreasing sets, $A_1 \supseteq A_2 \supseteq \cdots$, let $A = \bigcap_i A_i$. Then by the fact that ν does not attain ∞ and the previous proposition:

$$\infty > \nu(A) = \lim \nu(A_i) > \sum_{i=1}^{\infty} n_i^{-1}$$

since the series converges, it must be that $n_i \to \infty$. Finally, once again take $B \subseteq A$ such that $\nu(A) > \nu(B) + n^{-1}$ for some positive integer n. Since $n_i \to \infty$, there exists an i sufficiently large so that $n < n_i$ and $B \subseteq A_i$, meaning we can replace n_i by n in the above construction. However, by construction, we picked the *minimum* integer every-time, and so this contradicts minimality! Therefore, it cannot be that there exists a set $A \subseteq N$ such that $\nu(A) > 0$, and so N is indeed negative.

If P' and N' where two other sets satisfying the decomposition condition, then we have that $P \setminus P' \subseteq P$ (since it's a subset) and $P \setminus P' \subseteq N'$ (since it is P minus all the positive set P') , so it is both positive and negative, which can only be possible if it is a null-set (similarly for $N \setminus N'$)

Using the Hahn decomposition, we can define a positive measure on P and N. Before proceeding to that theorem, we quickly give a name to the relation between the two measures that will be defined:

Definition 3.1.2: Mutually Singular Signed Measures

Let ν and μ be be two signed measure on $(X, \mathcal{M})$. Then they are called *mutually singular* or μ *is singular with respect to* ν if there exists a $E, F \in \mathcal{M}$ such that $E \cup F = X$, $E \cap F = \emptyset$, and:

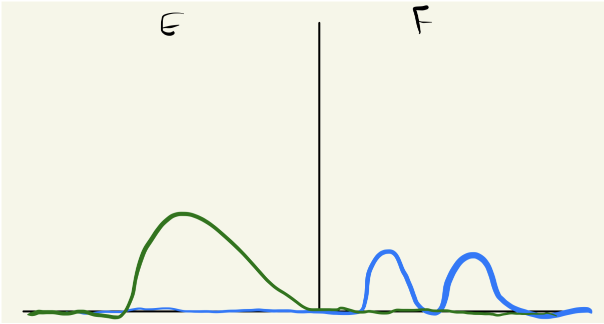
1. $\mu(E) = 0$

2. $\nu(F) = 0$

and we write mutually singular signed measures as:

$$\mu \perp \nu$$

In a sense, this means that these two sets are are meaningful on disjoint sets, in that sense they are "perpendicular" to each other. I like to keep an image like this in mind



Notice that it is possible that μ have zero values on F, and ν have zero values on E. They key is that they *must* have zero values on their respective sets.

Mutually singular measures can be applied more generally than in the context of decomposing a signed measure. For example, of δ_0 is the dirac-delta measure at 0 (i.e. $\delta(\{0\}) = 1$ and zero everywehre else), then $\delta_0 \perp \lambda$.

Theorem 3.1.2: Jordan Decomposition Theorem

If ν is a signed measure, there exists unique positive measure ν^+ and ν^- such that $\nu = \nu^+ - \nu^-$ and $\nu^+ \perp \nu^-$

Proof **:**
By theorem 3.1.1, let $X = P \cup N$ be the Hahn decomposition of X with respect to ν. Define $\nu^+(E) = \nu(E \cap P)$ and $\nu^-(E) = -\nu(E \cap N)$. Then by definition, $\nu = \nu^+ - \nu^-$ and $\nu^+ \perp \nu^-$.

If there was another decomposition $\nu = \mu^+ - \mu^-$ where $\mu^+ \perp \mu^-$, then we have $E, F \in \mathcal{M}$ such that $E \sqcup F = X$ and $E \cap F = \emptyset$, $\mu^+(F) = \mu^-(E) = 0$, so E, F is another Hahn decomposition for ν, and so $P \Delta E$ is null. Thus, for any $A \in \mathcal{M}$

$$\mu^+(A) = \mu^+(A \cap E) = \nu(A \cap E) = \nu(A \cap P) = \nu^+(A)$$

thus, $\mu^+ = \nu^+$, and similarly $\mu^- = \nu^-$, completing the proof.

Definition 3.1.3: Jordan Decomposition

Let ν be a signed measure on $(X, \mathcal{M})$. THen ν^+ and ν^- are caleld the *positive* and *negative variation* of ν, and $\nu = \nu^+ - \nu^-$ is called the *Jordan decomposition of* ν. The *total variation of* ν, denoted $|\nu|$ is defined as $|\nu| = \nu^+ + \nu^-$.

As an exercise, you should verify that $E \in \mathcal{M}$ is ν-null if and only if $|\nu|(E) = 0$, and $\nu \perp \mu$ if and only if $|\nu| \perp \mu$ if and only if $\nu^+ \perp \mu$ and $\nu^- \perp \mu$. With this definition, we will say that ν is finite (resp. σ-finite) if $|\nu|$ is finite (resp. σ-finite)

Finally, we can also get back to showing that any signed measure is of the form $\nu(E) = \int_E d\mu$ for appropriate μ, namely $\mu = |\nu|$ and $f = \chi_P - \chi_N$ for appropriate P and N from a Hahn decomposition. From this, we can define the integration with respect to a signed measure ν as follows:

Definition 3.1.4: Integration w.r.t. Signed Measures

Let ν be a signed measure, $\nu^+ - \nu^-$ be it's Hahn decomposition, and $f : X \to \mathbb{C}$ be a real function. Then we will say that f is *integrable with respect to the signed measure* ν if:

$$f \in L^1(\nu^+) \cap L^1(\nu^-) = L^1(\nu)$$

and define it's integral to be:

$$\int f d\nu = \int f d\nu^+ - \int f d\nu^- \qquad (f \in L^1(\nu))$$

3.2 Lebesgue-Radon-Nikodym Theorem

In this section, we show under what circumstances we can decompose a signed measure μ into $\nu + \rho$ where $\nu(A) = \int_A f$ and ρ is the "left-over" (also known as the singular part). We first define the "opposite" notion of mutually singular signed measures. The motivation for the vocabulary will come in section ref:HERE

Definition 3.2.1: Absolutely Continuous w.r.t. μ

Let ν be a signed measure on $(X, \mathcal{M})$ and μ is a positive measure[a] on $(X, \mathcal{M})$. We say that ν is *absolutely continous with respect to* μ when if $\mu(E) = 0$ (given $E \in \mathcal{M}$) then $\nu(E) = 0$, and we write is as:

$$\nu \ll \mu$$

[a]this can be extended to a signed measure by taking $\nu \ll \mu$ if and only if $\nu \ll |\mu|$, but we don't need this for this document

I like to remember the direction of the arrow by remembering $\mu(E) = 0 \Rightarrow \nu(E) = 0$, so the arrow is pointing towards ν, so we write the arrow's pointing towards ν: $\nu \ll \mu$. The ν is absolutely continuous with respect to μ, and so it's ν "w.r.t." μ: ν is on the left, the μ on the right.

Example 3.2: Absolutely Continuous Measures

1. If μ is any measure and 0 is the zero measure, it is always the case that $\mu \ll \mu$ and $0 \ll \mu$.
2. Given a measure μ and a function f, then $\nu(E) = \int_E f d\mu$ define's a measure, and clearly $\nu \ll \mu$, since if $\mu(E) = 0$, then $\int_E f d\mu = 0 = \nu(E)$.

In the following, We will show that all absolutely continuous measures can be "related" through an integral in this way.

It should be verified that $\nu \ll \mu$ if and only if $|\nu| \ll \mu$ if and only if $\nu^+ \ll \mu$ and $\mu^- \ll \mu$. Notice that absolute continuity is sort of the opposite of modular singularity in the sense that if $\nu \perp \mu$ and $\nu \ll \mu$, then it must be that $\nu = 0$. Note that $\mu \perp \nu$ does not imply $\mu(E) = 0 \Rightarrow \nu(E^c) = 0$: take ν to be the Lebesgue signed measure on $\mathbb{R}$, so that the positive and negative sets are $[0, \infty)$ and $(-\infty, 0)$ and experiment.

The following theorem should give you a taste on why the term "continuity" comes around when ν is finite. In particular, it should give you a feeling of uniformally continuous. Later on, we'll show that absolute continuity is stronger than uniform continuity, and so this feeling should make sense:

Theorem 3.2.1: Epsilon-Delta Absolute Continuity

Let ν be a finite signed meaure on $(X, \mathcal{M})$ and μ a positive measure on $(X, \mathcal{M})$. Then $\nu \ll \mu$ if and only if for every $\epsilon > 0$, there exists a $\delta > 0$ such that $\mu(E) < \delta \Rightarrow |\nu(E)| < \epsilon$.

***Proof* :**
Since $\nu \ll \mu$ if and only if $|\nu| \ll \mu$, and moreover $|\nu(E)| \leq |\nu|(E)$, it suffices to show it in the case where $\nu = |\nu|$ is positive.

($\Leftarrow$) Let the ϵ-δ definition hold so that for any $\epsilon > 0$, there exists a $\delta > 0$ such that for all $E \in \mathcal{M}$ that satisfy $\mu(E) < \delta$, then $\nu(E) < \epsilon$. If we pick all E such that $\mu(E) = 0$, then for any ϵ arbitrarily small, any δ will work, and so $\nu(E) < \epsilon$ for all ϵ, and so it must be that $\nu(E) = 0$

($\Rightarrow$) assume the contra-positive; that the ϵ-δ condition is not satisfied, meaning there exists a $\epsilon > 0$ such that for all $\delta > 0$, there is a E_δ such that $\mu(E_\delta) < \delta \Rightarrow \nu(E_\delta) \geq \epsilon$. In particular, for every $n \in \mathbb{N}$, $\mu(E_n) < 2^{-n} \Rightarrow \nu(E_n) \geq \epsilon$. let $F_k = \bigcup_{n=k}^{\infty} E_n$ and $F = \bigcap_k F_k$. Then

$$\mu(F_k) < \sum_{k=n}^{\infty} \frac{1}{2^n} = 2^{1-k}$$

and since $F \subseteq F_k$ for all F_k, it must be that $\mu(F) = 0$. However, by assumpiton $\nu(F_k) \geq \epsilon$ for all k. Since ν is finite, $\nu(F_1)$ is finite, so continuity from above implies

$$\nu(F) = \lim \nu(F_k) \geq \epsilon$$

and so $\nu \not\ll \mu$

An easy way to find a ν that is absolutey continuous with respect to μ is if we have a measure μ and a extended μ-integrable function f and define $\nu(E) = \int_E f \, d\mu$. If $f \in L^1$, then ν is finite, in which case we have the following corrollary:

Corollary 3.2.1: Absolute Continuity And L^1

If $f \in L^1(\mu)$, for every $\epsilon > 0$, there exists a $\delta > 0$ such that

$$\forall E \in \mathcal{M}, \; \mu(E) < \delta \; \Rightarrow \; \left| \int_E f \right| < \epsilon$$

***Proof* :**
apply the previous theorem on $\Re f$ and $\Im f$

The definition of ν with respect to $\int_E f d\mu$ is common enough that we define the following notation for it:

$$d\nu = f d\mu$$

This should give you the feeling of change of variables, and indeed the two concepts will be related. I like to remember this notation by thinking that if we add an integral with respect to E, $\int_E$ to both sides, and imagine there is the constant function in-front of the $d\nu$ on the left hand side, we get back our previous notation of $\nu(E) = \int_E f d\mu$.

We are now in a position to fully classify signed measures given positive measures. The way we will generlize this is to start by remember that an extended μ integrable function f has one part that allow's for an infinite value. Thus, we can think of the domain of f as being split between a signed measure and a positive measure. To avoid pathologies, we will work over σ-finite measures (as usual). We first prove some technical lemma before continuing to the main result:

Lemma 3.2.1

Let μ and ν be finite measures on $(X, \mathcal{M})$. Then either $\mu \perp \nu$, or there exists a $\epsilon > 0$ and $E \in \mathcal{M}$ such that $\mu(E) > 0$ and $\nu \geq \epsilon\mu$ on E (that is, ν and μ have domain $M_E = \{E \cap F \mid F \in \mathcal{M}\}$, and E is a positive set for $\nu - \epsilon\mu$)

Proof **:**
Let $X = P_n \sqcup N_n$ be the Hahn decomposition for $\nu - n^{-1}\mu$, and let $P = \bigcup_n^\infty P_n$ and $N = \bigcap_n^\infty N_n = P^c$. Then N is a negative set on $\nu - n^{-1}\mu$ for all n a.e., and

$$0 \leq \nu(N) \leq n^{-1}\mu(N)$$

for all n, so $\nu(N) = 0$. If $\mu(P) = 0$, then $\nu \perp \mu$. If $\mu(P) > 0$, then $\mu(P_n) > 0$ for some n, and P_n is a positive set for $\nu - n^{-1}\mu$

Lemma 3.2.2: Sum and Absolute Continuity

Suppose $\{\nu_i\}$ is a sequence of positive meaures. If $\nu_i \perp \mu$ for all i, then $\sum_i^\infty \nu_i \perp \mu$ and if $\nu_i \ll \mu$ for all i then $\sum_i^\infty \nu_i \ll \mu$

Proof **:**
Let $\nu_i \perp \mu$ for all i, so that $X = A_i \sqcup B_i$ where $\nu_i(A_i) = 0$ and $\mu(B_i) = 0$ for all i. Since $\mu(B_i) = 0$ for all i, then $\mu\left(\bigcup_i B_i\right) = 0$. Similarly, $\sum_i^n \nu_i\left(\bigcap_i A_i\right) = 0$ since $\cap_i A_i \subseteq A_i$ for each i. By some basic set theory

$$X = \left(\bigcap_i A_i\right) \sqcup \left(\bigcup_i B_i\right) \qquad \left(\bigcap_i A_i\right) \cap \left(\bigcup_i B_i\right) = 0$$

since each x is in either A_i or B_i and $A_i \sqcup B_i = X$ for each i. Therefore, we have:

$$\sum_i \nu_i \perp \mu$$

For the second statement, let's say $\nu_i \ll \mu$ for all i. This implies that:

$$\begin{aligned} \mu(E) = 0 &\Rightarrow \nu_1(E) = 0 \\ &\Rightarrow \nu_2(E) = 0 \\ &\vdots \\ &\Rightarrow \nu_n(E) = 0 \\ &\vdots \end{aligned}$$

meaning if $\mu(E) = 0$, then each $\nu_i(E) = 0$. But then $\sum_i \nu_i(E) = 0$, and so

$$\sum_i^\infty \nu_i \ll \mu$$

as we sought to show.

Theorem 3.2.2: Lebesgue-Radon-Nikodym Theorem

Let ν be a signed σ-finite measure on $(X, \mathcal{M})$ and μ be a positive σ-finite measure on $(X, \mathcal{M})$. Then, there exists unique σ-finite measure λ and ρ on $(X, \mathcal{M})$ such that

$$\lambda \perp \mu, \qquad \rho \ll \mu \quad \text{and} \quad \nu = \lambda + \rho$$

Furthermore, there is an extended μ-integrable function f such that $d\rho = f d\mu$, and any two such functions are equal μ-a.e. and so we get:

$$\nu(A) = \int_A f d\mu + \lambda(A)$$

In other words, λ and ρ "split" the measure μ based on where μ is zero or not , and together they can be combined to make ν. Notice that there is no prior relation between the signed measure ν and the positive measure μ. This means that we can replace μ with μ', and find $\lambda' \perp \mu'$ and $\rho' \ll \mu'$, and still get $\nu = \lambda' + \rho'$, and so this decomposition of a signed measure with respect to a measure μ heavily depends on μ. This decomposition might make more sense if we conider the special case where $\nu \ll \mu$, in which case we ar left with $\nu = \rho$. The last statement of the theorem tell's us that:

$$d\rho = f d\mu \qquad \text{which is shorthand for} \qquad \rho(E) = \int_E f d\mu$$

Notice that this looks like the change of varaibles we were talking about earlier. Thus, in this case, this theorem is ultimately about a "change of variables" into the μ measure!

Finally, notice that this "change of variables" is show's that the measure ν is sort of the "integral" of the measure of f, and so in a sense f is thus the "derivative" of ν since you "get back" ν by integrating. In other words, this should start giving you Fundamental Theorem of Calculus vibes. However, be weary that the function f is not [always] obtained by taking the limit of some function F, i.e. does does not always exist a function F such that $f(a) = F'(a) = \lim_{x \to a} \frac{F(x+a) - F(a)}{x-a}$. However, in the next seciton, we will see that in some sepcial cases we *can* define a function F (which we obtain through ν) where $F' = f$. Thus, the function f can be thought of as a sort of generalization of the notion of derivative.

***Proof* :**
We will break down this proof into 3 cases: when ν and μ are finite positive measures, when ν and μ are σ-finite positive meaures, and finally when ν is a signed measure.

For the first case, let:

$$\mathcal{F} = \left\{ f : X \to [0, \infty] \;\middle|\; \int_E f\, d\mu \leq \nu(E) = \int_E 1 d\nu \text{ for all } E \in \mathcal{M} \right\}$$

Remember that we can define a measure through the integral, so what we are asking for is if we can find all possible functions which can define us a measure that will be smaller than ν. We will end up taking the largest such function f, and labelling $d\rho = f d\mu$, and $d\lambda = d\nu - d\rho$. We can already see that $\rho \ll \mu$. What's left to show is that such an f exists, that $\lambda \perp \mu$ and that the result is unique.

Notice also how the construction can make you think of change of variables. Let's say for a moment that $\int_E f d\mu = \int_E 1 d\nu$. Then it is as though f is correcting the value so that the values are equal. This is actually what we are looking for, so we look for all f such that $\int_E f d\mu \leq \int_E 1 d\nu$ and take the supremum.

First, $\mathcal{F}$ is nonempty since $0 \in \mathcal{F}$. Thus, let $a = \sup \left\{ \int f\, d\mu \;\middle|\; f \in \mathcal{F} \right\}$ (notice that $a \leq \nu(X) < \infty$ since ν is finite), and choose a sequence $\{f_n\} \subseteq \mathcal{F}$ such that $\int f_n d\mu \to a$ and let $f = \sup_n f_n$. We would like to show that $a = \int f$. To that end, notice that if $f, g \in \mathcal{F}$, then $\max(f, g) \in \mathcal{F}$: if $A = \{x \mid f(x) > g(x)\}$, for any $E \in \mathcal{M}$:

$$\int_E \max(f, g) d\mu = \int_{E\cap A} f d\mu + \int_{E \setminus A} g d\mu \leq \nu(E \cap A) + \nu(E \setminus A) = \nu(E)$$

Let $g_n = \max(f_1, f_2, ..., f_n)$. Then $g_n \in \mathcal{F}$, g_n increases pointwise to f, and $\int g_n\, d\mu \geq \int f_n\, d\mu$ by monotonicity. Thus, since a is the supremum, it must be that $\lim \int g_n\, d\mu = a$, and so by the MCT

$$a = \lim \int g_n d\mu = \int \lim g_n d\mu = \int f$$

and so $f \in \mathcal{F}$ (i.e. the supremum is "achieved") and $\int f\, d\mu = a$. In particular, $f < \infty$ a.e., so we may take f to be real-valued everywhere. Let $d\rho = f d\mu$ (we will soon show this $d\rho$ satisfies the conditions of the theorem).

We know define the measure λ which is singular with respect to μ. Let $d\lambda = d\nu - f d\mu$ (i.e. $\lambda(E) = \nu(E) - \int_E f d\mu$). Then $d\lambda$ is positive, since $f \in \mathcal{F}$, and is singular with respect to μ. It it weren't, then by lemma 3.2.1, there exists a $\epsilon > 0$ and $E \in \mathcal{M}$ such that $\mu(E) > 0$ and $\lambda \geq \epsilon\mu$ on E. Then $\epsilon\chi_E d\mu \leq d\lambda = d\nu = f d\mu$, or rearranging $(f + \epsilon\chi_E)d\mu \leq d\nu$. Since $f \in \mathcal{F}$ and the $\epsilon\chi_E$ only effects the value on E, we just showed that $f + \epsilon\chi_E \in \mathcal{F}$. However, $\int f + \epsilon\chi_E d\mu = a + \epsilon\mu(E) > a$, contradicting the maximality of a. Hence it must be that $\lambda \perp \mu$.

Therfore, we have that $d\nu = f d\mu + d\lambda$ where $\lambda \perp \mu$. Recalling we set $d\rho = f d\mu$, we get that $d\nu = d\rho + d\lambda$, and by definition of ρ we have $\rho \ll \mu$ (see example 3.2). Converting into our more usual notation, we see that $\nu(E) = \rho(E) + \lambda(E)$ for all E, and so $\nu = \rho + \lambda$ with $\lambda \perp \mu$ and $\rho \ll \mu$. It remains to show that this decomposition is unique. To that end, let's say $d\nu = d\lambda' + f' d\mu$. Then

$d\lambda - d\lambda' = (f - f')d\mu$. Then (show that $d\lambda - d\lambda' \perp \mu$) and $(f' - f)d\mu \ll d\mu$, and since both sides must be equal, this forces that $d\lambda - d\lambda' = (f' - f)d\mu = 0$. Thus, $\lambda = \lambda'$, and by proposition 2.3.2, $f = f'$ μ-a.e. Thus, if ν and μ are of finite measure, the theorem is done.

Now let ν and μ be σ-finite measures so that X is a countable union of μ-finite sets and a countable union of ν-finite sets. By refining the two by taking intersections, we get a new collection $\{A_i\} \subseteq \mathcal{M}$ $\mu(A_i)$ and $\nu(A_i)$ is finite for all i and $X = \bigcup_i^\infty A_i$. We can define

$$\mu_i(E) = \mu(E \cap A_i) \qquad \nu_i(E) = \nu(E \cap A_i)$$

Then by what we've just shown, we have the decomposition:

$$d\nu_i = d\lambda_i - f_i d\mu$$

where $\lambda_i \perp \mu_i$. Now $\mu_i(A_i^c) = \nu_i(A_i^c) = 0$, so $\int_{A_i^c} f d\mu = 0$ and

$$\lambda_i(A_i^c) = \nu_i(A_i^c) - \int_{A_i^c} f d\mu = 0$$

for our construciton, let f_i be 0 on A_i^c. Now, let $\lambda = \sum_{i=1}^\infty \lambda_i$ and $f = \sum_{i=1}^\infty f_i$. Then by lemma 3.2.2:

$$d\nu = d\lambda - f d\mu$$

and $d\lambda$ and $f d\mu$ are σ-finite, as desired.

Finally, taking the case of ν being a signed measure, using the Jordan decomposition, apply the preceeding argument to ν^+ and $\nu-$, compelting the proof

Since the decomposition exists, it is given a name:

Definition 3.2.2: Lebesgue Decomposition

Let ν be a signed σ-finite measure and μ a positive σ-finite measure. Then the decomposition

$$\nu = \lambda + \rho \qquad \lambda \perp \mu \quad \rho \ll \mu$$

is called the *Lebesgue decomposition of ν with respect to μ.*

if $\nu \ll \mu$, then the Lebesgue-Radon-Nikodym theorem tell's us that $d\nu = f d\mu$. If the extra condition that $\nu \ll \mu$ is added, the theorem is often called *Radon-Nikodym's Theorem*, and f is given a special name:

Definition 3.2.3: Radon-Nikodym Derivative

Let $\nu \ll \mu$ and f be the function obtain from the Lebesgue Decomposition such that $d\nu = f d\mu$. Then f is called the *Radon-Nikodym derivative of ν with respect to μ.* We often dentoe f by:

$$\frac{d\nu}{d\mu} := f$$

As a consequence of this notation, we get the following representatin of $\nu(E) = \rho(E)$ that looks like a fraction cancellaton:

$$d\nu = \frac{d\nu}{d\mu} d\nu$$

Strictly speaking, $\frac{d\nu}{d\mu}$ should really be the class of function's that are equal to f μ-a.e., however just like with L^1, we often abuse notation and pick some function in this class when we use this notation). Most of the time, we will be working with the case of $\nu \ll \mu$. In Distribution theory, we will be studying the case where $\nu \not\ll \mu$.

Writting f in this form is very fitting, as in f respects the same properties as the derivative! Once we reach the special case of the measures being Lebesggue measures and the functions being differnetiable functions, then f will in fact be the derivative of f.

Without yet doing conrete computations of f, if we take the Radon-Nikodym of $\nu_1 + \nu_2$ where $\nu_1 + \nu_1 \ll \mu$ then we have

$$\frac{d(\nu_1 + \nu_2)}{d\mu} = \frac{d\nu_1}{d\mu} + \frac{d\nu_1}{d\mu}$$

to see this, let $d(\nu_1 + \nu_2)/d\mu = f$. Since $\nu_1 + \nu_2 \ll \mu$, $\nu_1 \ll \mu$ and $\nu_2 \ll \mu$. Then taking $d\nu_1 = f_1 d\mu$ and $d\nu_2 = f_2 d\mu$, then:

$$\begin{aligned}
&= \int \frac{d\nu_1}{d\mu} + \frac{d\nu_2}{d\mu} d\mu \\
&= \int \frac{d\nu_1}{d\mu} d\mu + \int \frac{d\nu_2}{d\mu} d\mu \\
&= \int 1 d\nu_1 + \int 1 d\nu_2 \\
&= \nu_1(E) + \nu_2(E) \\
&= (\nu_1 + \nu_2)(E) \\
&= \int 1 d(\nu_1 + \nu_2) \\
&= \int_E \frac{d(\nu_1 + \nu_2)}{d\mu} d\mu
\end{aligned}$$

and since the two integrals are equal over every measurable set, we get:

$$\frac{d\nu_1}{d\mu} + \frac{d\nu_2}{d\mu} = \frac{d(\nu_1 + \nu_2)}{d\mu}$$

which is the same property of differentiable functions. In particular, notice that Similarly, we have the following chain-rule property:

Proposition 3.2.1: Chain Rule For Measures

Let ν be a signed σ-finite measure on $(X, \mathcal{M})$, μ and λ positive σ-finite measures on $(X, \mathcal{M})$ such that $\nu \ll \mu$ and $\mu \ll \lambda$. Then:

1. If $g \in L^1(\nu)$, then $g(d\nu/d\mu) \in L^1(\mu)$ and
$$\int g d\nu = \int g(d\nu/d\mu)d\mu$$
2. Absolute continuity, $\ll$, is transitive: $\nu \ll \lambda$ and
$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda}$$

Proof **:**

1. This proof uses (what now should be) the standard measure-theory way to generalize results. First, we can consider ν^+ and ν^- seperately, so without loss of generality let's say $\nu \geq 0$. If we let $g = \chi_E$, then:
$$\int g\frac{d\nu}{d\mu}d\mu = \int \int \chi_E f d\mu = \int_E f d\mu$$
which is true by definition of f. Thus, it is true for any simple functions since the integral and the construction is linear, and so it is also true for non-negative function's by the MCT, and so it is true for functions in $L^1(\nu)$ by linearity again.
2. If we take ν to be μ and μ to be λ and set $g = \chi_E \frac{d\nu}{d\mu}$, we get:
$$\nu(E) = \int_E \frac{d\nu}{d\mu}d\mu = \int_E \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda}d\lambda$$
for all $E \in \mathcal{M}$, including X, and so by proposition 2.3.2
$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda} \quad \lambda\text{-a.e.}$$
as we sought to show

As a consequence, if $\mu \ll \lambda$ and $\lambda \ll \mu$, then $(d\lambda/d\mu)(d\mu/d\lambda) = 1$ λ-a.e. and μ-a.e. Note that all of this is for the special case of $\nu \ll \mu$. If this is not the case (say $\nu \ll \mu$) we get a different notion of "$d\nu/d\mu$ (the example of $\delta_0 \perp \lambda$ has a "Radon-Nikodym derivative" of the dirac delta function).

Note that σ-finiteness is necessary:

Example 3.3: Withiout σ-Finiteness

Let $X = [0,1]$, $\mathcal{M} = \mathcal{B}_{[0,1]}$, m the Lebesgue measure and μ the counting measure on $\mathcal{M}$

1. $m \ll \mu$ but $dm \neq f d\mu$ for any f

Proof. First, if $\mu(E) = 0$, then $|E| = 0$, which we have shown that this implies $m(E) = 0$, and so $m \ll \mu$.

However, there does not exist an f such that $m(E) = \int_E f d\mu$. To see this, consider any $p \in [0, 1]$. Then

$$0 = m(\{p\}) = \int_{\{p\}} f d\mu = \int_{[0,1]} \chi_{\{p\}} f d\mu \stackrel{!}{=} f(p)\mu(\{p\}) = f(p)$$

where $\stackrel{!}{=}$ comes from the fact that $\chi_{\{p\}} f$ is in fact a simple funciton (being of the form $\sum_i c_i \chi_{E_i}$ with $i = 1$, $E_i = \{p\}$ and $c_i = f(p)$), and so matches the definition of the integral for simple functions (which we have shown in class is the value of the usual integl). Hence, we have that $f(p) = 0$ for all $p \in [0, 1]$. However, this implies that $f \equiv 0$. This leads to a contardiction, since:

$$1 = m([0, 1]) = \int_{[0,1]} f d\mu = \int_{[0,1]} 0 d\mu = 0$$

but $1 \neq 0$ □

Next, show that μ has no Lebesgue decomposition with respect to m

Proof. For the sake of contradiction, let $\mu = \lambda + \rho$ where $\lambda \perp m$ and $\rho \ll m$. Then for any $c \in [0, 1]$

$$\mu(\{c\}) = \lambda(\{c\}) = 1$$

since $m(\{c\}) = 0$ and $\rho \ll m$, $\rho(\{c\}) = 0$. But then, λ is never null on non-zero sets by monotonicity, implying that m must *always* be null. However, this is not true since m is the Lebesgue measure – a contradiction. Thus, μ does not have such a decomposition with respect to m. □

Before we move on to our last result of the section, I want to take a moment to take about generalizations of the Radon-Nikodym theorem. In the theorem, we assumed that the measures must be σ-finite. However, in contrast to previous example where without σ-finiteness the theorem fails badly, we seem to have these assumption because the generalization requires more build-up. In particular, we would have to define the notion of beign *truly continuous*. If μ is σ-finite, being absolutely continuous and truly continuous are equivalent. The statemnt of the Radon-Nikodym can be reformulated in terms of truly continuous meaures, and I found a stack-exchagne post on this:

`https://mathoverflow.net/questions/258936/radon-nikodym-theorem-for-non-sigma-finite-measures`

Finally, we conclude with this final proposition we'll use later on:

Proposition 3.2.2

Let $\mu_1, ..., \mu_n$ be measures on $(X, \mathcal{M})$. Then there exists a measure μ (namely $\sum_i \mu_i$) such that $\mu_i \ll \mu$

Proof **:**
exercise

3.3 Complex Measure

We know show how we can generalize the results when the codomain is $\mathbb{C}$. The reason to define complex measure is generalize the notion of the Radon-Nikodym derivtaive to the complex case so that we may define complex results. Most of the work we have done before transaltes one-to-one for the complex measures, with the exception of defining the absolute variation which we will sepdn most of this section covering.

Definition 3.3.1: Complex Measure

Let $(X, \mathcal{M})$ be a measure space. A function $\nu : \mathcal{M} \to \mathbb{C}$ is called a *complex measure* if

1. $\nu(\emptyset) = 0$
2. If $\{E_i\} \subseteq \mathcal{M}$ is a pair-wise disjoint sequence, then if $\sum_i^\infty \nu(E_i)$ converges absolutely, then
$$\nu\left(\bigcup_i^\infty E_i\right) = \sum_i^\infty \nu(E_i)$$

Just like for signed measures, the absolute convergence is so that the order in which we measure the sets does not matter.

Notice that the sequence $X, \emptyset, \emptyset, ...$ is a valid disjoint sequence, and so $\sum_i \nu(X_i) = \nu(X)$ must converge absolutely, and so ν must be a finite measure. Thus, a measure is a complex measure if it is finite. Another way of saying it is that if μ is a positive measure and $f \in L^1(\mu)$ (recall f is a complex function), then $f d\mu$ is also a complex measure. As is usually done with complex valued function, we define the signed measures $\nu_r(E) = \pi_r(\nu(E))$ and $\nu_i(E) = \frac{\pi_i(\nu(E))}{i}$ to take on the real and complex values of ν, so ν_i and ν_r are *finite* signed measures and $\nu = \nu_r + i\nu_i$. As a consequence, the range of ν is a bounded subset of $\mathbb{C}$.

Many of the previous concepts we've seen naturally generalize to the complex case:

1. We define $L^1(\nu)$ to be $L^1(\nu_r) \cap L^1(\nu_i)$
2. For $f \in L^1(\nu)$, set $\int f d\nu := \int f d\nu_r + i \int f d\nu_i$
3. If ν and μ are complex measures, then we that $\mu \perp \nu$ if $\nu_a \perp \nu_b$ for all $a, b \in \{r, i\}$.
4. If λ is a positive measure, we say $\nu \ll \lambda$ if $\nu_r \ll \lambda$ and $\nu_i \ll \lambda$
5. The Lebesgue-Radon-Nikodym Theroem generlizes to complex measures nicely; you would apply theorem 3.2.2 to the real and imaginary part seperately. To make sure there is no ambiguity in understanding the generalization:

   Let ν be a complex measure and μ a σ-finite measure on $(X, \mathcal{M})$. Then there exists a complex measure λ and a $f \in L^1(\mu)$ such that
   $$d\nu = d\lambda + f d\mu \qquad \lambda \perp \mu$$

Furthermore, if there existed another λ' such that $\lambda' \perp \mu$ and $d\nu = d\lambda' + f' d\mu$, then $\lambda = \lambda'$ μ-a.e. and $f = f'$ μ-a.e.

If $\nu \ll \mu$, then we denote f a $\frac{d\nu}{d\mu}$

We next want to generalize the notion of *total variation* to the complex case. In the signed measure case, this was an easy task since $\mathbb{R}$ has an inherent order. In the complex case, this is not so straight forward: Given that a surface $S \subseteq \mathbb{C}$ is the image of some complex measure $\nu(E) = S$, what does it mean to capture the "total-variation" on that surface? There is more than one sensible answer to this question, namely since there is no "natural order" to $\mathbb{C}$, and so it would be nice to have succint way to capture it's *total variation*. There is in fact a nice and concise way of defining this which also generalizes the previous way of looking at total variation that captures the idea in 1 formula, and so we will build-up to it.

First, let ν be a complex measure and let $\mu = |\nu_r| + |\nu_i|$. Next, notice that $\nu \ll \mu$, since if $\mu(E) = 0$, then it must be that both $|\nu_r|$ and $|\nu_i|$ are zero, and so $\nu_r(E) = \nu_i(E) = 0$, and so $\nu(E) = 0$. Therefore, the Lebesgue decomposition of ν with respect to μ gives us a f such that $d\nu = f d\mu$. As an example of such an f, let's say that ν was simply a finite signed measure (which is a special case of a complex measure), so $\mu = |\nu|$. Then $f = (\chi_P - \chi_N)$ and:

$$d\nu = (\chi_P - \chi_N) d\mu = (\chi_P - \chi_N) d|\nu|$$

Which is the Jordan decomposition theorem we have shown earlier! Now, notice that $|\chi_P - \chi_N| = 1$ so that we can write:

$$d|\nu| = |\chi_P - \chi_N| d\mu \quad \text{which is shorthand for} \quad |\nu|(E) = \int_E |\chi_P - \chi_N| d\mu$$

If we re-write the last equation as $d|\nu| = |f| d\mu$, then we have that the measure $|\nu|$ is defiend by the absolute value of the measure defined by the the derivative with respect to $|\nu|$. If we come back to the case of ν being a general complex measure, then we would get $\mu = |\nu_r| + |\nu_i|$, giving us a less obvious, but certainly existent decomposition $d|\nu| = |f| d\mu$ (i.e. $|\nu|(E) = \int_E |f| d\mu$). Thus, it seems that this propety somehow "captures" the idea of total variation. We chose a particular μ for both decomposition, however it turns out that this property we have stated is independent of choice of μ up to an appopraite zero-measure set, and so we will define total-variation with respect to the following property:

Definition 3.3.2: Total Varitaion And Complex Differentiation

Let ν be a compelx measure. Then $|\nu|$ is defined such that when $d\nu = f d\mu$ where μ is a positive measure, then $d|\nu| = |f| d\mu$

What we have shown with the μ we saw earlier is that $d\nu$ is always decomposable into $d\nu = f d\mu$. We must now show that $|\nu|$ is in-fact independent of choice of μ (i.e. any choice will get the same result up zero an appropraite zero-meaure set) given μ has the property of the definition. Let's say $d\nu = f_1 d\mu_1 = f_2 d\mu_2$ satisfy the total-variation property. Let $\rho = \mu_1 + \mu_2$. By proposition 3.2.1, we have that:

$$f_1 \frac{d\mu_1}{d\rho} d\rho = d\nu = f_2 \frac{d\mu_2}{d\rho} d\rho$$

Thus $f_1 \frac{d\mu_1}{d\rho} = f_2 \frac{d\mu_2}{d\rho}$ ρ-a.e. Since $\frac{d\mu_1}{d\rho}$ and $\frac{d\mu_2}{d\rho}$ are both non-negative, we have:

$$|f_1| \frac{d\mu_1}{d\rho} d\rho = \left| f_1 \frac{d\mu_1}{d\rho} d\rho \right| = \left| f_2 \frac{d\mu_2}{d\rho} d\rho \right| = |f_2| \frac{d\mu_2}{d\rho} d\rho \quad \rho\text{-a.e.}$$

Thus:

$$|f_1|d\mu_1 = |f_1|\frac{d\mu_1}{d\rho}d\rho = |f_2|\frac{d\mu_2}{d\rho}d\rho = |f_2|d\mu_2$$

and so, ν is independent of choice of μ. We next give some useful properties of total variation:

Proposition 3.3.1: Properties Of Total Variation

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

1. $|\nu(E)| \leq |\nu|(E)$ for all $E \in \mathcal{M}$
2. $\nu \ll |\nu|$ and $\frac{d\nu}{d|\nu|}$ has absolute value 1 $|\nu|$-a.e.
3. $L^1(\nu) = L^1(|\nu|)$ and if $f \in L^1(\nu)$, then $|\int f d\nu| \leq \int |f| d|\nu|$

***Proof* :**
Pick f such that $d\nu = f d\mu$ (as we've shown exists when defining $|\nu|$). Then:

$$|\nu(E)| = \left|\int_E f d\mu\right| \leq \int_E |f| d\mu = |\nu|(E)$$

showing the first claim, and also showing that $\nu \ll |\nu|$. For the next claim, let $g = \frac{d\nu}{d|\nu|}$ so

$$f d\mu = d\nu = g d|\nu|$$

so $g|f| = f$ μ-a.e. and so $|\nu|$-a.e. Since $|f| > 0$ $|\nu|$-a.e. as well, it must be that $g = 1$ $|\nu|$-a.e. Next, (left as an exercise by Folland)

The next property we introduce shows that it interacts well with the triangle inequality

Proposition 3.3.2: Triangle Inequality Of Total Variation

Let ν_1 and ν_2 be complex measures on $(X, \mathcal{M})$. Then

$$|\nu_1 + \nu_2| \leq |\nu_1| + |\nu_2|$$

***Proof* :**
By proposition 3.2.2, we can write $\nu_1 = f_1 d\mu$ and $\nu_2 = f_2 d\mu$ for the same μ. Then:

$$d|\nu_1 + \nu_2| = |f_1 + f_2| d\mu \leq |f_1| d\mu + |f_2| d\mu = d|\nu_1| + d|\nu_2|$$

as we sought to show

Finally, though this characterization of total variation is a good definition, it can be hard to work with. The following gives an equivalent definition of total variation that is easier computationally:

Proposition 3.3.3: Equivalence Of Total Variation

Let ν be a complex measure on $(X, \mathcal{M})$. If $E \in \mathcal{M}$, define

1. $\mu_1(E) = \sup\{\sum_{i=1}^n |\nu(E_i)| \mid n \in \mathbb{N},\ E_1, ..., E_n \text{ disjiont},\ E = \bigsqcup_{i=1}^n E_i\}$
2. $\mu_2(E) = \sup\{\sum_{i=1}^\infty |\nu(E_i)| \mid E_1, ... \text{ disjiont},\ E = \bigsqcup_{i=1}^\infty E_i\}$
3. $\mu_3(E) = \sup\left\{\left|\int_E f d\nu\right| \mid |f| \leq 1\right\}$

Then $\mu_1 = \mu_2 = \mu_3 = |\nu|$

Proof **:**
Throughout this proof, let:

$$S_1^E = \left\{\sum_{i=1}^n |\nu(E_i)| \;\middle|\; n \in \mathbb{N},\ E_1, ..., E_n \text{ disjiont},\ E = \bigsqcup_{i=1}^n E_i\right\}$$

and

$$S_2^E = \left\{\sum_{i=1}^\infty |\nu(E_i)| \;\middle|\; E_1, ... \text{ disjiont},\ E = \bigsqcup_{i=1}^\infty E_i\right\}$$

($\mu_1 \leq \mu_2$) Notice that $S_1^E \subseteq S_2^E$ (since for any finite disjoint sequence, we can extend it to an infinite disjoint sequence by letting $E_i = \emptyset$ for the rest of the terms), we get $\sup S_1^E \leq \sup S_2^E$, and since this is true for all E, $\mu_1 \leq \mu_2$

($\mu_2 \leq \mu_3$) In the process we will use the fact that there exists an f such that:

$$d\nu = f d|\nu|$$

By proposition 3.3.1, $\left|\frac{d\nu}{d|\nu|}\right| = |f| = 1$ $|\nu|$-a,e, and so $|f| \leq 1$ $|\nu|$-a.e. For the $|\nu|$-zero-measure points that are not less than or equal to 1, we can simply choose them to be equal to 0. This also implies that $|\overline{f}| \leq 1$. With this information we get that:

$$\overline{f} d\nu = \overline{f} f d|\nu| = |f|^2 d|\nu| = 1 d|\nu| = d|\nu| \quad |\nu|\text{-a.e.}$$

and so:

$$\begin{aligned}
\sum_i^\infty |\nu(E_i)| &\leq \sum_i^\infty |\nu|(E_i) \\
&= \int_E d|\nu| \\
&= \int_E \overline{f} d\nu \\
&\leq \left|\int_E \overline{f} d\nu\right|
\end{aligned}$$

and since $|f| \leq 1$, this is an element of $\left\{\left|\int_E f d\nu\right| \mid |f| \leq 1\right\}$, and since this was true for arbitrary E, we get: $\mu_2 \leq \mu_3$

($\mu_3 = |\nu|$) We'll first show $\mu_3 \leq |\nu|$. By the definition of the supremum, for all $\epsilon > 0$, there exists an f where $|f| \leq 1$ such that

$$\begin{aligned}
\mu_3 &\leq \left|\int_E f d\nu\right| + \epsilon \\
&\leq \int_E |f| d|\nu| + \epsilon && \text{prop. 3.13[c]} \\
&\leq \int_E d|\nu| + \epsilon && |f| \leq 1 \\
&\leq |\nu|(E) + \epsilon
\end{aligned}$$

and as $\epsilon \to 0$, we have that $\mu_3 \leq |\nu|(E)$. Going the other way, we can again define $\frac{d\nu}{d|\nu|}$, which by proposition 3.13 its absolute value is 1 $|\nu|$-a.e. Let $\overline{f} = \frac{d\nu}{d|\nu|}$ so that $1 = |f|^2 = f\overline{f}$ $|\nu|$-a.e. If f ever has value greater than 1, we can simply redefine them to be 0 since those value have measure 0. Then:

$$\begin{aligned}
|\nu|(E) &= \int_E d|\nu| \\
&= \int_E 1 d|\nu| \\
&= \int_E |f|^2 d|\nu| \\
&= \int_E |f| d|\nu| && \text{since } |f| = 1 \ |\nu|\text{-a.e.} \\
&= \int_E f\overline{f} d|\nu| \\
&= \int_E f \frac{d\nu}{d|\nu|} d|\nu| \\
&= \int_E f d\nu && \text{prop. 3.9} \\
&\leq \left|\int_E f d\nu\right|
\end{aligned}$$

and since $|f| \leq 1$,

$$\left|\int_E f d\nu|\right| \in \left\{\left|\int_E f d\nu\right| \;\middle|\; |f| \leq 1\right\}$$

and since this is true for arbitrary E:

$$|\nu| \leq \mu_3$$

and so, combining the two inequalities, we get $\mu_3 = |\nu|$

($\mu_3 \leq \mu_1$) We will do as the hint suggets and use simple functions to approximate f, in particular we will use the standard representation from theorem 2.10, let's label the increasing sequence $\{\phi_k\}_{k=1}^{\infty}$. To take advantage of these functions, first break up ν into positive meaures, $\nu_r^+, \nu_i^+, \nu_r^-, \nu_i^-$ so that:

$$\left|\int_E f d\nu\right| = \left|\int_E f d\nu_r^+ - \int_E d\nu_r^- + i\left(\int_E f d\nu_i^+ - \int_E f d\nu_i^-\right)\right|$$

Then by the definition of integral, for all $\epsilon/4 > 0$, there must exists an $n \in \mathbb{N}$ such that

$$\begin{aligned}
&= \left|\int_E f d\nu_r^+ - \int_E f d\nu_r^- + i\left(\int_E f d\nu_i^+ - \int_e f d\nu_i^-\right)\right| \\
&\leq \left|\int_E \phi_n d\nu_r^+ + \frac{\epsilon}{4} - \int_E \phi_n d\nu_r^- - \frac{\epsilon}{4} + i\left(\int_E \phi_n d\nu_i^+ + \frac{\epsilon}{4} - \int_e \phi_n d\nu_i^- - \frac{\epsilon}{4}\right)\right| \\
&\leq \left|\int_E \phi_n d\nu_r^+ - \int_E \phi_n d\nu_r^- + i\left(\int_E \phi_n d\nu_i^+ - \int_E \phi_n d\nu_i^-\right)\right| + \epsilon \\
&= \left|\int_E \phi_n d\nu\right| + \epsilon
\end{aligned}$$

Then by the definition of a simple function, $\phi_n = \sum_{i=1}^n c_i \chi_{F_i}$ for appropriate F_i and c_i. Note that $c_i \leq 1$ since $|\phi_n| \leq |f| \leq 1$. Also, note that $\chi_E \chi_{F_i} = \chi_{E \cap F_i}$. Then by definition of an integral:

$$\left|\int_E \phi_n d\nu\right| + \epsilon = \left|\sum_{i=1}^n c_i \nu(F_i \cap E)\right| + \epsilon \leq \sum_{i=1}^n |c_i| \, |\nu(F_i \cap E)| + \epsilon$$

This is almost the result we need. The problem is that the collection $\{F_i \cap E\}$ is not guaranteed to cover E since there might be portions that are yet to be "precise" enough to bring it into the approximation. Since we only need to get bigger, we can solve this by adding $\bigcap_{i=1}^n F_i^c \cap E$, which is disjoint of every $F_i \cap E$, and so:

$$\begin{aligned}
\sum_{i=1}^n |c_i| \, |\nu(F_i \cap E)| + \epsilon &\leq \sum_{i=1}^n |c_i| \, |\nu(F_i \cap E)| + \left|\nu\left(\bigcap_i^n F_i^c \cap E\right)\right| + \epsilon \\
&\leq \sum_{i=1}^n |\nu(F_i \cap E)| + \left|\nu\left(\bigcap_i^n F_i^c \cap E\right)\right| + \epsilon
\end{aligned}$$

and so, combining all the inqualities, we get:

$$\left|\int_E f d\nu\right| \leq \sum_{i=1}^n |\nu(F_i \cap E)| + \left|\nu\left(\bigcap_i^n F_i^c \cap E\right)\right| + \epsilon$$

and so, by the definition of the supremum:

$$\left|\int_E f d\nu\right| \leq \sup S_1^E$$

and since this is true for arbitrary $\left|\int_E f d\nu\right|$:

$$\mu_3 \leq \mu_1$$

and so, we've established $\mu_1 \leq \mu_2 \leq \mu_3 [= |\nu|] \leq \mu_1$, and so they are all equal, as we sought to show

3.4 Differentiation of Euclidean Space

We know analyze the special case where $\mu = m$ is the Lebesgue measure on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$. Since we are working over the Borel sets of $\mathbb{R}^n$, we have a metric available to us, and so we can define the pointwise derivative of ν with resepct to m at x:

Definition 3.4.1: Pointwise Derivative

Let $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}, m)$ be our measure space, ν a signed or complex measure on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$, and $x \in \mathbb{R}^n$. Then if $B_r(x)$ is a ball of radius r around x, then if the following limit exists:

$$\lim_{r \to 0} \frac{\nu(B_r(x))}{m(B_r(x))} = F(x)$$

then we say that $F(x)$ is the *pointwise derivative of* ν *with respect to* m *at* x

We can replace balls of radius x with sets that shrink suitably quickly in some "regular" way. This will be advantageous to us later, since it will turn out we can make these "nicely shrinking" sets to be $(x, x+h)$ (or the appropriate generalisation in the $\mathbb{R}^n$ case) which will give us

$$\lim_{r \to 0} \frac{\nu(x+h)}{h}$$

notice that if we define ν in terms of a distribution as we've done for the outer measure in section 1.3, say G is our distribution, then our numerator becomes $G(x+h) - G(x)$, so in this cse the definition *exacly* matches be the definition of the derivative! We will get back to this important case after some more examinatino of the general case.

If $\nu \ll m$ so that $d\nu = f dm$, then $\nu(B_r(x))/m(B_r(x)) = \frac{1}{m(B_r(x))} \int_{B_r(x)} f d\mu$ is the average value of f on $B_r(x)$ (recall the 1-dimensional intuition of $\frac{1}{b-a} \int_a^b f(x)dx$). Since $r \to 0$, we can hope that $F = f$ m-a.e. (since the averaging is getting more precise with the smaller range), telling us that we can recover a density F with information about f. This will indeed be the case if $\nu(B_r(x))$ is finite for all r, x; this will be what we explore in this section. Essentially, from the point of view of f, what we've showing is that the derivative of the integral of f is f, i.e., the Fundamental Theorem of Calculus (FTC). For the rest of this section, the term integrable and a.e. is with respect to the Lebesgue measure m.

We start with an important technical lemma:

Lemma 3.4.1: build-up lemma I

Let $\mathcal{C}$ be a collection of open balls in $\mathbb{R}^n$ and let $U = \bigcup_{B \in \mathcal{C}} B$. Then for any $c \in \mathbb{R}$ such that $c < m(U)$, there exists a fininte disjoint subset of $\mathcal{C}$, $B_1, ..., B_k$, such that

$$\sum_{i=1}^{k} m(B_i) > \frac{c}{3^n}$$

In other words, for any value c less than $m(U)$, we can show that we only need to pick a finite number of balls from $\mathcal{C}$ and still keep the inequality, up to a correction factor of 3^{-n} where n is the dimension of our space, $\mathbb{R}^n$, and 3 comes form making sure the radius of the balls will be big enough to cover what we need.

***Proof* :**
We essentially sneakily get compactness by using theorem 2.6.1[1]. Using this theorem, we know that there exists a compact set $K \subseteq U$ such that $c < \mu(K)$, so there exists a finite subcover of balls from $\mathcal{C}$, say $A = \{A_1, ..., A_m\}$, that cover K. Since balls have finite radius and there are finitely many balls, let B_1 be the ball with largest radius from A. Then let B_2 be the largest ball from A that are disjoint from B_1. Continuing until we have exhausted all possibilites, we get a collection of balls $B = \{B_1, ..., B_k\}$. If there is an A_i that is not in B, then there must be a B_j such that $A_i \cap B_j \neq \emptyset$, and if j is the smallest integer with this property, then by our construction A_i can have radius that is at most that of B_j (or else we would have picked A_i instead of B_j in our construction). Thus, we can take a ball B_j^* whose radius is 3 times that of B_j and have the same center as B_i (i.e., they are concentric) so that $A_i \subseteq B_j^*$. Then by this construction, $K \subseteq \bigcup_j^k B_j^*$, and so

$$c < \mu(K) \leq \sum_j^k m(B_j^*) = 3^n \sum_j^k m(B_j)$$

and so $\frac{c}{c^n} \leq \sum_j^k m(B_j)$, as we sought to show

With this lemma ready to be referenced, we define another important concept that we will use

Definition 3.4.2: Locally Integrable

Let $f : \mathbb{R}^n \to \mathbb{C}$ be a function. Then we say that f is *locally integrable* (with respect to the Lebesgue measure) if:

$$\forall K \in \mathcal{B}_{\mathbb{R}^n}, m(K) < \infty, \quad \int_K |f(x)|dx < \infty$$

and is denoted L^1_{loc}

Definition 3.4.3: Average

Let $f \in L^1_{\text{loc}}$. Then we define the average $A_r f(x)$ on $B_r(x)$ to be

$$A_r f(x) = \frac{1}{B_r(x)} \int_{B_r(x)} f(y)dy$$

This is simply the observation we made in the previous paragraph but made into a definition. We next prove that varying r or x by a bit will only change $A_r f(x)$ by a bit, i.e., $A_r f(x)$ is continuous:

Lemma 3.4.2: Build-Up Lemma II

Let $f \in L^1_{\text{loc}}$. Then $A_r f(x)$ is continuous on both r ($r > 0$) and x ($x \in \mathbb{R}^n$)

Notice that f certainly need not be continuous, while $1/m(B_r(x))$ is clearly continuous. It is the $\int$

operation that is forcing this result to be continuous.

Proof **:**
First, taking advantage that we're working over the Lebesgue measure in $\mathbb{R}^n$, so that if $c = B_1(0)$, then $cr^n = B_r(x)$ and $m(\partial B_r(x)) = 0$ (where $\partial B_r(x) = \{y \mid \|x - y\| = r\}$). This will come back soon.

To actually prove continuity, we'll show that $A_{\lim_{r\to r_0}} f(x) = \lim_{r\to r_0} A_r f(x)$ and $A_r f(\lim_{x\to x_0}) = \lim_{x\to x_0} A_r(x)$. To that end, take a sequene $r \to r_0$ and $x \to x_0$, and notice that the $\chi_{B_r(x)} \to \chi_{B_{r_0}(x_0)}$ pointwise on $\mathbb{R}^n \setminus \partial B_{r_0}(x_0)$ (imagine $x = x_0$ is fixed and the radius goes to r_0 by oscilating around r_0 so that the pointwise limit doesn't exist for the border). Thus, $\chi_{B_r(x)} \to \chi_{B_{r_0}(x_0)}$ a.e., and so we have that $\chi_{B_{r_0+1}(x_0)}$ dominates the $\chi_{B_r(x)}$ if $r < r_0 + 1/2$ and $|x - x_0| < 1/2$. Thus, by the Dominated Converting Theorem:

$$\begin{aligned}
A_{\lim\limits_{r\to r_0} r} f(\lim_{x_0\to x} x) &= \int_{B_{\lim\limits_{r\to r_0} r}(\lim\limits_{x\to x_0} x)} f(y)dy \\
&= \int_{B_r(x)} \lim_{r,x} f(y)dy \\
&= \lim_{r,x} \int_{B_r(x)} f(y)dy && \text{DCT}
\end{aligned}$$

and so $\int_{B_r(x)} f(y)dy$ is continuous. Since $B_r(x) = cr^n$ and r^n is a continuous funciton, and the product of two continuous function is continuous, we have that

$$A_r f(x) = \frac{1}{cr^n} \int_{B_r(x)} f(y)dy$$

is also continuous, as we sought to show

Since $A_r f$ is continuous, it is meaurable. Next, we can define the largest possible average given a point x:

Definition 3.4.4: Hardy-Littlewood Maximal Function

Let $f \in L^1_{\text{loc}}$. Then the *Hardy-Littlewood Maximal function* is defined as

$$Hf(x) := \sup_{r>0} A_r|f|(x) = \sup_{r>0} \frac{1}{m(B_r(x))} \int_{B_r(x)} |f(y)|dy$$

Note that Hf is measurable since $(Hf)^{-1} = \bigcup_{r>0}(A - r|f|)^{-1}((a, \infty))$ is open for any $a \in \mathbb{R}$ by the previous lemma. What the next theorem shows us is that if we fix some number $\alpha \in \mathbb{R}_{>0}$, and try to take the measure of all points where $Hf(x) > \alpha$ (i.e., we are measuring the area of the largest possible average over an area of points with that larger average), then this value is in fact bounded by $\int |f(x)|$ provided an appropraite constant C/α:

Theorem 3.4.1: The Maximal Theorem

Let $(\mathbb{R}^n, \mathcal{L}, m)$ be a measure space. There exists a constant C such that for all $f \in L^1$ and all $\alpha > 0$

$$m(\{x \mid Hf(x) > \alpha\}) \leq \frac{C}{\alpha} \int |f(x)| dx$$

Proof **:**
Let $E_\alpha = \{x \mid Hf(x) > \alpha\}$. By suprmum property of $Hf(x)$, there exists a $r_x > 0$ such that

$$\frac{1}{m(B_{r_x}(x))} \int_{B_{r_x}(x)} |f(y)| dy = A_{r_x}|f|(x) > \alpha \quad \Leftrightarrow \quad m(B_{r_x}(x)) < \frac{1}{\alpha} \int_{B_{r_x}(x)} |f(y)| dy \tag{3.1}$$

Clearly, the balls $B_{r_x}(x)$ cover E_α. So for any $c < m(E_\alpha)$, by lemma 3.4.1, there exists a finite collection $r_{x_1}, ..., r_{x_k} \in E_\alpha$ such that

$$\sum_{i=1}^{n} m(B_{r_{x_j}}(x_j)) > \frac{c}{3^n}$$

and so, combining the inequalities we have and moving values around, we get:

$$c < 3^n \sum_{i=1}^{n} m(B_{r_{x_j}}(x_j)) \overset{\text{eq. 3.1}}{\leq} \frac{3^n}{\alpha} \int_{B_{r_{x_j}}} |f(y)| dy \leq \frac{3^n}{\alpha} \int_{\mathbb{R}^n} |f(y) dy|$$

Since this is true for all c, letting $c \to m(E_\alpha)$, we get the desired result, as we sought to show.

With all this build-up, we are ready to show the result we stated at the beginning, that $A_r f(x) \to f(x)$ a.e.

Before stating the theorem, it is useful to define the notion of limit superior for real-valued functions:

Definition 3.4.5: Limit Suprerior

Let $\phi : \mathbb{R} \to \mathbb{R}$. Then:

$$\limsup_{r \to R} \phi(r) := \lim_{\epsilon \to 0} \sup_{0 < |r-R| < \epsilon} \phi(r) = \inf_{\epsilon > 0} \sup_{0 |r-R| < \epsilon} \phi(r)$$

As an exercise, show that:

$$\lim_{r \to R} \phi(r) = c \quad \text{iff} \quad \limsup_{r \to R} |\phi(r) - c| = 0$$

Theorem 3.4.2: Almost Lebesgue Differentiation Thoerem

If $f \in L^1_{\text{loc}}$. Then
$$\lim_{r\to 0} A_r f(x) = f(x)$$
for a.e. $x \in \mathbb{R}^n$. Phrased another way:
$$\lim_{r\to 0} \frac{1}{mB_r(x)} \int_{B_r(x)} f(x) - f(y)dy = 0$$
for a.e. $x \in \mathbb{R}^n$.

Proof **:**
First, let's show we can limit our attention to function in L^1. In particular, If $A_r f(x) \to f(x)$ for a.e. x, for any $N \in \mathbb{N}$, we can bound what x's we choose by $|x| \le N$. Since for any $r \le 1$, the values $A_r f(x)$ depend only on $f(y)$ for $|y| \le N+1$, we can replace f with $f\chi_{B_{N+1}(0)}$, so $f \in L^1$

By theorem 2.6.2, for every $\epsilon > 0$, there exists a continuous integrable function g such that $\int |g(y) - f(y)|dy < \epsilon$. Since g is continuous, for every $x \in \mathbb{R}^n$, for all $\delta > 0$, there exists an $r > 0$ such that $|y - x| < r$ implies $|g(y) - g(x)| < \delta$. From this, we get the result to work for g:
$$\begin{aligned}
|A_r g(x) - g(x)| &= \left| \frac{1}{m(B_r(x))} \int_{B_r(x)} g(y)dy - g(x) \right| \\
&= \left| \frac{1}{m(B_r(x))} \int_{B_r(x)} g(y)dy - \frac{m(B_r(x))}{m(B_r(x))} g(x) \right| \\
&= \frac{1}{m(B_r(x))} \left| \int_{B_r(x)} g(y) - g(x)dy \right| \\
&< \frac{1}{m(B_r(x))} \left| \int_{B_r(x)} \delta dy \right| \\
&= \frac{1}{m(B_r(x))} |m(B_r(x))\delta| \\
&= \delta
\end{aligned}$$
and so, since δ can be anything, we hve that $A_r g(x) \to g(x)$ for a.e. x.

Now we use this result to prove it for f. We will use the Maximal theorem to bound the result. Let:
$$\begin{aligned}
\limsup_{r\to 0} |A_r f(x) - f(x)| &= \limsup_{r\to 0} |A_r f(x) - A_r g(x) + A_r g(x) - g(x) + g(x) - f(x)| \\
&= \limsup_{r\to 0} |A_r (f-g)(x) + (A_r g - g)(x) + (g-f)(x)| \\
&\le H(f-g)(x) + 0 + |f-g|(x)
\end{aligned}$$
Or, in terms of sets, if
$$E_\alpha = \left\{ x \;\middle|\; \limsup_{r\to 0} |A_r f(x) - f(x)| > \alpha \right\} \qquad F_\alpha = \{x \mid |f-g| > \alpha\}$$

then

$$E_\alpha \subseteq F_{\alpha/2} \cup \{x \mid H(f-g)(x) > \alpha/2\}$$

Furthermore, we have that

$$(\alpha/2)m(F_{\alpha/2}) \leq \int_{F_{\alpha/2}} |f(x)-g(x)|dx < \epsilon$$

Thus, by the maximal theroem:

$$m(E_\alpha) \leq \frac{2\epsilon}{\alpha} + \frac{2C\epsilon}{\alpha}$$

and since ϵ is arbitrary, $m(E_\alpha) = 0$ for all $\alpha > 0$. However, $\lim_{r\to 0} A_r f(x)$ for all $x \notin \bigcup_{i=1}^{\infty} E_{1/n}$, finishing the proof.

In the theorem, we proved it for $f(x)$; it can in fact be shown that it can be proven for $|f|$. To show this, we define the following concept for reference:

Definition 3.4.6: Lebesgue Set

Let f be a measurable function. Then the *Lebesgue set* is :

$$L_f := \left\{ x \;\middle|\; \lim_{r\to 0} \frac{1}{m(B_r(x))} \int_{B_r(x)} |f(x)-f(y)|dy = 0 \right\}$$

The following theorem shows that the points on which the Lebesgue set fails on is measure 0, and so the result is true for L_f as well

Theorem 3.4.3: Upgrade To Lebesgue Set

Let $f \in L^1_{\text{loc}}$. Then $m((L_f)^c) = 0$

Proof **:**
This proof will take advantage of the zero measure set in theorem 3.4.2, and that we can form a dense set of $\mathbb{C}$ which will correspond the zero-measure set from this previous theorem, meaning we can make a limit argument that will work for the proof.

First, for each $c \in \mathbb{C}$, we can apply theorem 3.4.2 to $g_c = |f(x) - c|$ to get that:

$$\lim_{r\to 0} \frac{1}{m(B_r(x))} \int_{B_r(x)} |f(y)-c|dy = |f(x)-c|$$

up to some Lebesgue zero measure set, say N_c. Now, Let D be any countable dense subset of $\mathbb{C}$ and let $N = \bigcup_{c\in D} N_c$ (so $m(N) = 0$). If $x \notin N$, then for any $\epsilon > 0$, we can choose some $c \in D$ such that $|f(x) - c| < \epsilon$. Thus:

$$|f(y) - f(x)| < |f(y) - c| + \epsilon$$

Thus, we have that:

$$\limsup_{r\to 0} \frac{1}{m(B_r(x))} \int_{B_r(x)} |f(y)-f(x)|dy \leq |f(x)-c| + \epsilon \leq 2\epsilon$$

and since ϵ can be arbitrary small, we have that $\limsup_{r\to 0} \frac{1}{m(B_r(x))} \int_{B_r(x)} |f(y) - f(x)| dy = 0$, and so if $x \notin N$, the theorem applies. But then $0 = m(N) = m((L_f)^c)$, as we sought to show.

We have essentially proven the main results of this section; the last thing we want to do is generalize the types of sets that shrink to x. So far, we have been using balls, however that is not necessary. Thus, we will generalize the types of sets and state the most general version of this "averaging" theorem

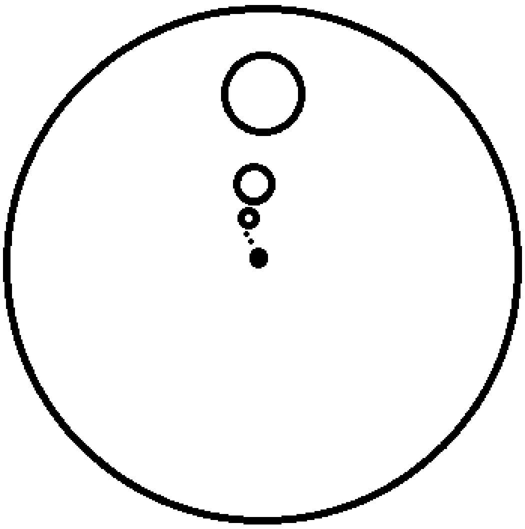
Definition 3.4.7: Sets That Shrink Nicely

A family of set $\{E_r\}_{r>0}$ of Borel subsets of $\mathbb{R}^n$ is said to *shrink nicely* if :

1. $E_r \subseteq B_r(x)$ for each r (so $m(E_r) \leq m(B_r(x))$)
2. There is a constant $\alpha > 0$ that is indepenent of r such that $m(E_r) > \alpha m(B_r(x))$

Example 3.4: Nicely Shrinking Sets

1. Let $U \subseteq B_1(0)$ be any borel subset. such that $m(U) > 0$, and consider $E_r = \{x + ry \mid y \in U\}$. You can imagine that U is some circle of $B_1(0)$ and see that, if $x = 0$:



   Then we see that E_r shrinks to to x (even if $x \notin B_1(0)$)

2. in $\mathbb{R}$, the set $E_r = (x, x + r]$ also shrink nicely.

With this, we can finally state the result we've building up to in full generality:

Theorem 3.4.4: The Lebesgue Differentiation Theorem

Let $f \in L^1_{\text{loc}}$. Then for every x in the Lebesgue set, that is almost every x:

$$\lim_{r\to 0} \frac{1}{m(E_r)} \int_{E_r} |f(y) - f(x)| dy = 0$$

and

$$\lim_{r\to 0} \frac{1}{m(E_r)} \int_{E_r} f(y) dy = f(x)$$

for every family of nicely shrinking sets $\{E_r\}_{r>0}$ that shrink nicely to x

Written in terms of measure's this tells us that if $d\nu = f d\mu$, then:

$$\lim_{r\to 0} \frac{\nu(E_r)}{m(E_r)} = f(x)$$

***Proof* :**
To prove the first equality, we simply use the definition of nicely shrinking sets:

$$\begin{aligned}\frac{1}{m(E_r)} \int_{E_r} |f(y) - f(x)| dy &\leq \frac{1}{m(E_r)} \int_{B_r(x)} |f(y) - f(x)| dy \\ &\leq \frac{1}{\alpha m(B_r(x))} \int_{E_r} |f(y) - f(x)| dy\end{aligned}$$

so we can use theorem 3.4.3.

For the second equality, intead of putting absolute value, can start without them, in which case we would do the same think and then use theorem 3.4.2, and move the $f(x)$ to the other side.

3.4.1 Regular Measures

We will apply these results to a special class of measures which, as is common with such definitions, we take some aspects that we like from the Lebesgue measure on $\mathbb{R}^n$, and want to study measure's more generally (in this case, borel measures on $\mathbb{R}^n$) that have the same property:

Definition 3.4.8: Regular Measure

Let $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}, \nu)$ be a borel measure space. Then ν is called *regular* if:

1. $\nu(K) < \infty$ for all compact subset $K \subseteq \mathbb{R}^n$
2. $\nu(E) = \inf \{\nu(U) \mid U \text{ open}, E \subseteq U\}$ for every $E \in \mathcal{B}_{\mathbb{R}^n}$

It is in fact provable that the first condition implies the second. We have shown this for $\mathbb{R}$ (i.e. $n = 1$) in theorem 1.4.1 and 1.4.2. In the study of Radon measures, this proof can be done for arbitrary n (Folland chapter 7.2).

Notice that a regular measure is automatically σ-finite, since $\mathbb{R}^n$ is σ-compact. As usual, a signed or complex measure ν is regular if $|\nu|$ is regular.

If $f \in L^+(\mathbb{R}^n)$, then the measure fdm is regular if and only if $f \in L^1_{\text{loc}}$. This means that we can generalize the Lebesgue differentiation theorem to regular measures in general instead of measures defined on densities which are locally integrable. Let's first check that that assertion is indeed the case. First, it should be clear that $f \in L^1_{\text{loc}}$ is equivalent to $\nu(K) < \infty$ for all compact sets. To show 2, we'll first show it works for bounded borel sets E. By theorem 2.6.1, for any $\delta > 0$, there exists an open set $U \supseteq E$ such that $\mu(U) < \mu(E) + \delta$, thus $\mu(U \setminus E) < \delta$. By theorem 3.2.1, for all $\epsilon > 0$, there exists a $U \supseteq E$ such that $\int_{U\setminus E} fdm < \epsilon$, so $\int_U fdm < \int_E fdm + \epsilon$, thus $\nu(U) < \nu(E) + \epsilon$. This completes the requirement of the infimum. For the unbounded case, take $E = \cup_i E_i$ where each E_i is finite, finding it in each case while choosing $\int_{U_i\setminus E_i} fdm < \epsilon 2^{-i}$.

With this equivalence, it becomes worthwile exploring Lebesgue Differentiation Theorem in the case of regular measures:

Theorem 3.4.5: Lebesgue Differentiation With Regular Measure

Let ν be a regular signed or complex Borel measure on $\mathbb{R}^n$, and let $d\nu = d\lambda + fdm$ be its Lebesgue decomposition. Then for m-almost every $x \in \mathbb{R}^n$:

$$\lim_{r\to 0} \frac{\nu(E_r)}{m(E_r)} = f(x)$$

for every nicely shrinking family $\{E_r\}_{r>0}$ that nicely shrinks to x

Proof **:**
We have essentially already done this proof, except now we have to consider we are adding $d\lambda$, $d\nu = d\lambda + fdm$, and so we would like to show that $\frac{\lambda(E_r)}{m(E_r)} \to 0$.

First, notice that $d|\nu| = d|\lambda| + |f|d\mu$ (or $|f|dm$?). Thus, regularity in ν implies regularity of λ and fdm (this should be proven, though it is a bit tedious[a]). Since $f \in L^1_{\text{loc}}$, it is enough to show that if $\lambda \perp m$, then for all x, $\lambda(E_r)/m(E_r) \to 0$ m-a.e as $r \to 0$. Without loss of generality, we can let $E_r = B_r(x)$ and assume λ is positive since:

$$\left|\frac{\lambda(E_r)}{m(E_r)}\right| \le \frac{|\lambda|(E_r)}{m(E_r)} \le \frac{|\lambda|(B_r(x))}{m(E_r)} \le \frac{|\lambda|(B_r(x))}{\alpha m(B_r(x))}$$

for some $\alpha > 0$. Now, let's say $\lambda \geq 0$ (if $\lambda = 0$, then we'd be done), and take $A \in \mathcal{B}_{\mathbb{R}^n}$ such that $\lambda(A) = m(A^c) = 0$ (considering that $\lambda \perp m$). We will do the usual trick when it comes to showing that something has measure zero by assuming it doesn't, havign some sets define define in terms of a " $1/k$ sequence", so to speak, and show we reach a contradiction. Consider:

$$F_k = \left\{ x \in A \;\middle|\; \limsup_{r\to 0} \frac{\lambda(B_r(x))}{m(B_r(x))} > \frac{1}{k} \right\}$$

We will show that $m(F_k) = 0$ for all k, meaning the set of points for which $\lambda(E_r)/m(E_r) \to 0$ is 0 m-a.e. First, since λ is regular, for all $\epsilon > 0$, there exists a $U_\epsilon \supseteq A$ such that $\lambda(U_\epsilon) < \epsilon$. Since U_ϵ is open, for each point $x \in U_\epsilon$, there exists a ball such that $B(x) \subseteq U_\epsilon$. By definition of F_k, $\lambda(B(x)) > \frac{m(B(x))}{k}$. or $k\lambda(B(x)) > m(B(x))$. Next, take $V_\delta = \cup_x B(x)$. Then by construction

$F_k \subseteq V_\epsilon$. We'll show that $m(V_\epsilon) \leq 3^n k\epsilon$ for all ϵ, and so $m(F_k) = 0$ or all k. To do this, notice that by assumption there is a constant c such that $c < m(V_\epsilon)$. Thus, by lemma 3.4.1, there exist disjoint balls around points $x_1, ..., x_n$ such that

$$c < 3^n \sum_1^n m(B_{x_i}) \leq 3^n k \sum_1^n \lambda(B_{x_i}) \leq 3^n k\lambda(V_\epsilon) \leq 3^n k\lambda(U_\epsilon) \leq 3^n k\epsilon$$

showing that $m(F_k) \leq m(V_\epsilon) < 3^n k\epsilon$, and since ϵ is arbitrary, we have that $m(F_k) = 0$ for all k, as we sought to show.

[a]It was homework 8, q26

3.5 Functions of Bounded Variation

We will now work towards the Fundamental Theorem of Calculus in its generlized form. This generalization is not in the direction of higher dimensions, as is the case with Stokes Theorem which is a generalization into higher dimensions, but requires much nicer conditions like working over smooth manifolds and differential forms. Rather, we want to stay in $\mathbb{R}$ (or $\mathbb{C}$) and see if we can weaken the conditions (i.e. weaker than differentiable) and see when we can still apply FTC.

In the case of $\mathbb{R}$, we have alraedy explored in section 1.4 that Borel measures defined on $\mathbb{R}$ are defined through an associated distribution F, which is a right-continuous increasing function. In particular, we defined μ_F based on a increasing right-continuous function F (i.e. a distribution) to be $\mu([a, b]) = F(b) - F(a)$. We will use the tools we build up in the last section to finally prove the Fundamental theorem of Calculus in full generality. We will then explore when we can define distributions when we have a complex function, and substitute the notion of increasing function with a function of *bounded variation*.

The first thing we will do is show how increasing functions are a.e. differentiable (which to me is sort of crazy)!!

Proposition 3.5.1: Increasing Then a.e. Differentiable

Let $F : \mathbb{R} \to \mathbb{R}$ be an increasing function. Then the set of points where F is discontinuous is countable.
Furthermore, if $G(x) = \lim_{y\to x^+} F(y) = F(x+)$, then F and G are differentiable a.e. and $F' = G'$ a.e.

Proof **:**
First, if x is continuous at x, then $F(x+) = F(x-)$, so the interval $(F(x-), F(x+))$ is empty. If it is not empty, then since F is increasing, each $(F(x-), F(x+))$ must be disjoint. So, for any $N \in \mathbb{N}$, for all $|x| \leq N$, we have that $(F(x-), F(x+)) \subseteq (F(-N), F(N))$. Now, here is the trick of this proof: notice that

$$\sum_{|x|<N} [F(x+) - F(x-)] \leq F(N) - F(-N) < \infty$$

since the $F(x-)$ is equal to $F(x_+)$ at appropriate locations, and so the sum cancels out (note that this sum is absolutely continuous since $F(x+) \geq F(x-)$). To see more clearly what's going on ,if we label all the points in $|x| < N$ through an indexing set A, and let $x_\alpha = (F(x_\alpha-) - F(x_\alpha-))$, then:

$$\sum_{\alpha < A} x_\alpha < \infty$$

From here, we use a the fact that this the sum of uncountably many positive x_α, and use the fact that the sum of uncountably many positive elements must be infinite if more than countably elements are non-zero, but since the sum *must* be finite $\{x \in (-N, N) \mid f(x+) \neq F(x-)\}$ is countable. Since this is true for all N, the set of discontinuities of F must be countable, proving the first assertion.

For the second assertion, first note that G is increasing and right continuous and $G = F$ at all points excepts where F is discontinuous. To show G is differentaible a.e, by proposition 1.4.1

$$G(x+h) - G(x) = \begin{cases} \mu_G((x, x+h]) & \text{if } h > 0 \\ \mu_G((x+h, x]) & \text{if } h < 0 \end{cases}$$

Notice that $\{(x-r, x]\}$ and $\{(x, x+r]\}$ shrink nicely to x as $r = |h| \to 0$. Furhtermore μ_G is regual by theorem 1.4.1. Thus, by theorem 3.4.5, we have that:

$$G'(x) = \lim_{r \to 0} \frac{G(x+h) - G(x)}{h} = \lim_{r \to 0} \frac{\mu_G(E_r)}{m(E_r)}$$

exists for a.e. x. The final part of this proof is to show that if $H = G - F$, then H' exists and is equal to zero a.e.

Let $\{x_i\}$ be an enumeration of the points where $H \neq 0$. Since $H = G - F$, $H(x_i) > 0$. Let's say $\delta_i : \mathbb{R} \to \mathbb{R}$ is a function where $\delta_i(x_i) = 1$ and zero eveywehre else. Then for any $N \in \mathbb{N}$, since H is increasing we have $\sum_{\{i \mid |x_i| < N\}} H(x_i) < \infty$. Thus, we can define a new measure:

$$\mu = \sum_i H(x_i)\delta_i$$

By what we've shown, μ is finite on all compact sets by the previous lines reasoning, and so by the theorem's we've been using μ is regular. Furthermore, $\mu \perp m$ since $m(E) = \mu(E^c) = 0$ where $E = \{x_i\}_{i=1}^\infty$. Thus:

$$\left| \frac{H(x+h) - H(x)}{h} \right| \leq \frac{H(x+h) + H(x)}{|h|} \overset{!}{\leq} \frac{\mu\left(x - 2|h|, x + 2|h|\right)}{|h|}$$

where $\overset{!}{\leq}$ comes from trying to ingulf $H(x+h)$ and $H(x)$ into an interval.

Since all measure are regular, by theorem 3.4.5 this limit exists and $H' = 0$ a.e., completing the proof.

With this, we now have the tools to try and explore how to think about "increasing function" of the form $F : \mathbb{R} \to \mathbb{C}$. Since $\mathbb{C}$ has no inheret order, we must be more precise on how we define "increasing" and how do we make sure it doesn't explode (just like how when defining the borel measures through a distribution)

To understand the motivation behind the following definitions, imagine the following scenario: let F

be the function representing the movement of a particle as it travels accross in time with respect to t (the graph being the position of the particle relative to, say, where it started). The *total variation* from time a to time b of the particle would be the total distance the particle has travelled in that time frame. if F were differentiable, this would simply be:

$$\int_a^b |F'(x)|dx$$

However, what if F is not differentiable? In that case, we will take the following approach: partition $[a, b]$ into $[a, t_1], [t_1, t_2], ..., [t_{n-1}, b]$. Then, linearly approximate F by a straight line from $(t_{i-1}, F(t_{i-1}))$ to $(t_i, F(t_i))$. Then, take the limit over all partitions (in particular, as partitions get finer). If you are worried that such a limit is not well-defined, remember that polynomials can arbitrarily well approximate continuous functions on $\mathbb{R}$, and in most cases F is right-continuous. This heuristic might be more rigorous, but it at least make me feel comfortable with doing this idea

We will do a similar idea for the complex case, except we will take the norm (which we'll also call absolute value) of the distances:

Definition 3.5.1: Total Variation

Let $f : \mathbb{R} \to \mathbb{C}$ and $x \in \mathbb{R}$. Then define the *total variation of F* to be:

$$T_F(x) = \sup\left\{\sum_{i=1}^{n} |F(x_i) - F(x_{i-1})| \;\middle|\; n \in \mathbb{N},\ -\infty < x_0 < x_1 < \cdots < x_n = x\right\}$$

I should be clear by construction that T_F is an increasing function. Since we put absolute values inside the sum, then as we refine the partition, the sum will get bigger. Therefore, If we have $a < b$ and consider $T_F(b)$, it is harmless to assume a is one of the subdivision points (since we can refine and get a partition with a being part of it). It follows that we have:

$$T_F(b) - T_F(a) = \sup\left\{\sum_{i=1}^{n} |F(x_i) - F(x_{i-1})| \;\middle|\; n \in \mathbb{N},\ a = x_0 < x_1 < \cdots < x_n = x\right\}$$

Using this fact, we can define the following:

Definition 3.5.2: Bounded Variation

Let T_F be the total variation of the function F. Then if:

$$\lim_{t\to\infty} T_F(t) = T_F(\infty) < \infty$$

Then F is said to be of *bounded variation* (on $\mathbb{R}$). We denote the set of of all functions of bounded variation as BV. If we are concerned with all function of bounded variation on an interval $[a, b]$, we will denote this as $BV([a, b])$.

Note that if $F \in BV$, then $F \in BV([a, b])$, since it's total variation on $[a, b]$ is $T_F(b) - T_F(a)$ which is less thatn the total vriation of T_F. If $F \in BV([a, b])$, then we can extend F to be in BV by letting $F(x) = F(a)$ for all $x \leq a$ and $F(b) = F(x)$ for all $b \leq x$. Further note that we rarely actually compute $T_F(b)$, but we use the fact that a function is in BV to get many theoretical results:

Example 3.5: Bounded Variation

1. If $F : \mathbb{R} \to \mathbb{R}$ is bounded and increasing, then $F \in BV$ $(T_f(x) = F(x) - F(-\infty))$
2. If $F, G \in BV$ and $a, b \in \mathbb{C}$ then $aF + bG \in BV$
3. If F is differentiable on $\mathbb{R}$ and F' is bounded, then $F \in BV([a,b])$ for $-\infty < a < b < \infty$ (by the mean value theorem)
4. If $F(x) = \sin(x)$, then $F \in BV([a,b])$ for all $-\infty < a < b < \infty$, but $F \in BV$
5. if $F = x\sin\left(\frac{1}{x}\right)$ for $x \neq 0$ and $F(0) = 0$, then $F \notin BV([a,b])$ for $a \leq 0 < b$ and $a < 0 \leq b$

Before we prove stuff about BV functions, let's first prove a lemma:

Lemma 3.5.1: Sum Is Increasing

Let $F \in BV$ be a real valued function. Then $T_F + F$ and $T_F - F$ are increasing functions.

***Proof* :**
Take $x < y$ and $\epsilon > 0$. Since T_F is a supremum, take appropriat $x_0 < \cdots x_n = x$ such that

$$\sum_i^n |F(x_i) - F(x_{i-1})| \geq T_F(x) - \epsilon$$

Next, notice that we can make this into an element of the supremum set of $T_f(y)$ (also known as an approximating sum) by doing

$$\sum_i^n |F(x_i) - F(x_{i-1})| + |f(y) - f(x)|$$

With this sum, take $F(y) = [F(y) - F(x)] + F(x)$. With some manipulation, we get:

$$T_f(y) \pm F(y) \geq \sum_i^n |F(x_i) - F(x_{i-1})| + |f(y) - f(x)| \pm [F(y) - F(x)] \pm F(x)$$

Since F is a real valued function, we cna simply take cases to notice that:

+	F(x) - F(y)	+	F(x) - F(y)	2|F(y) - F(x)|
+	F(x) - F(y)	-	F(x) - F(y)	0
-	F(y) - F(x)	+	F(x) - F(y)	0
-	F(y) - F(x)	-	F(x) - F(y)	2|F(y) - F(x)|

$$\begin{aligned} T_f(y) \pm F(y) &\geq \sum_i^n |F(x_i) - F(x_{i-1})| + |f(y) - f(x)| \pm [F(y) - F(x)] \pm F(x) \\ &\geq T_f(x) - \epsilon \pm F(x) \end{aligned}$$

letting $\epsilon \to 0$, we see that:
$$T_F(y) \pm F(y) \geq T_F(x) \pm F(x)$$
and since x and y was arbitrary, we have establish that $T_F + F$ and $T_F - F$ are both increasing functions, we as sought to show

Theorem 3.5.1: Properties Of BV

1. $T_F \in BV$ is an increasing function
2. $F \in BV$ if and only if $\Re F \in BV$ and $ImF \in BV$
3. If $F : \mathbb{R} \to \mathbb{R}$ then $F \in BV$ if and only if F is the difference of two bounded increasing functions; for $F \in BV$ these functions may be taken to be $\frac{1}{2}(T_F + F)$ and $\frac{1}{2}(T_F - F)$
4. If $F \in BV$, then $F(x+) = \lim_{y \to x^-} F(y)$ and $F(x-) = \lim_{y \to x^+} F(y)$ exist for all $x \in \mathbb{R}$, as do $F(\pm\infty) = \lim_{y \to \pm\infty} F(y)$
5. If $F \in BV$, the set of points at which F is discontinuous is countable
6. If $F \in BV$ and $G(x) = F(x+)$, then F' and G' are exist and are equal a.e.

***Proof* :**

1. This essentially comes from the definition
2. This is essentially immediate, and so is left as exam review
3. Look at example 3.5 to see how the $\Rightarrow$ direction goes. For the $\Leftarrow$ direction, by lemma 3.5.1, the equation $F = \frac{1}{2}(T_F + F) - \frac{1}{2}(T_F - F)$ express F as the difference of two increasing functions. To show boundedness, we showed that $T_F(y) \pm F(y) \geq T_F(x) \pm F(x)$ for $y > x$. Re-arranging and taking advantage of the fact that $F \in BV$, we get:
$$|F(y) - F(x)| \leq T_F(y) - T_F(x) \leq T_F(\infty) - T_F(-\infty) < \infty$$
showing that F (and hence T_F) is bounded

4,5,6 follow from 1,2,3 along with proposition 3.5.1.

Note that (6) is particularly powerful since we now have that a function simply needs to be of bounded variation to have a derivative a.e. (not necessarily increasing anymore). The converse is false: being differentiable a.e. (even everywhere) does not imply BV (think of $\sin(1/x)$ on $(0, 1]$).

The equation in part 2 are important enough to be given a name:

Definition 3.5.3: Jordan Decomposition

Let $F \in BV$ be a real-valued function. Then the *jordan decomposition* of F is:

$$F = \frac{1}{2}(T_F + F) - \frac{1}{2}(T_F - F)$$

where $\frac{1}{2}(T_F + F)$ and $\frac{1}{2}(T_F - F)$ are called the *positive variation of* F and *negative variation of* F respectively.

Since $x^+ = \max(x, 0) = \frac{1}{2}(|x| + x)$ and $x^- = \max(-x, 0) = \frac{1}{2}(|x| - x)$ for $x \in \mathbb{R}$, we get that:

$$\frac{1}{2}(T_F \pm F)(x) = \sum \left\{ \sum_i^n [F(x_i)F(x_{i-1})]^\pm \;\middle|\; x_0 < \cdots < x_n = x \right\} \pm \frac{1}{2}F(-\infty)$$

We are almost at a stage to connect borel measures to BV functions. There is one more peice of the puzzle that we need:

Definition 3.5.4: Normalized Bounded Variation (NBV)

Let $F \in BV$ be a function of bounded variation. Then if F is right-continuous and $F(-\infty) = 0$, then F is said to be a function of *normalized bounded variation*. We will denote the collection of such functions as:

$$NBV := \{F \in BV \mid F \text{ is right continuous and } F(-\infty) = 0\}$$

Example 3.6: NBV

Let $F \in BV$. Define $G = F(x+) - F(-\infty)$. Then $G \in NBV$, and $F' = G'$ a.e.

First, to show that $G \in BV$, by theorem 3.5.1[2,3], if F is real and $F = F_1 - F_2$ where F_1, F_2 are increasing, then $G(x) = F_1(x+) - [F_2(x+) - F(-\infty)]$, which is the difference of two increasing functions. That $G(-\infty) = 0$ comes by definition. Hence $G \in NBV$.

In fact, every $F \in BV$ almost defines an NBV function through T_F:

Lemma 3.5.2: BV, then T_F almost NBV

If $F \in BV$, then $T_F(-\infty) = 0$. If F is also right continuous, then T_F is also right-continuous (making $T_F \in NBV$)

Proof **:**
Let $\epsilon > 0$ so that for any x and $x_0 < \cdots < x$ we ahve:

$$\sum F(x_{i+1}) - F(x_i) = T_f(x) - T_f(x_0) > T_F(x) - \epsilon$$

So $T_f(x_0) < \epsilon$ or more generally for any $y \le x_0$ $T_f(y) < \epsilon$. Since epislon is arbitray, as we let $y \to 0$ we get $T_f(-\infty) = 0$

For the next assertion, suppose F is right-continuous. I'll finish this proof later:

Now suppose that F is right continuous. Given $x \in \mathbb{R}$ and $\epsilon > 0$, let $\alpha = T_F(x+) - T_F(x)$, and choose $\delta > 0$ so that $|F(x+h) - F(x)| < \epsilon$ and $T_F(x+h) - T_F(x+) < \epsilon$ whenever $0 < h < \delta$. For any such h, by (3.24) there exist $x_0 < \cdots < x_n = x + h$ such that

$$\sum_1^n |F(x_j) - F(x_{j-1})| \geq \tfrac{3}{4}[T_F(x+h) - T_F(x)] \geq \tfrac{3}{4}\alpha,$$

and hence

$$\sum_2^n |F(x_j) - F(x_{j-1})| \geq \tfrac{3}{4}\alpha - |F(x_1) - F(x_0)| \geq \tfrac{3}{4}\alpha - \epsilon.$$

Likewise, there exist $x = t_0 < \cdots < t_m = x_1$ such that $\sum_1^n |F(t_j) - F(t_{j-1})| \geq \frac{3}{4}\alpha$, and hence

$$\begin{aligned}
\alpha + \epsilon &> T_F(x+h) - T_F(x) \\
&\geq \sum_1^m |F(t_j) - F(t_{j-1})| + \sum_2^n |F(x_j) - F(x_{j-1})| \\
&\geq \tfrac{3}{2}\alpha - \epsilon.
\end{aligned}$$

Thus $\alpha < 4\epsilon$, and since ϵ is arbitrary, $\alpha = 0$. ∎

We finally get to establish the connection between complex meaures and functions in BV:

Theorem 3.5.2: Complex Meaures And BV

If μ is a complex Borel meaure on $\mathbb{R}$ (i.e. a finite signed measure) and $F(x) = \mu((-\infty, x])$, then $F \in NBV$. Conversely, if $F \in NBV$, then there is a unique complex Borel measure μ_F such that $F(x) = \mu_F((-\infty, x])$

***Proof* :**

Let μ a complex measure. Then we can decompose it into $\mu = \mu_1^+ - \mu_1^- + i(\mu_2^+ - \mu_2^-)$ where $\mu_i^\pm$ are positive finite measures. If $F_i^\pm(x) = \mu_i^\pm((-\infty, x])$, then $F_i^\pm$ is increasing, right continuous, $F_i^\pm(-\infty) = 0$, and $F_i^\pm(\infty) = \mu_i^\pm(\mathbb{R}) < \infty$. Thus, by theorem 3.5.1[2,3], the function

$$F = F_1 + -F_1^- + i(F_2^+ - F_2^-) \in NBV \tag{3.2}$$

Conversely, by theorem 3.5.1 and lemma 3.5.2, any $F \in NBV$ can be written in the from of equation (3.2) with the $F_i^\pm$ increasing and in NBV. By Theorem 1.4.1, each $F_i^\pm$ gives a measure $\mu_i^\pm$, so $F(x) = \mu_F((-\infty, x])$ where $\mu_F = \mu_1^+ - \mu_1^- + i(\mu_2^+ - \mu_2^-)$.

Finally, to show $|\mu_F| = \mu_{T_F}$ (this was left as an outlined exercise 28)

Corollary 3.5.1: Total Variation Of μ_F

Let $F \in NBV$, μ_F be the associated complex measure on $\mathbb{R}$, and $G(x) = |\mu_F|((-\infty, x])$. Then:

$$|\mu_F| = \mu_{T_F}$$

***Proof* :**
We break down the proof in the following steps:

1. From the definition of T_F, $T_F \leq G$

   Proof. We need to show $T_F(x) \leq G(x)$ In the previous homework, we showed that:

   $$|\mu_F|((-\infty, x]) = \sup\left\{\sum_i^{\infty} |\mu_F(E_i)| \;\middle|\; \bigsqcup_i E_i = (-\infty, x]\right\}$$

   In particular, we can choose are disjoint sets to be h-intervals and get:

   $$\begin{aligned} |\mu_F|((-\infty, x]) &= \sup\left\{\sum_i^{\infty} |\mu_F((x_i, x_{i+1}])| \;\middle|\; \bigsqcup_i (x_i, x_{i+1}] = (-\infty, x]\right\} \\ &= \sup\left\{\sum_i^{\infty} |F(x_{i+1}) - F(x_i)| \;\middle|\; \bigsqcup_i (x_i, x_{i+1}] = (-\infty, x]\right\} \end{aligned}$$

   and by definition of T_F

   $$T_F(x) = \sup\left\{\sum_i^{N} |F(x_{i+1}) - F(x_i)| \;\middle|\; N \in \mathbb{N},\ x_0 < \cdots < x_N = x\right\}$$

   Notice that the set for $|\mu_F|$ contains that of $T_F(x)$ since we can set the rest of the infinite value for T_F to 0, and so $T_F \leq G = |\mu_F|$ $\square$

2. $|\mu_F(E)| \leq \mu_{T_F}(E)$ when E is an interval, and hence E is a borel set.

   Proof. Without loss of generality, we will work with intervals of the form $(a, b]$. We only require these type of intervals for the σ-algebra that will be generatedi n part (c).

   First, since $F \in NBV$, then it is right-continuous, and so by lemma 3.28 T_F is right-continuous, and so by lemma 3.28 again $T_F \in NBV$. Thus, by theorem 3.29, there is an assocaited complex borel meaure such that $T_F(x) = \mu_{T_F}((-\infty, x])$. So $\mu_{T_F}((a, b]) = T_F(b) - T_F(a)$. With that, we proceed with the proof.

   First, the emptyset (the trivial interval) is immediate since $|\mu_F(\emptyset)| = 0 = \mu_{T_F}(\emptyset)$. Next, to show the inequality is true for bounded intervals, notice that:

   $$|\mu_F((a, b])| = |\mu_F((-\infty, b]) + \mu_F((-\infty, a]) = |F(b) - F(a)|$$

Next, recall that:

$$\mu_{T_F}((a,b]) = T_F(b) - T_F(a) = \sup\left\{ \sum_i^n |F(x_{i+1}) - F(x_i)| \;\middle|\; n \in \mathbb{N}\ a = x_0 < \cdots < x_n = b \right\}$$

So, $|F(b) - F(a)|$ is inside the supremum set, and so:

$$= |F(b) - F(a)| \leq T_F(b) - T_F(a) = \mu_{T_F}((a,b])$$

Next, for rays, notice we can split a ray into the countable union of h-intervals:

$$|\mu_F((-\infty,b])| = \left|\sum_{i=0}^{\infty} \mu_F((b-i-1, i-k])\right| \leq \sum_{i=0}^{\infty} |\mu_F((b-i-1, i-k])|$$

where we can now apply the results for bounded intervals to get that:

$$\leq \sum_{i=0}^{\infty} |\mu_F((b-i-1, i-k])| \leq \sum_{i=0}^{\infty} \mu_{T_F}((a,b]) = \mu_{T_F}((-\infty,b])$$

Thus $|\mu_F((-\infty,b])| \leq \mu_{T_F}((-\infty,b])$, and similarly for rays of the form (a,∞).

To show it's true for all Borel sets, first recall that h-intervals form an algebra (their disjoint union forms an elementary family, and hence by proposition 1.7). Next, let $\mathcal{C} = \{E \in \mathcal{B}_{\mathbb{R}} \mid |\mu_F(E)| \leq \mu_{T_F}(E)\}$. This set clearly contains all h-intervals by what we've just shown. We'll show that this set is in fact a monotone class, and hence by the monotone class lemma contains all Borel sets.

This is simply a matter of using the appropraite definitions. Let $\{E_i\}_{i=1}^{\infty} \subseteq \mathcal{C}$ be an increasing sequence. Then:

$$\begin{aligned}
\left|\mu_F\left(\bigcup_i^{\infty} E_i\right)\right| &= \left|\lim_{i\to\infty} \mu_F\left(\bigcup_i^{\infty} E_i\right)\right| && \text{definition}\\
&= \lim_{i\to\infty}\left|\mu_F\left(\bigcup_i^{\infty} E_i\right)\right| && \text{limit property}\\
&\leq \lim_{i\to\infty} \mu_{T_F}(E_i) && \text{all } E_i \in \mathcal{C}\\
&= \mu_{T_F}\left(\bigcup_i^{\infty} E_i\right) && \text{definition}
\end{aligned}$$

and hence $\mathcal{C}$ is closed under increasing sequences. A similar argument works for decreasing sequences (in particular since μ_{T_F} is finite, it will hold true for any decreasing sequence). Hence, by the monotone class lemma, $\mathcal{C}$ contains the Borel set, showing that $|\mu_F(E)| \leq \mu_{T_F}(E)$ for all $E \in \mathcal{B}_{\mathbb{R}}$, completing the proof.

□

3. $|\mu_F| \leq \mu_{T_F}$, and hence $G \leq T_F$ (use exercise 21.)

Proof. Recall that

$$|\mu_F|(E) = \sup \left\{ \sum_i^\infty |\mu_F(E_i)| \;\middle|\; E_i \in \mathcal{B}_\mathbb{R}, \bigsqcup_i E_i = E \right\}$$

where I modified what's inside the parenthesis to make it more explicit that the sets are inside $\mathcal{B}_\mathbb{R}$. Using what we have just proven:

$$\begin{aligned}
|\mu_F|(E) &= \sup \left\{ \sum_i^\infty |\mu_F(E_i)| \;\middle|\; E_i \in \mathcal{B}_\mathbb{R}, \bigsqcup_i E_i = E \right\} \\
&\leq \sup \left\{ \sum_i^\infty \mu_{T_F}(E_i) \;\middle|\; E_i \in \mathcal{B}_\mathbb{R}, \bigsqcup_i E_i = E \right\} \\
&\leq \sup \left\{ \mu_{T_F}(\sqcup_i E_i) \;\middle|\; E_i \in \mathcal{B}_\mathbb{R}, \bigsqcup_i E_i = E \right\} && \text{disjoint} \\
&\leq \sup \left\{ \sum_i^\infty \mu_{T_F}(E) \;\middle|\; E_i \in \mathcal{B}_\mathbb{R}, \bigsqcup_i E_i = E \right\} \\
&= \mu_{T_F}(E)
\end{aligned}$$

Thus, $|\mu_F|(E) \leq \mu_{T_F}(E)$. If we take $E = (-\infty, x]$, then we get $|\mu_F|(E) = G(E) \leq \mu_{T_F}(E)F$, completing the proof.

In particular, since we've shown the other direction, we have that $G = \mu_{T_F}$. □

With the connection established for between NBV functions and complex measures, it is natural to ask which function in NBV correspond to measure μ such that $\mu \perp m$ or $\mu \ll m$? The following proposition answers that:

Proposition 3.5.2: NBV And Radon-Nikodym with respect to m

If $F \in NBV$, then $F' \in L^1(m)$. Moreover:

1. $\mu_F \perp m$ if and only if $F' = 0$ a.e.
2. $\mu_F \ll m$ if and only if $F(x) = \int_{-\infty}^x F'(t)dt$

***Proof* :**
First, note that μ_F is regular by theorem 1.4.2. Next, observe that:

$$F'(x) = \lim_{r \to 0} \frac{\mu_F(E_r)}{m(E_r)}$$

where $E_r = (x, x+r]$ or $(x-r, x]$ by applying theorem 3.4.5

Since we have now characterized μ_F in terms of F, we can characterize $\mu_F \ll m$ in terms of F:

Definition 3.5.5: Absolute Continuity w.r.t. F

A function $F : \mathbb{R} \to \mathbb{C}$ is said to be *absolutely continuous* if

$$\forall \epsilon > 0,\ \exists \delta > 0 \text{ such that}$$

$$\text{for any set of disjoint intervals } (a_1, b_1), ..., (a_N, b_N),$$

$$\sum_1^N (b_i - a_i) < \delta \Rightarrow \sum_i^N |F(b_i) - F(a_i)| < \epsilon$$

More generally, F is absolutely continuous on $[a, b]$ if all the open intervals lie in $[a, b]$.

Notice that if F is absolutely continuous, then F is automaticaly uniform (pick $N = 1$). If F is everywhere differentiable and F' is bounded, then F is absolutely continuous, since $|F(b_i) - F(a_i)| \leq (\max |F'|)(b_i - a_i)$ by the Mean Value Theorem. More generally, on a compact set, we have that:

$$\text{absolutely continuous } \subseteq \text{ uniformally continuous } \subseteq \text{ continuous}$$

and on a compact interval:

$$\text{continosly differentiable } \subseteq \text{ Lipchitz continuous } \subseteq \text{ absolutely continuous } \subseteq \text{ Bounded Variation } \subseteq \text{ differentiable a.e}$$

where Lipchitz continuous means for all $x, y \in \mathbb{R}$, there exists an M such that $|f(x) - f(y)| \leq M|x - y|$. After going over the FTC for Lebesgue integrals, you can prove as an exercise that Lipchitz continuous implies abslutely continuous. Usually, Lipchitz continuity is used as a simple way of finding absolutely continuous functions.

To make sure this new notion links with the absolute continuity of the measure:

Proposition 3.5.3: Linking Absolute Continuity Of F And Measures

Let $F \in NBV$. Then F is absolutely continuous if and only if $\mu_F \ll m$

***Proof* :**

($\Leftarrow$) Let $\mu_F \ll m$. Then since μ_F is finite, we can apply theorem 3.2.1 to the sets $E = \bigcup_i^N (a_i, b_i]$, and get that F is absolutely continuous

($\Rightarrow$) Let E be a borel set such that $m(E) = 0$. Let $\epsilon > 0$ so that we can choose a δ that satisfies the definition of absolute continuity of F. By theorem 1.4.2, we can find open sets $U \supseteq U_1 \cdots \supseteq E$ such that $m(U_1) < \delta$ (which also implies $m(U_i) < \delta$ by monotonicity) and $\mu_F(U_i) \to \mu_F(E)$. Since each U_i is the countable union of disjoint intervals $(a_{i,j}, b_{i,j})$, for all $N \in \mathbb{N}$:

$$\sum_k^N |\mu_F((a_{i,j}, b_{i,j}))| \leq \sum_k^N |F(b_{i,j}) - F(a_{i,j})| < \epsilon$$

by absolute continuity. Letting $N \to \infty$, we get that $|\mu_F(U_i)| < \epsilon$, and so $|\mu_F(E)| < \epsilon$. Since this is true for all epsilon, it must be that $|\mu_F(E)| = 0$, showing that $\mu_F \ll m$, as we sought to show

With this, we can relate L^1 functions to NBV functions:

Corollary 3.5.2: Relating L^1 And NBV Functions

Let $f \in L^1$. Then the function $F(x) = \int_{-\infty}^{x} f(t)dt$ is in NBV, is absolutely continuous, and $f = F'$ a.e. Conversely, if $F \in NBV$, is absolutely continuous, then $F' \in L^1(m)$ and $F(x) = \int_{-\infty}^{x} F'(t)dt$

***Proof* :**
This follows from proposition 3.5.2 and 3.5.3.

If the function is bounded, then we get a slightly stronger result:

Lemma 3.5.3

If F is absolutely continuous on $[a, b]$, then $F \in BV([a, b])$

***Proof* :**
We will find an upper bound for the total variation of F, that is, we'll find an N such that $T_f(b) < N$. Choose $\epsilon = 1$ and let δ be the appropriate choice to satisfy absolute continuity: for any finite set of disjoint intervals (a_i, b_i):

$$\sum(b_i - a_i) < \delta \Rightarrow \sum |F(b_i) - F(a_i)| < \epsilon$$

Choose N to be the greatest integer that is smaller than $\delta^{-1}(b-a)+1$. Then, for any subdivision $a = x_0 < x_1 < \cdots < x_n = b$, we can collect the subintervals (x_i, x_{i+1}) into at most N groups such that the sum of each group is less than δ (adding subdivision's if necessary). Then the sum $\sum |F(x_{i+1}) - F(x_i)|$ over each group is 1, and so the total variation of F on $[a, b]$ is at most N

Is the converse true: If $F \in BV$, then does it imply that F is aboslutely continuous? The answer is no. Even better, F can even be NBV, uniformally continuous, and increasing (so differentiable a.e.), but still not absolutely continuous:

Example 3.7: NBV, Continuous, But Not Absolutely Continuous
Let F be the cantor function. Note that F is increasing, $F(-\infty) = 0$, and $T_F(\infty) = 1/2$ (since it's continuous, it is integrable, and integarting give 1/2). Thus μ_F is defined. However, it is not absolutely continuous with respect to m. However, F is not absolutely continuous (you should prove this).

Theorem 3.5.3: Fundamental Theorem Of Calculus For Lebesgue Integrals

Let $-\infty < a < b < \infty$ and $f : [a, b] \to \mathbb{C}$. Then the following are equivalent:

(a) F is absolutely continuous

(b) $F(x) - F(a) = \int_a^x f(t)dt$ for some $f \in L^1([a, b], m)$

(c) F is differentiable a.e. on $[a, b]$, $F' \in L^1([a, b], m)$ and $F(x) - F(a) = \int_a^x F'(t)dt$

Proof **:**
First, to show $(a) \Rightarrow (c)$, we can subtract some constant from F so that $F(a) = 0$. Doing the usual extension of setting $F(x) = 0$ for $x < a$ and $F(x) = F(b)$ for $x > b$, then $F \in NBV$. Thus, by lemma 3.5.3 using corrollary 3.5.2. Then, that $(c) \Rightarrow (b)$ comes immetidaly since (c) is a stronger condition than (b). For $(b) \Rightarrow (a)$, if we set $f(t) = 0$ for $t \notin [a, b]$ then $f \in L^1(m)$ and so by corrollary 3.5.2 we get the final result.

If $F \in NBV$, then we usually write an integral of a function g with respect to the associated μ_F as $\int gdF$ or $\int g(x)dF(x)$. These types of integrals are called *Lebesgue-Stieltjes integrals*. I'm not sure why they were introduced, but this theorem was presented for generalizing integration by parts to these types of integrals

Theorem 3.5.4: Integration By Parts For Lebesgue-Stieltjes Integrals

Let $F, G \in NBV$ where at least one of them is continuous. Then for $-\infty < a < b < \infty$

$$\int_{()a,b]} FdG + \int_{(a,b]} GdF = F(b)G(b) - F(a)G(a)$$

***Proof* :**

F and G are linear combinations of increasing functions in NBV by Theorem 3.27(a,b), so a simple calculation shows that it suffices to assume F and G increasing. Suppose for the sake of definiteness that G is continuous, and let $\Omega = \{(x, y) : a < x \le y \le b\}$. We use Fubini's theorem to compute $\mu_F \times \mu_G(\Omega)$ in two ways:

$$\begin{aligned}\mu_F \times \mu_G(\Omega) &= \int_{(a,b]} \int_{(a,y]} dF(x)\, dG(x) = \int_{(a,b]} [F(y) - F(a)]\, dG(y) \\ &= \int_{(a,b]} F\, dG - F(a)[G(b) - G(a)],\end{aligned}$$

and since $G(x) = G(x-)$,

$$\begin{aligned}\mu_F \times \mu_G(\Omega) &= \int_{(a,b]} \int_{[x,b]} dG(y)\, dF(x) = \int_{(a,b]} [G(b) - G(x)]\, dF(x) \\ &= G(b)[F(b) - F(a)] - \int_{(a,b]} G\, dF.\end{aligned}$$

Subtracting these two equations, we obtain the desired result. ∎

Finally, we end this chapter with a quick discussion about different terminology when it comes down to decomposing measures.

Definition 3.5.6: Discrete And Continuous Measures

Let μ be a complex Borel measure on $\mathbb{R}^n$

1. μ is called *discrete* if if there is a coutnable subset $\{x_i\}_i \subseteq \mathbb{R}^n$ such that $\sum_i c_i < \infty$ and $\mu = \sum c_i \delta_{x_i}$ where $\delta_{x_i}(x_i) = 1$ and is zero everywhere else (i.e. the pointmass at x).
2. μ is called *continuous* if for all $x \in \mathbb{R}^n$, $\mu(\{x\}) = 0$

It is easy to see that any measure can be broken down into $\mu = \mu_d + \mu_c$ where μ_d is discrete and μ_c is continuous. The main thing is to insure that μ_d is over a countable set: let $E = \{x \mid \mu(\{x\}) \neq 0\}$. Taking any countable subset $F \subseteq E$, then $\sum_{x \in F} \mu(\{x\})$ converges absolutely to $\mu(F)$ so $\left\{x \in E \;\middle|\; \mu(\{x\}) > k^{-1}\right\}$ is finite, and so it must be that E is countable, so $\mu_d(A) = \mu(A \cap E)$ is discrete and $\mu_c(A) = \mu(A \setminus E)$ is continuous.

If μ is discrete, then $\mu \perp m$, and if μ is continuous then $\mu \ll m$. Thus, we can further break down the decomposition into:

$$\mu = \mu_d + \mu_{ac} + \mu_{sc}$$

where μ_d is the discrete part, μ_{ac} is the absolutely continuous part, and μ_{sc} is the *singularly continuous* part. So far, we have mainly worked with measures with zero singularly continuous parts. There do exist measures with nonzero singularly continuous parts (Folland says the surface measure of the sphere for $n > 1$ is an example). An example for $n = 1$ is rarer, but they do exist. By proposition 3.5.2, we know that we are looking for a corresponding $F \in NBV$ such that $F' = 0$ a.e. In fact, the cantor function is such an example (when extended in the way we keep doing to make sure functions are in NBV). In fact, using the cantor function, we can construct an increasing function F that is continuous, $F' = 0$ a.e., and that is *strictly increasing*:

Example 3.8: Interesting Singularly Continuous Function

1. If $\{F_i\}_{i=1}^{\infty}$ is a sequence on nonnegative increasing functions on $[a, b]$ such that $F(x) = \sum_1^{\infty} F_i(x) < \infty$ for all $x \in [a, b]$, then $F'(x) = \sum_1^{\infty} F_i'(x)$ for a.e. $x \in [a, b]$ (It suffices to assume $F_i \in NBV$. Consider the measure μ_{F_i})

Proof **:**
First, notice that since the F_i are nonnegative, by letting $F(x) = 0$ for $x < a$ and $F(b) = F(y)$ for $b < y$, we have that $F_i(-\infty) = 0$. Furthermore, the the function $G(x) = F(x+)$ and $G_i(x) = F_i(x+)$ is equal a.e. with the same derivative wherever it is defined (which it is a.e.), and so if G and G_i has a particualr derivative, so do F and F_i. Thus, Without loss of generality, we'll assume $F, F_i \in NBV$.

Since they are in NBV, consider the induced complex borel measure μ_F and μ_{F_i} given by theorem 3.29. Since these measures are finite, they are σ-finite, and so admit a Radon-Nykodym derivative:

$$d\mu_F = f dm + d\lambda \qquad d\mu_{F_i} = f_i dm + d\lambda_i$$

where $f dm \ll m$, $f_i dm \ll m$ and $d\lambda \perp m$, $d\lambda_i \perp m$. We will show that $f dm = \sum f_i dm$. First, we need to show that $\mu_f = \sum \mu_{F_i}$, and then use the previous homework exercise (question 8) to conclude the equality.

By definition of F, $\mu_F((-\infty, b]) = \sum_i^\infty \mu_{F_i}((-\infty, b])$. Since the two measures agree on a generating set (closed rays form a generated set), they agree everywehre, meaning $\mu_F = \sum_i^\infty \mu_{F_i}$. Thus, we have that (using absolute convergence to re-arrange the sum)

$$f dm + d\lambda = d\mu_F = \sum_{i=1}^{\infty} d\mu_{F_i} = \sum f_i dm + \sum \lambda_i$$

Since $(\sum d\lambda_i) \perp m$ and $(\sum f_i dm) \ll m$, by question 8:

$$d\lambda = \sum d\lambda_i \qquad f dm = \sum f_i dm$$

Finally, we can invoke theorem 3.22 to connect the information about f and f_i to derivative information. By letting $E_r = (x, x + r]$ (since the F and F_i's are always positive), we get:

$$\begin{aligned}
F'(x) &= \lim_{r\to 0} \frac{F(x+r) - F(x)}{r} \\
&= \lim_{r\to 0} \frac{\mu_F(E_r)}{m(E_r)} \\
&= f(x) && \text{thm. 3.22} \\
&= \sum f_i(x) && m\text{-a.e.} \\
&= \sum \left(\lim_{r\to 0} \frac{\mu_{F_i}(E_r)}{m(E_r)} \right) && \text{thm 3.22} \\
&= \sum \left(\lim_{r\to 0} \frac{F_i(x+r) - F_i(x)}{r} \right) \\
&= \sum F'(x)
\end{aligned}$$

as we sought to show.

2. Let F denote the cantor function on $[0, 1]$ (see 1.5), and set $F(x) = 0$ for $x < 0$ and $F(x) = 1$ for $x > 1$. Let $\{[a_n, b_n]\}$ be an enumeration of the closed subintervals of $[0, 1]$ with rational

endpoints, and let $F_n(x) = F\left(\frac{x-a_n}{b_n-a_n}\right)$. Then $G = \sum_1^\infty 2^{-n}F_n$ is continuous and strictly increasing on $[0,1]$ and $G' = 0$ a.e. (use exercise 39)

Proof **:**
We break this proof down in many steps.

First, F is differentiable a.e. Notice that F is increasing and bounded, and so by theorem 3.27(e) F is a.e. differentiable. Notice that we can extend F as $F(x) = F(0)$ $x \leq 0$ and $F(1) = F(y)$ $1 \leq y$. In this way, $F \in NBV$.

Next, $F' = 0$ a.e. To see this, notice that by definition of F, the function is constant on C^c where C is the cantor set. Since the Cantor set is closed, C^c is open. On any open interval of C^c, F is constant, and so $F' = 0$ on C^c.

As a consequence, every F_n has zero derivative almost everywhere. At any point where F is differentiable:

$$F_n' = F'\left(\frac{x-a_n}{b_n-a_n}\right)\left(\frac{x-a_n}{b_n-a_n}\right)' = 0\left(\frac{x-a_n}{b_n-a_n}\right)' = 0$$

Showing that F_n' is a.e. 0.

Using this result, we can apply exercise 39 to get $G' = \sum_n 2^{-n}F_n$. Since each $2^{-n}F_n$ is a non-negative increasing function has derivative 0 a.e.:

$$G' = \sum_{n=1}^\infty 2^{-n}F_n' = 0 \qquad \text{a.e.}$$

Which gives us our first result.

Next, let's prove G is continuous. To see this, let $G_n = \sum_{k=1}^n 2^{-k}F_k$. Notice that $2^{-k}F_k$ is continuous being the product of two continuous functions (F_k is continuous being the composition of two continuous functions). By definition of F_k, the largest value if 1, and so $\left|\sum_1^n 2^{-k}F_n\right| \leq \sum_1^n 2^{-k} = 1 - 2^{-n}$. Then we can combine these two pieces of information to show that G_n uniformally converges to G. In particular, for all $\epsilon > 0$, chosoe $n > \frac{1}{2^\epsilon}$ so that $|G - G_n| < \epsilon$ for all $x \in [0,1]$. Thus, G is continuous.

Finally, to show G is strictly increasing, choose $x, y \in [a,b]$ such that $x < y$. Since x is strictly less than y, there exists an ϵ such that $x \notin [y-\epsilon, y]$. Since the rational's are dense, we can choose $[a_n, b_n]$ from our enumeration such that $x < a_n < b_n < y$ so that $x, y \notin [a_n, b_n]$. As a consequence, notice that $G(x)$ must be at least 2^{-n} away from $G(y)$, since:

$$\sum_n^\infty 2^{-n}F_n(x) < \sum_n^\infty 2^{-n}F_n(y)$$

and so G is strictly increasing.

I also want to give a quick word on convexity. I'd like to write more about convex functions, like their definition and some properties, but for now I'll just write down in bullet form what I found interesting:

1. Convex functions on compact intervals are absolutely continuous

2. As a consequence, FTC applies. In fact convex functions can be characterized like so : f is convex if and only if f' is strictly increasing. If f is twice differentiable, this leads to the common notion of $f''(x) \geq 0$

3. In a sense, linear functions are a sepcial type of convex function. Convex functions are closed under addition and sclara multiplication. Thus, linear function being both convex and concave mean they are closd under subraction, which is why they form a vector-space, but convex functions do not. The inequalities of linear function are also all “trivial”, preciesly because of linearity, while those of convex functions are more complex.

Chapter 4

Function Analysis

(This entire article is so enlightening: Intuition on Function Spaces)

Recall in section 2.3 that the space L^1 is a vector-space. As an exercise, it should be shown that it is an infinite dimensional vector-space[1]. Since it is a real vector-space, it is isomorphic to $\mathbb{R}^N$ for appropriate cardinality N. The question becomes "how would we define a measure on $\mathbb{R}^N$"? This is quite a different question, and we usually don't meaure a subset of L^1 (or function space more generally), but instead assign a meaure to a single function, and then define a notion of how "close" other functions are to this function.

The term Functiona Analysis or Functional Analysis is the traditional name for the study of infinte-dimensional vector-spaces over $\mathbb{R}$ or $\mathbb{C}$ and the linear maps between them. When in finite dimension, we have a lot of very nice algebraic results such as: Let V be a finite dimensional vector space,

1. If $W \subseteq V$ and $f \in W^*$, then f can be extended to be a linear functional on V
2. Bijective homomorphisms are isomorphisms
3. Others

What makes functional analysis different form linear algebra over finite dimensional vector space is that in finite dimension, there is only one "resonable" topology that makes linear functions automatically continuous. This is not the case for ininite dimensional spaces (recall the amount of interpretations we saw of how a sequence $\{f_n\}$ of function on $\mathbb{R}$ can convege $f_n \to f$). This chapter will be an introduction to this world.

(Why do we want maps to be continuous? Is it that we saw that many spaces are Banach spaces, and we want to thus apply algebraic results, but they are infinite dimensional meaning many finite-dimensional results don't hold, and so we limit our attention to "continuous" continuous functions to fix this?)

(When does a metric induce a norm: `https://math.stackexchange.com/questions/166380/not-every-metric-is-induced-from-a-norm`)

(I don't konw if this is the right place to put this, but it is really really really cool. It is bassically the classical duality between algebra and geoemtry. Given a topological space X, we can define

[1]answer: the functions $\chi_{[n,n+1)}$ are linearly independent for all $n \in \mathbb{N}$

an algebra by taking the set of all continuous functions $C(X)$ with the operations being pointwise. Now, given just the information $C(X)$, does there exist a X such that $C(X)$ is isomorphic to A? In other words, can we recover X? The answer is yes if X is a locally compact topological space! This is the algebra/topology duality. The same is true if X is a Riemann manifold, though you need more information, something called the *spectral triple*. Later, Grothendick generlized this and we got to create form any algebra A a topological space $\operatorname{Spec}(A)$!!! Given the appropriate topological properties on $\operatorname{Spec}(A)$, we can recover A!! In other words, there is this way to go back and forth between algebra and geometry. This same exists in analysis too: the dual of $C_c(X)$ for a LCH X is intimately related to Radon measure, and the dual of $C_c(X)$ with the appropriate conditions let's us generlize the notion of derivatives to locally integrable functions!!!)

4.1 Normed Vector Spaces

One of the key problems a norm tries to introduce is making sure there is still a reasonable "finiteness" to your space. Though a Normed vector-space doesn't have a ordering, there is something simlar to it: in the same way the distance between any two real numbers if a real numbers (i.e. finite) the *norm* of any element in a Banach space is *finite*. Thus, though a space might be infinite like L^1, there is still some notion of finiteness allowing us to bring in many of the concepts we have worked with before and avoid the pathologies of the infinite.

Let k be either $\mathbb{R}$ or $\mathbb{C}$, let V or X denote a vector-space over k, with 0 representing the zero vector. Let kx denote the subspace of X spanned by x. If $M, N \subseteq X$, then $M + N = \{m + n \mid m \in M, n \in N\}$.

Definition 4.1.1: Semi-Norm And Norm

Let V be a vector-space. A *semi-norm* on V is a function $x \mapsto \|x$ from V to $[0, \infty)$ satisfying the propeties:

1. $\|x + y\| \leq \|x\| + \|y\|$ (the triangle inequality)
2. $\|\lambda x\| = |\lambda|\|x\|$ (homogeneity of scalars)

If furthermoer:

(3) $\|x\| = 0$ if and only if $x = 0$

then the function is called a *norm*

From the axioms of a semi-norm, $\|0\| = 0$. There are norms for which $x \neq 0$ but $\|x\| = 0$, as we'll soon show.

Definition 4.1.2: Normed Vector-Space

Let V be a vector space. Then if there exists a function $\|\cdot\|$ that is defined on V, then $(V, \|\cdot\|)$ is called a *normed vector-space.*

If V is a normed vector-space, we can induce a metric $\rho(x, y) = \|x - y\|$, namely since:

$$\|x - z \leq \|x - y\| + \|y - z\| \qquad \|x - y\| = \|(-1)(x - y)\| = \|y - x\|$$

Since any metric induces a topology, the induced topology by the metric induced by a norm $\|\cdot\|$ is called the *norm topology* on V. A nice consequence of the induced norm topology is that addition and scalar multiplication is continuous. Two norms $\|\cdot\|_1$ and $\|\cdot\|_1$ are said to be *equivalent* if there exits some $C_1, C_2 > 0$ st

$$C_1\|\cdot\|_1 \leq \|\cdot\|_2 \leq C_2\|\cdot\|_1$$

Equivalent norms define equivalent metricsand ehcne the same topology and so the same Cauchy sequences. As a different perspective on norms, since their codomain is $\mathbb{R}$

Definition 4.1.3: Banach Space

Let V be a normed vector space. Then V is a *Banach space* if it is complete with respect to it's induced metric.

Example 4.1: Banach Spaces

1. Every finite dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$ is a banach space. In this space, all linear map are continuous.
2. Let K be a compact space, and take $C_0(K)$. Then $C_0(K)$ is a Banach space with the operator norm (see definition 4.1.6). We will later show (Banach–Mazur theorem) that all real separable banach spaces are isometric (the isomrophisms in the category of banach spaces) to a subsapce of $C_0([0, 1])$, giving us a universal way of thinking about Banach spaces. If the space is not separable, if the smallest dense set D of a Banach space X has density of cardinality α, then X is isometric to $C_0([0, 1]^\alpha, \mathbb{R})$[a]
3. If X is a topological space, then $B(X, k)$ and $BC(X, k)$ (all bounded functionals and bounded continuous funcitonals) form a Banach sapce with the uniform norm $\|f\|_u = \sup_{x\in X} |f(x)|$.
4. For p where $1 \leq p \leq \infty$, the space $L^p(X)$ for some measurable space X is a Banach space with the L^p-norm (recall that $L^p(\mu)$ is the set of equivalence classes functions that agree a.e.; if $L^p(\mu)$ represented only functions, then $\|\cdot\|_p$ would be a semi-norm)
5. Two other common examples of Banach spaces we will encounter later will be special spaces of differentiable functions, Banach Algebras, and Hilbert Spaces.

There are also examples of spaces that are not Banach Spaces:

1. Let $U \subseteq \mathbb{R}^n$ be open and take $C_0(U, \mathbb{C})$. Then This space is *not* a Banch space
2. The space $H(U, \mathbb{C})$ of all holomorphic functions from U to $\mathbb{C}$ is *not* a Banach Space
3. The sapce $C_c^\infty(\mathbb{R}^n)$ is *not* a Banach space
4. (If you know some PDE's) The sapce of distributions $\mathcal{D}'$ is *not* a Banach Space

These are all examples of a weaker structure known as a *topological vector space* which we shall cover in section 4.4.

[a]An interesting result using Banach-Mazur's theorem by Luis Rodríguez-Piazza shows that the space of smooth functions with the uniform norm is isometric to the space of nowhere differentiable functions

There is a nice way to check whether a vector space V is a Banach space which requires the following definition:

Definition 4.1.4: Absolutely Convergent

Let V be a normed vector space. and $\{x_k\}_k^\infty \subseteq V$ a sequence. Then the sequence is said to *converge absolutely* if

$$\sum_k^\infty \|x_k\| < \infty$$

Lemma 4.1.1: Cauchy and Absolute Convegence

Let V be a normed vector space. Then V is complete if and only if every absolutely convergent series in V converges.

Proof **:**

Let's first say that X is complete and that $\sum_{n=1}^\infty \|x\| < \infty$. To show it converges, we take advantage of completeness and show there is a Cauchy sequence. In particular, let $S_n = \sum_{k=1}^n x_k$ be the partial sum. Then since the series is absolutely convergent, there for all $n > m$,

$$\|S_n - S_m\| \leq \sum_{m+1}^n \|x_k\| \to 0 \quad \text{as } m, n \to \infty$$

so the sequence $\{S_n\}$ is Cauchy, and so converges, so $\{x_k\}$ converges.

Conversely, let's say an absolutely convergent series converges, and choose any Cauchy sequence $\{x_k\}$. Since it's a Cauchy sequence, choose a subsequence such that

$$\left\|x_{k_{i+1}} - x_{k_i}\right\| < 2^{-i}$$

Take $y_1 = x_{k_1}$ and $y_i = x_{i+1} - x_i$ for $i > 1$ so that $\sum_{i=1}^n y_i = x_{k_n}$. Then notice that $\{y_i\}$ is in fact absolutely convergent:

$$\sum_{i=1}^\infty \|y_i\| \leq \|y_1\| + \sum_{i=1}^\infty 2^{-i} = \|y_1\| + 1 < \infty$$

So $\sum_{i=1}^\infty y_i = \lim x_{k_n}$ exists. Since the subsequence of every Cauchy sequence approaches the same "point" (just like convergent sequences), we see that $\{x_n\}$ also converges to the same point, showing that Cauchy sequences are in fact convergent sequences, as we sought to show.

Next, we see what happens with common vector-space constructions like products and quotients. There are 3 equivalent norms we can put on the product of two normed vector-spaces:

1. $\|(x, y)\| = \max(\|x\|, \|y\|)$
2. $\|(x, y)\| = \|x\| + \|y\|$
3. $\|(x, y)\| = (\|x\|^2 + \|y\|^2)^{1/2}$

For quotients V/W, we usually define the *quotient norm*:

$$\|x+W\| = \inf_{w\in W} \|x+w\|$$

Next, we would want to study which linear functions between two vector-sapces are *continuous*. It will *not* be the case that all linear functions will be continuous. The following builds up to the classification of continuous linear functions:

Definition 4.1.5: Bounded Linear Map

Let $T : X \to Y$ be a linear map between two normed vector space. Then if there exists some C such that for all $x \in X$

$$\|T(x)\| \leq C\|x\|$$

Then T is said to be *bounded*, or is a *bounded linear operator*.

Note how this is different then boundedness of a set (ex. the range is bounded if $\|T(x)\| \leq C$). This definition is usefless to us since no nonzero linear map would satisfy it. Furthermore, all linear map over finite dimensional vector spaces clearly satisfy the condition (can you find the C?), so this concept gets interesting in infinite dimesion

Proposition 4.1.1: Bounded $\Rightarrow$ Continuous

Let $T : X \to Y$ be a linear map between two normed vector spaces. Then the following are equivalent:

1. T is continuous
2. T is continuous at 0
3. T is bounded

***Proof* :**
If T is the zero map, then we're done, so assume $T \neq 0$

(1) implies (2) by definition. For (2) implies (3), Since T is continuous at 0, for for all $\epsilon > 0$, there exists a $\delta > 0$ such that if $\|x - 0\| < \|x\| < \delta$, then $\|T(x) - T(0)\| = \|T(x) - 0\|\|T(x)\| < \epsilon$. Choosing $\epsilon = 1$, we get a δ_1 such that

$$\|x\| < \delta_1 \Rightarrow \|T(x)\| < 1$$

we know want to show there exists a C such that $\|T(y)\| \leq C\|y\|$ for all y in the domain. We will do this by taking advantage of linearity and shrinking element so that we can take advantage of continuity.

Define $x(y) = \frac{\delta_1}{2\|y\|}y$ so that

$$\|x(y)\| = \left\|\frac{\delta_1}{2\|y\|}y\right\| = \frac{\delta_1}{2}\frac{1}{\|y\|}\|y\| = \frac{\delta_1}{2} < \delta_1$$

therefore, by continuity and linearity:

$$\|T(x(y))\| = \left\|T\left(\frac{\delta_1}{2\|y\|}y\right))\right\| = \frac{\delta_1}{2\|y\|}\|T(y)\| < 1$$

since this is true for all y, we if we re-arrange we have:

$$\|T(y)\| < \frac{2}{\delta_1}\|y\|$$

showing that T is indeed bounded.

Finally, if $\|T(x)\| \leq C\|x\|$, then for all $\epsilon > 0$, choose $\delta = C^{-1}\epsilon$ so that

$$\|T(y) - T(x)\| = \|T(y-x)\| \leq \epsilon$$

completing the proof.

We will denote the set of bounded linear functions (which we will call linear operators) as $L(X,Y)$, $\mathcal{L}(X,Y)$, or $\mathrm{Hom}_k(X,Y)$. This space is itself a normed space with the following norm:

Definition 4.1.6: Operator Norm

$L(X,Y)$ is a normed vector space with norm:

$$\begin{aligned}\|T\| &= \sup \|T(x)\| \|x\| = 1 \\ &= \sup \frac{\|T(x)\|}{\|x\|} \|x\| \neq 0 \\ &= \inf \{C \mid \|T(x)\| \leq C\|x\| \text{ for all } x\}\end{aligned}$$

If $\|T(x)\| = \|x\|$, then T is called an *isometry*

It is useful to establish when is $L(X,Y)$ a Banach space

Proposition 4.1.2: Completness Of $L(X,Y)$

Let X, Y be normed vector-space and $L(X,Y)$ the normed vector-space of all bounded linear functions from X to Y. If Y is complete, so is $L(X,Y)$.

Proof **:**
This is essentially done by defining a function for which a Cauchy sequence T_n can converge too (Folland p. 154)

Finally, we can find a way to multiply functions. if $T \in L(X,Y)$, and $S \in L(Y,Z)$, then

$$\|ST(x)\| \leq \|S\|\|T(x)\| \leq \|S\|\|T\|\|x\|$$

and so $ST \in L(X,Z)$. In particular, that makes $L(X,X)$ into an algebra, and when X is complete, $L(X,X)$ is a *Banach Algebra*. Note that for $T \in L(X,Y)$ to be an isomorphism, we require that T^{-1} also be bounded (example HERE). If $\|T(x)\| = \|x\|$ for all x, then T is called an *isometry*.

4.2 Linear Functionals

This looks cool! Will get back to this antoher day (Folland p. 157)

4.3 Baire Category Theorem

the open mapping theorem and closed graph theorem are really cool consequences

4.4 Topological Vector Space

(This post seems like an excellent post to follow! link here)

There are many topologies on which the operations are continuous (i.e. + and scalar multiplicaiotn is continuous) that do not arrise from norms. We examine these here

Definition 4.4.1: Topological Vector Space (TVS)

Let V be a vector-space over $\mathbb{R}$ or $\mathbb{C}$. Then V is a *topological vector space* (TVS) if the operation of addition and scalar multiplication are continuous. A topological vector-space is said to be *locally convex* if there is a base for the topology consisting of convex sets.

The most common way of forming TVS is through a collection of semi-norms:

Theorem 4.4.1: Semi-Norms Generate TVS

Let $\{p_\alpha\}_{\alpha \in A}$ be a family of norms on a vector space V. If $x \in X$, $\alpha \in A$, and $\epsilon > 0$, let

$$U_{x,\alpha,\epsilon} = \{y \in V \mid p_\alpha(y - x) < \epsilon\}$$

and let $\mathcal{T}$ be the topology generated by the sets $U_{x,\alpha,\epsilon}$. Then:

1. For each $x \in X<$ the inite intersections of the sets $U_{x,\alpha,\epsilon}$ $(\alpha \in A, \epsilon > 0)$ form a neigborhood base at x
2. if $\langle x_i \rangle_{i \in I}$ is a net in V ,then $x_i \to x$ if and only if $p_\alpha(x_i - x) \to 0$ for all $\alpha \in A$.
3. $(V, \mathcal{T})$ is a locally convex topological vector-space

***Proof* :**
Folland p. 166

(theorems on how to characterize continuous maps and when a spac is hausdorff, and also if it is hausdorff, it is equivalent to define a translation-invariant metric)

(build-up)

Definition 4.4.2: Fréchet Space

A complet eHausdorff topological vector-space whose topology is defined by a countable family of seminroms is called a *Fréchet space.*

Example 4.2: Topological Vector-Spaces

1. Let X be a LCH space. On $\mathbb{C}^X$, the topology of uniform convergence on compact sets is defined by the seminorms $p_K(f) = \sup_{x \in K} |f(x)|$ as K rangers over compact subsets of X. If X is σ-compact and $\{U_n\}$ are as in proposition ref;HERE (folland p. 168), this topology is defined by the seminorms $p_n(f) = \sup_{x \in \overline{U_n}} |f(x)|$. In this case, $\mathbb{C}^X$ is easly seen to be complete, so it is a Fréchet space. By proposition ref:HERE, so is $C(X)$.

2. The space $L^1_{\text{loc}}(\mathbb{R}^n)$ is a Fréchet space with the topology defined by the seminroms $p_k * f) = \int_{|x| \leq k} |f(x)|dx$ (completeness follows from completness of L^1).

3. Here is another useful place where TVS are defined. Let's say we want to study a space in which d/dx is continuous (and hence bounded) on $C^\infty([0,1])$. However, this is essentially impossible to define the topology through a norm. Indeed, if $f_\lambda(x) = e^{\lambda x}$, then $(d/dx)f_\lambda = \lambda f_\lambda$, and so $\|d/dx\| \geq \lambda$ for all λ, no matter what norm is used on $C^\infty([0,1])$. There are a few ways to rectify this:

   (a) One can consider differentiation as an unbounded operator form V to W where W is a suitable banace space and X is a dense subpsac eof W (exercise 30 in Folland)

   (b) One can consider differentiation as a bounded linear map from one Banach space V to a differnet one W such that $V = C^k([0,10])$ and $W = C^{k-1}([0,1])$ (execise 9)

   (c) One can consider differentiation as a continuous operator on a locally convex space V whose topology is not given by a norm. It is easy to construct families of seminorms on the space of smooth functions st differentiation becmeso continuous almost by definition. For example, the seminorms:

   $$p_k(f) = \sup_{0 \leq x \leq 1} |f^{(k)}(x)|$$

   where $k \in \in \{0, 1, 2, ...\}$ maeks $C^\infty([0,1])$ into a Fréchet space (the completeness is proved in exercise 9), and d/dx is continuous by proposition 5.15, since $p_k(f') = p_{k+1}(f)$. Other examples are in exer exercsie 45 in chatper 9 of Folland.

(Modify this so that it works for TVS)

Definition 4.4.3: Weak Convergence

Let $\{x_k\}_k^\infty \subseteq \mathcal{H}$ be a sequence in an inner product space. Then it is said to converge *weekly* to x if for any $y \in \mathcal{H}$:

$$\langle x_k, y \rangle \to \langle x, y \rangle$$

We denote this as $x_k \rightharpoonup x$ or $x_n \xrightarrow{\text{w}} x$

4.5 Hilbert Spaces

(Another intuition to incorporate: Since Hilbert spaces are self-dual due to Riesz represnetation theorem, they have a very simple duality theory. Think of how C^∞ does is not self-dual, and that's why distributions *can* be more copmlicated, or how $(L^p)^*$ has a very easy isomorphism to L^q. In the case of a hilbert space H, was have $H^* \cong H$, and so any extra information we might try to get form a dual (ex. notion of size, boundedness, oscillation, regulartiy, decay, etc.) can be represented by an element of the Hilbert space! This makes the dual theory easier to study!!)

(Another important property of Hilbert spaces is that we can always find a dense set that is *linear*. That is, any element $x \in H$ is equal to a convegent sequence $x = \sum_i^\infty x_i$. This is in a sense the best type of dense set!)

When dealing with Euclidean Geometry, we have many useful conceptions such as the triangle inequality, the notion of perpendicularity, the idea notion of "area" and the translation invariance of said area. We generalized these notion to vector spaces over $\mathbb{R}$ or $\mathbb{C}$, which preservd many of these intuition's in the finite-dimensional case. However, moving up to the infinite dimensional case, these intuitions are no longer necessarily preserved (examples here?). Furthermore, even in the finite dimensional case over $\mathbb{R}$ and $\mathbb{C}$, without a topology that permits the notion of completeness, we cannot do many of the constructions we've bene donig thus far that are dependent upon completion (essentially, evertyhing we have been working with).

The goal of this chapter is to be able to talk about infinite dimensional vector-spaces which still have some geometric notions, as well as some topological notions. Perhaps the most common such example would be functions spaces (some of which we will explore in this section), where many are infinite dimensional. If would be nice if our notion of geometry alos applied to them, and so the goal of this section is to see whether they do.

Note that the notion of a metric and a norm is not sufficiently strong enough to capture the geometrical ideas we will present: the inner product really is necessary to induce these ideas, as we shall soon see.

(also gotta integarte the two following things somehow) In the following, I will give results for when we are working over inner product spaces:

Definition 4.5.1: Weak Convergence in Hilbert Spaces

Let $\{x_k\}_k^\infty \subseteq \mathcal{H}$ be a sequence in an inner product space. Then it is said to converge *weekly* to x if for any $y \in \mathcal{H}$:

$$\langle x_k, y\rangle \to \langle x, y\rangle$$

We denote this as $x_k \rightharpoonup x$ or $x_n \xrightarrow{\text{w}} x$

We can define the operator norm on $\|T\|$ differently using the inner product:

$$\|T\| = \sup_{\|x\|=1,\ \|y\|=1} |\langle T(x), y\rangle|$$

Proof **:**
For $\geq$, using Cauchy-Swartz we get:

$$\sup_{\|x\|=1,\ \|y\|=1} |\langle T(x), y\rangle| \leq \|T(x)\| \leq \|T\|$$

For the other direction

$$\begin{aligned}\|T\| &= \sup_{\|x\|=1} \|T(x)\| \\ &= \sup_{\|x\|=1,\ T(x)\neq 0} \frac{\langle T(x), T(x)\rangle}{\|T(x)\|} \\ &\leq \sup_{\|x\|=\|y\|=1} |\langle T(x), y\rangle|\end{aligned}$$

4.5.1 Basic definition and properties

Definition 4.5.2: Inner Product Space

Let H be a complex vector space with an inner product $\langle \cdot, \cdot \rangle : H \times H \to \mathbb{C}$ where

1. $\langle x, y\rangle = \overline{\langle y, x\rangle}$
2. $\langle x + y, z\rangle = \langle x, z\rangle + \langle x, z\rangle$
3. $\langle ax, y\rangle = a\langle x, y\rangle$, $a \in \mathbb{C}$
4. $\langle x, x\rangle \in (0, \infty)$ for all nonzero $x \in H$
5. $\langle x, x\rangle = 0$ if and only if $x = 0$

We'll define $\|x\| = \sqrt{\langle, x, x\rangle}$. This is naturally a norm since:

1. $\|x\| = \sqrt{\langle x, x\rangle} = 0$ if and only if $x = 0$
2. $\|ax\| = \sqrt{\langle ax, ax\rangle} = \sqrt{a\bar{a}\langle x, x\rangle} = \sqrt{a^2\langle x, x\rangle} = |a|\sqrt{\langle x, x\rangle} = |a|\|x\|$

Proving the triangle inequality is a bit more difficult: IF we square both sides, we it is equivalent to show:

$$\|x + y\|^2 \leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2$$

expanding out the left hand side, we get:

$$\|x + y\|^2 = \|x\|^2 + 2\Re\langle x, y\rangle + \|y\|^2$$

so we have to show that $\Re\langle x, y\rangle \leq \|x\|\|y\|$. In fact, a bit more is true: $|\langle x, y\rangle| \leq \|x\|\|y\|$. Though we are using this inequality to prove the triangle inequality, notice that this in fact gives us a general bound on the value of the inner product! This inequality is one of the central tools used in the study of Hilbert spaces:

Theorem 4.5.1: Cauchy-Schwartz Inequality

$$|\langle x, y\rangle| \leq \|x\|\|y\|$$

***Proof* :**
If $\|y\| = 0$ then $y = 0$, and then $|\langle x, 0\rangle| = 0$, so assume $\|y\| \neq 0$. Then for all $t \in \mathbb{R}$

$$\begin{aligned} 0 &\leq \langle x - ty, x - ty\rangle \\ &= \|x\|^2 + 2t\Re\langle x, y\rangle + t^2\|y\|^2 \\ &= \|x\|^2 + \frac{2\Re\langle x, y\rangle^2}{\|y\|^2} + \frac{\Re\langle x, y\rangle^2}{\|y\|^2} & t = \frac{\Re\langle x, y\rangle}{\|y\|^2} \\ \Rightarrow |\Re\langle x, y\rangle| &\leq \|x\|\|y\| \end{aligned}$$

To finish the proof, pick $\alpha \in \mathbb{C}$ with $|\alpha| = 1$ such that

$$|\Re(\alpha\langle x, y\rangle)| = |\langle x, y\rangle|$$

Notice this does not break the inequality, and so we can repeat the entire process but with α taken into account with t, and so:

$$|\langle x, y\rangle \leq \|x\|\|y\|$$

as we sought to show

With this, the rest of the proof that $\|\cdot\|$ inducd by $\langle\cdot,\cdot\rangle$ is a norm follows

Theorem 4.5.2: Triangle Inequality For Induced Norm

$$\|x + y\| \leq \|x\| + \|y\|$$

***Proof* :**

$$\begin{aligned} \|x + y\|^2 &= \langle x + y, x + y\rangle \\ &\leq \|x\|^2 + 2\Re\langle x, y\rangle + \|y\|^2 \\ &\leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 \\ &= (\|x\| + \|y\|)^2 i \end{aligned}$$

The Cauchy-Shwartz inequality is the tool we use to bring in geometry as we know it into Hilbert space. For example, in real inner product spaces, we will define the angle between two vectors to be θ where θ satisfies:

$$\langle u, v\rangle = \cos(\theta)\|u\|\|v\|$$

Many other cool properties of the Cauchy-Shwartz inequality can be found here: (commented)

Now, notice we can define a metric by doing $\rho(x, y) = \|x - y\|$:

1. $\rho(x, x) = \|x - x\| = 0$ if and only if $x - x = 0$ so $x = x$
2. $\rho(x, y) = \|x - y\| \geq 0$
3. $\rho(x, y) = \|x - y\| = \|x - y + y - z\| \leq \|x - y\| + \|y - z\| = \rho(x, y) = \rho(y, z)$

Thus, we can define a natural topology on any pre-hilbert space. If that topology is complete, we give it a name:

Definition 4.5.3: Hilbert Space

Let H be an inner product space. Then H is a Hilbert space if it is *complete* with respect to the induced metric.

Example 4.3: Hilbert Spaces

1. $L^2(\mu)$ for a complex measure μ is a Hilbert space with the inner product being:

$$\langle f, g\rangle_{L^2} = \int_X f(x)\overline{g(x)}d\mu(x)$$

   each axiom is simple to verify:

2. $\mathbb{R}^n$ is a Hilbert space with the inner product being:

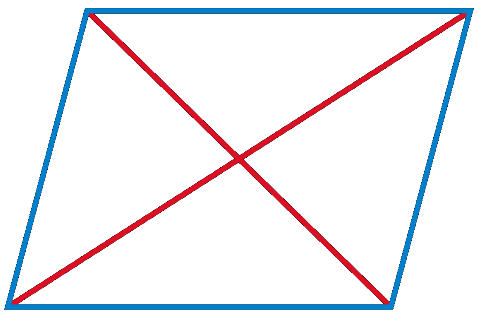
$$\langle x, y\rangle = \sum_i x_i y_i$$

Hilbert spaces are very euclidean in nature, for example:

Theorem 4.5.3: Parralelogram

$$\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2)$$

Visually, this menas that the "area" of a parralelogram is can be measured in two ways:



***Proof* :**

$$\begin{aligned}\|x+y\|^2+\|x-y\|^2 &= \langle x+y,x+y\rangle+\langle x-y,x-y\rangle \\ &= 2\langle x,x\rangle+\langle x,y\rangle-\langle x,y\rangle+\langle y,x\rangle-\langle y,x\rangle+2\langle y,y\rangle \\ &= 2\langle x,x\rangle+2\langle y,y\rangle \\ &= 2\|x\|^2+2\|y\|^2\end{aligned}$$

It turns out that a norm satisfying the parralelogram law will in fact define an inner product. This shows how norms define geometries that are more general than euclidean:

Example 4.4: Exploring Parralelogram Law

1. (The polarization Identity)$\mathcal{H}$ be a Hilbert space. For any $x, y \in \mathcal{H}$,

$$\langle x,y\rangle = \frac{1}{4}(\|x+y\|^2-\|x-y\|^2+i\|x+iy\|^2-i\|x-iy\|^2)$$

Proof. This is simply a matter of computation:

$$\begin{aligned}&\frac{1}{4}(\|x+y\|^2-\|x-y\|^2+i\|x+iy\|^2-i\|x-iy\|^2) \\ &=\frac{1}{4}(\langle x+y,x+y\rangle-\langle x-y,x-y\rangle+i\langle x+iy\rangle-i\langle x-iy,x-iy\rangle) \\ &=\frac{1}{4}(\langle x,x\rangle+\langle x,y\rangle+\langle y,x\rangle+\langle y,y\rangle-(\langle x,x\rangle-\langle x,y\rangle-\langle y,x\rangle+\langle y,y\rangle) \\ &\quad+i(\langle x,x\rangle+\langle x,iy\rangle+\langle iy,x\rangle+\langle iy,iy\rangle-(\langle x,x\rangle-\langle x,iy\rangle-\langle iy,x\rangle+\langle iy,iy\rangle)) \\ &=\frac{1}{4}(2\langle x,y\rangle+2\langle y,x\rangle+i(2\langle x,iy\rangle+2\langle iy,x\rangle)) \\ &=\frac{1}{4}(2\langle x,y\rangle+2\langle y,x\rangle+i(-i)2\langle x,y\rangle+2ii\langle y,x\rangle) \\ &=\frac{1}{4}(2\langle x,y\rangle+2\langle y,x\rangle+2\langle x,y\rangle-2\langle y,x\rangle) \\ &=\frac{1}{4}4\langle x,y\rangle \\ &=\langle x,y\rangle\end{aligned}$$

□

2. If $\mathcal{H}'$ is another Hilbert space, a linear map from $\mathcal{H}$ to $\mathcal{H}'$ is unitary if and only if it is isometric and surjective

Proof. Let U be unitary so that $\langle Ux,Uy\rangle_2 = \langle x,y\rangle_1$ for any $x, y \in \mathcal{H}$. Letting $x = y$ gives us $\|Ux\|_2 = \|x\|_1$, giving us that it's an isometry. Since U is invertible, it is surjective, and so U is a surjective isometry.

Conversely, let U be a surjective isometry. We need to show its invertible (it suffices to show it's injective), an isometry, and that U and U^{-1} are bounded. Clearly, U is bounded since

$\|U(x)\| \leq 1\|x\|$. For injectivity, let $Ux = Uy$, so $Ux - Uy = 0$. Taking the norm on both sides, we get:

$$0 = \|0\|_2 = \|Ux - Uy\|_2 = \|U(x-y)\|_2 = \|x-y\|_1$$

where the last equality is due to U being an isometry. Since $\|z\| = 0$ if and only if $z = 0$, we have $x - y = 0$, i.e. $x = y$, proving that U is injective. Along with surjectivity, that makes U invertible. With this, we can also show that U^{-1} is bounded: for any $y \in \mathcal{H}'$, there exists an $x \in \mathcal{H}$ such that $U(x) = y$. Thus:

$$\|U^{-1}(y)\| = \|x\| \stackrel{!}{=} \|U(x)\| = \|y\|$$

where $\stackrel{!}{=}$ comes from U being an isometry, showing that $\|U^{-1}(y)\| \leq 1 \cdot \|y\|$, making it also bounded.

Finally, to show that $\langle Ux, Uy\rangle_2 = \langle x, y\rangle_1$, we will use part (a):

$$\begin{aligned}
\langle x, y\rangle &= \frac{1}{4}(\|x+y\|^2 - \|x-y\|^2 + i\|x+iy\|^2 - i\|x-iy\|^2) \\
&\stackrel{!}{=} \frac{1}{4}(\|U(x+y)\|^2 - \|U(x-y)\|^2 + i\|U(x+iy)\|^2 - i\|U(x-iy)\|^2) \\
&= \frac{1}{4}(\|U(x)+U(y)\|^2 - \|U(x)-U(y)\|^2 + i\|U(x)+iU(y))\|^2 - i\|U(x)-iU(y)\|^2) \\
&= \langle U(x), U(y)\rangle
\end{aligned}$$

where $\stackrel{!}{=}$ comes from U being an isometry. Thus:

$$\langle Ux, Uy\rangle = \langle x, y\rangle$$

As we sought to show. □

The next concept we consider is a way of defining two vectors being somehow "unrelated" or going in some notion of "direction" that doesn't give information for the other vector:

Definition 4.5.4: Orthogonal

We say that x is *orthogonal* to y if $\langle x, y\rangle = 0$. We will write $x \perp y$

If two vectors are orthogonal, the triangle inequality becomes instant:

Theorem 4.5.4: Pythagoras

$$x \perp y \Rightarrow \|x+y\|^2 = \|x\|^2 + \|y\|^2$$

Proof **:**

$$\|x+y\|^2 = \langle x+y, x+y\rangle = \|x\|^2 + \|y\|^2$$

We might wonder if we take $E \subseteq H$, can we find all vectors in H that are somehow "dual" to E?

Define it's *orthogonal compliment* to be:

$$E^\perp = \{x \in H \mid \langle x, y\rangle = 0 \ \forall y \in E\}$$

Notice that E is *always* a closed subspace:

Proof. The fact it's a subspace follows from linearity of the second component.t
for closed, suppose there is a sequecne $\{x_n\} \subseteq E^\perp$ where $x_n \to x \in H$. Then:

$$|\langle x, y\rangle| \leq |\langle x - x_n, y\rangle| \leq \|x - x_n\|\|y\| \to 0 \quad \text{as } n \to \infty$$

□

We can also find closed (and not closed) subspace of hilbert spaces:

Example 4.5: Subspaces Examples

1. If $\mathcal{H}$ is a finite-dimensional Hilbert space, then all subspace are closed
2. If $\mathcal{H}$ is infinite-dimensional, consider $\mathcal{H} = L^1([0,1])$ with the L^2 inner product. and take $P([0,1]) \subseteq L^1([0,1])$ to be the subspace of all polynomial functions. This is is certainly not equal to $L^1([0,1])$, however polynomial functions are dense in $L^1([0,1])$ (they can arbitraily approximate continuous bump characteristic functions, and hence simple functions). However, their closure is $L^1([0,1])$, showing that $P([0,1])$ cannot be closed.
3. Assume $E \subseteq \mathcal{H}$ is an open subspace. Then since $0 \in E$, there exists a $r > 0$ such that

   $$B_r(0) \subseteq E$$

   However, for any nonzero element $x \in \mathcal{H}$,

   $$\frac{r}{2\|x\|}x \in B_r(0) \subseteq E$$

   Since E is closed under scalar multiplication, this shows that $x \in E$, so $\mathcal{H} \subseteq E$, so $\mathcal{H} = E$, showing that the only open subspace of $\mathcal{H}$ is $\mathcal{H}$ itself.

Thus, we are interested in closed subspace since are closed under limits, meaning we can use our tools from analysis to get the following decomposition result:

Theorem 4.5.5: Hilbert Space Decomposition

Let $E \subseteq \mathcal{H}$ be a closed subspace. Then:

$$\mathcal{H} = E \oplus E^\perp$$

In particular, for any $x \in H$, there is a unique $y \in E$ and $z \in E^\perp$ such that $x = y + z$.

***Proof* :**
Let $x \in \mathcal{H}$, and let $\delta = \inf\{\|x-y\| = \rho(x,y) \mid y \in E\}$ (which exists by completeness of $\mathbb{R}^n$ and $\mathcal{H}$). let $\{y_n\} \subseteq E$ be a sequence such that $\|x - y_n\| \to \delta$. We will show that $\{y_n\}$ is a cauchy sequence, and then find a limit point for it. By the parralelogram law:

$$2(\|x-y_n\| + \|x-y_m\|) = \|y_n - y_m\|^2 + \|(y_n + y_n) - 2x\|^2$$

Since $y_n + y_m \in E$, for sufficinetly large N so that $n, m > N$:

$$\begin{aligned}\|y_n - y_m\|^2 &= 2\|y_n - x\|^2 + 2\|y_m - x\|^2 - 4\|\frac{1}{2}(y_n + y_n) - x\|^2 \\ &\leq 2\|y_n - x\|^2 + 2\|y_m - x\|^2 - 4\delta^2\end{aligned}$$

and as $n, m \to \infty$, this sequence goes to $4\delta^2 - 4\delta^2 = 0$, and so $\|y_n - y_m\|\rho(y_n, y_m) \to 0$, meaning $\{y_i\}_i$ is a cauchy sequence. Take $y \lim y_n$. Then since E is closed, $y \in E$. By construction $\|x - y\| = \delta$. Now let $z = x - y$.

I claim that $z \in E^\perp$, that is $\langle z, u\rangle = 0$ for all $u \in E$. So, consider $\langle u, z\rangle$ for any $u \in E$. If $\langle u, z\rangle$ is complex, multiply it by a [complex] unit-length scalar to make it real ($\langle u, z\rangle = 0$ if and only if $a\langle u, z\rangle = 0$). With the fact that it's real, we will do a sneaky trick where we take advantage of optimization of differentiable functions. In particular, let

$$f : \mathbb{R} \to \mathbb{R} \quad f(t) = \|z + tu\|^2 = \|z\|^2 + 2t\langle z, u\rangle + \|u\|^2$$

f is certainly a differentiable function, being a polynomial with a square-root on $\mathbb{R}$. Furthermore, $z = x - y$ is the minimal value of $\|x - y\|$, and so the minimal value of $\|x - y\|^2$. It follows that $z + ut$ is not minimal, and so this equation is has a minimum when $t = 0$. Thus, we get $0 = f'(0) = 2\langle u, z\rangle$, implying that $\langle u, z\rangle = 0$. Since u was arbitrary, we get that $z \in E^\perp$.

To show that z is the minimal distance from x, let's say there was another $z' \in E^\perp$ that satisfied the equation. Then by the Pythagorean theorem:

$$\begin{aligned}\|x - z'\| &= \|(x - z) + (z - z')\|^2 \\ &= \|x - z\|^2 + \|z - z'\|^2 && x - y \in E, z - z' \in E^\perp \\ &\geq \|x - z\|^2\end{aligned}$$

So equality holds if and only if $z = z'$, meaning z is of minimal distance to x. A similar reasoning shows the same for y.

To show uniqueness, let's say $y + z = x = y' + z'$. Then $y - y' = z - z' \in E \cap E^\perp = \{0\}$, so $y - y' = z - z' = 0$, which can be used to show that $y = y'$ and $z = z'$, completing the proof.

We said that orthogonal vector are somehow dual to each other. However, there is the more classical notion of "dual" vector spaces. Let $\mathcal{H}^*$ be set of all linear functions from $\mathcal{H}$ to the underlying field (usually $\mathbb{C}$, but sometimes $\mathbb{R}$); such linear functions are called functionals. It turns out that all linear functionals can be written in the form $f_y(x) = \langle x, y\rangle$, and we can define a norm on $\mathbb{H}^*$ such that $\|f_y\| = \|y\|$. Thus, $y \mapsto f_y$ will define a conjugate-linear isometry from $\mathcal{H}$ to $\mathcal{H}^*$. The following theorem shows tha this map is surjective:

Theorem 4.5.6: Natural Isomorphism To $\mathcal{H}^*$ (Riesz Representation)

Let $f \in \mathcal{H}^*$. Then there exists a $y \in \mathcal{H}$ such that $f(x) = \langle x, y \rangle$ for all $x \in X$.

Proof **:**
It easy easy to check uniqueness: if $\langle x, y \rangle = \langle x, y' \rangle$ for all x, then let $x = y - y'$. Manipulating the equation, we get that $\|y - y'\| = 0$, so $y = y'$.

If f is the zero functional, then take $y = 0$. If otherwise, then let $\mathcal{M} = \{x \mid f(x) = 0\}$ be the kernel of f. It is clearly a subspace, and any squence will also necesarily converge within $\mathcal{M}$, and so it is also closed. Furthermore, $\mathcal{M}$ is a proper subspace, since $f \neq 0$. Thus, $\mathcal{M}^\perp \neq \{0\}$ by the previous proof.

Now, pick any $z \in \mathcal{M}^\perp$ with $\|z\| = 1$. Then if we set $u = f(x)z - f(z)x$, we have that $u \in \mathcal{M}$. Hence:

$$0 = \langle u, z \rangle = f(x)\|z\|^2 - f(z)\langle x, z \rangle = f(x) - \langle x, \overline{f(z)}z \rangle$$

re-arrangine and setting $y = \overline{f(z)}z$, we get

$$f(x) = \langle y, z \rangle$$

as we sought to show.

The uniqueness makes this map to be injective, and by linearity of f and the first component of the inner product makes it is an isomorphism. Furthermore, since $\|f_y\| = \|y\|$, limits are also preserved, and so this in fact defines an isomorphism between $\mathcal{H} \cong \mathcal{H}^*$. Notice how this differs from vector-spaces where there is no natural isomorphism between V and V^* (there is one between V and V^{**}).

Since we are working over a vector-space, we should be able to define a basis. Since we are studying infinite dimensional vector-spaces, we often care more about "countable linear combination" rather than "finite linear combination" as we have for vector-spaces. To distinguish the two, we will call *Hamel basis* a set of linearly independent elements for which all *finite linear combinations* can represent every element of the vector space. We will call a *Schauder basis* a set of linearly independent elements for which all *countable inear combinations* can represent every element of the vector space. If every element in a Schauder basis has norm 1 and is orthogonal, we call it an *orthonormal basis*. Since we have an inner-product that is complete, it would be nicest if this vector-space has an orthonormal basis. We strive to this end in the following build-up:

Theorem 4.5.7: Bessel's Inequality

If $\{u_\alpha\}_{\alpha \in A}$ is an orthonormal set in $\mathcal{H}$, then for any $x \in \mathcal{H}$:

$$\sum_{\alpha \in A} |\langle x, u_\alpha \rangle|^2 \leq \|x\|^2$$

As a consequence, $\{\alpha \mid \langle x, u_\alpha \rangle \neq 0\}$ is countable.

Proof :
It suffices to show that $\sum_{\alpha \in F} |\langle x, u_\alpha \rangle|^2 \leq \|x\|^2$ for any finite subset $F \subseteq A$, since if it fails for some countable amount, it will fail for some finite amount. Then we see that:

$$\begin{aligned}
0 &\leq \left\| x - \sum_{\alpha \in F} \langle x, u_\alpha \rangle u_\alpha \right\| \\
&= \|x\|^2 - 2\Re \left\langle x, \sum_{\alpha \in F} \langle x, u_\alpha \rangle u_\alpha \right\rangle + \left\| \sum_{\alpha \in F} \langle x, u_\alpha \rangle u_\alpha \right\| \\
&= \|x\|^2 - 2 \sum_{\alpha \in F} |\langle x, u_\alpha \rangle|^2 + \sum_{\alpha \in F} |\langle x, u_\alpha \rangle|^2 && \text{Pythagorean Theorem} \\
&= \|x\|^2 - \sum_{\alpha \in F} |\langle x, u_\alpha \rangle|^2
\end{aligned}$$

and manipulating the final equation we get:

$$\sum_{\alpha \in F} |\langle x, u_\alpha \rangle|^2 \leq \|x\|^2$$

as we sought to show

This in particular is important for uncountable sets. In the following we establish when equality holds, and thus when an orthonormal set is an orthonormal basis

Theorem 4.5.8: Orthonormal Basis Condition

Let $\{u_\alpha\}_{\alpha \in A}$ be an orthonormal set in $\mathcal{H}$. Then the following are equivlent:

1. (Completenss) If $\langle x, u_\alpha \rangle = 0$ for all α, then $x = 0$
2. (Parseval's Identity) $\|x\|^2 = \sum_{\alpha \in A} |\langle x, u_\alpha \rangle|^2$ for al l$x \in \mathcal{H}$
3. For each $x \in \mathcal{H}$, $x = \sum_{\alpha \in A} \langle x, u_\alpha \rangle u_\alpha$ where the right hand side sum has only countably many nonzero terms and converges in the norm topology no matter the order of the terms (so converges absolutely?)

Proof :
p.175 Folland

Definition 4.5.5: Orthonormal Basis

Let $\{u_\alpha\}_{\alpha \in A}$ be an orthonormal set. Then if the set satisfies any of the 3 conditions in theorem 4.5.8, it is called an *orthonormal basis.*

Theorem 4.5.9: Hilbert Space Basis

Every Hilbert space has a basis

***Proof* :**
Application of Zorn's lemma: The collection of ordered orthonormal sets ordered by inclusion has a maximal element, and maximality is equivalent to the first property of theorem 4.5.8.

Now that we have established a basis for Hilbert spaces, we will seperate them into 3 categories: Those with finite, countable, and uncountable basis. In most cases, we will be working with countable basis due to the following property:

Proposition 4.5.1: Seperable Hilbert Spaces

A Hilbert space $\mathcal{H}$ is seperable if and only if it has a countable orthonormal basis, in which case every orthonormal basis for $\mathcal{H}$ is countable.

Here, we are using the word seperable to mean countable dense subset.

***Proof* :**
p. 176

We next talk about the isomorphisms in the category of Hilbert spaces:

Definition 4.5.6: Unitary Map

Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be two hilbert spaces with inner products $\langle \cdot, \cdot \rangle_1$ and $\langle \cdot, \cdot \rangle_2$, and let $U : \mathcal{H}_1 \to \mathcal{H}_2$. Then U is called a *unitary map* if if it is an invertible linear map that perserves inner products:

$$\langle Ux, Uy \rangle_2 = \langle x, y \rangle_1$$

Note that by taking $x = y$, we get that $\|x\|_1 = \|Ux\|_2$, i.e. it is an isometry. In fact, the converse is true too: every surjective isometry is unitary. With the definition of an isomorphism, we can see that all Hilbert spaces are isomorphic to $\ell^2(A)$ given basis A:

Theorem 4.5.10: Classifying Hilbert Spaces

Let $\{e_\alpha\}_{\alpha \in A}$ be an orthonormal basis for $\mathcal{H}$. Then the correspondence

$$x \mapsto \hat{x} \qquad \hat{x}(\alpha) = \langle x, u_\alpha \rangle$$

is a unitary map from $\mathcal{H}$ to $\ell^2(A)$.

***Proof* :**
We'll show it's a surjective isometry, which is equivalent to it being unitary.

First off, the map $x \to \hat{x}$ is linear since the inner product is linear in each component, and it is an isometry by Parseval identity: $\|x\|^2 = \sum |\hat{x}(x)|^2$.

For surjectivity, let $f \in \ell^2(A)$, so $\sum_{\alpha \in A} |f(\alpha)^2| < \infty$, so the Pythagorean shows that all partial sumns of the series $\sum f(\alpha)u_\alpha$ are Cauchy [a], so $x = \sum_\alpha f(\alpha)u_\alpha$ exists in $\mathcal{H}$ and $\hat{x} = f$.

[a]Note that only countably many terms are nonzero

Chapter 5

L^p space

Look at this blog!! `https://terrytao.wordpress.com/2009/01/09/245b-notes-3-lp-spaces/` In this chapter, we will be studying Banach spaces which are defined through the integral: L^p spaces. These spaces are a generalization of L^1 spaces, where if $f \in L^p$, then it must decay "faster" at infinity, and and f doesn't blow up too fast at any finite point. In other words, L^p spaces capture a type of "speed of growth". As we have seen in the previous chapters, limits of integrals are central to many constructions. For some constructions, being in L^1 was enough, but for others (like Fourier series) we will require that functsion be more than finite in integral (i.e. more than $f \in L^1$). Furthermore, they L^p spaces will give a nice set of examples of Banach spaces that are relatively simple to define. Finally, they are also sometimes treated as the intermediate to being able to bound a function: it is sometimes nontrivial to immediately say a function is bounded, but we can build-up to it by saying it's L^3, which will imply L^5, which will imply L^7, and so on allowing us to conclude it's L^∞, which means a functins is bounded up to a zero-measure set.

5.1 Basics

So far, we have limited our attention to measurable functions f such that $\int |f| < \infty$ and have labelled them L^1. Often, what we are looking are function with more precise "convergent properties", in particular:

$$\int |f|^p < \infty$$

for some $p \geq 1$, including ∞ (which we will define shortly). We start by labelling the colletion of all these functions:

Definition 5.1.1: L^p Space

Let f be measurable. for $p \in (1, \infty)^a$, define the *p-norm* to be:

$$\|f\|_p = \left[\int |f|^p \, d\mu\right]^{1/p}$$

From this norm, define:

$$L^p(X, \mathcal{M}, \mu) := \left\{ f : X \to \mathbb{C} \;\middle|\; f \text{ is measurable and } \|f\|_p < \infty \right\}$$

[a] we will define L^∞ later

As with L^1, we often abbreviate to $L^p(\mu)$, $L^p(X)$, or just L^p. As with L^1, we'll let L^p be the equivalence class of the set of functions such that $|f|^p$ is integrable where $f \sim g$ if and only if $\int |f - g|^p = 0$. Just like L^1, two functions will be equal in L^p (i.e. belong in the same equivalence class) if they are equal almost everywhere. Furthermore, we claimed that $\|\cdot\|_p$ is a norm. That $\|f\|_p = 0$ if and only if $f = 0$ a.e. comes almost immediately, since

$$\left(\int |f|^p\right)^{1/p} = 0 \quad \text{iff} \quad \int |f|^p = 0 \quad \text{iff} \quad \int |f| = 0 \quad \text{iff} \quad f = 0 \text{ a.e.}$$

For $\|cf\|_p = |c| \, \|f\|_p$, we take advantage of our normalisation:

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

However, the triangle inequality is not so trivial. We will in fact take good part of this section to show that the triangle inequality is satisfied. For now, we take a moment to explore an example of L^p spaces and L^p functions to gain an intuition:

Example 5.1: L^p spaces

1. Let $f : [0, \infty) \to \mathbb{R}$ $f(x) = \frac{1}{x^{1/p}}$. Then $f \in L^q$ for $q > p$. This comes down to the fact that:

$$x > x^{1/p} \;\Leftrightarrow\; \frac{1}{x} < \frac{1}{x^{1/p}}$$

   and by the fact that $\int \left|\frac{1}{x^n}\right| < \infty$ if $n > 1$. We have therefore found a class of functions that fits into all possible $p \in [1, \infty)$. In fact, we can find functions that fit into any connected subinterval of $[1, \infty)$! In particular, let $(X, \mathcal{M}, \mu)$ be a measure space and $f : X \to \mathbb{C}$ a complex measurable function on X. Let

$$E := \left\{ p \in [1, \infty) | \;\middle|\; \int_X |f|^p < \infty \right\}$$

   (a) Then E is connected, that is, if $p, q \in E$ such that $p < r < q$ for some r, then $r \in E$

   Proof. Clearly, if $E = \emptyset$ or $|E| = 1$, then E is connected because the only non-trivial clopen set is $\emptyset$ and E, so let $|E| > 1$.

Pick $p, q \in E$ such that $p < q$, and consider a point r such that $p < r < q$. We'll show that $\int_X |f|^r < \infty$ using the following trick:

$$|f| = \chi_{|f|\geq 1}|f| + \chi_{|f|<1}|f|$$

that is, we're splitting $|f|$ to where f is greater than or equal to 1 and strictly less than 1. The equality holds since $A = \{x \mid |f(x)| \geq 1\}$ and $B = \{x \mid |f(x)| < 1\}$ are disjoint. Importantly, notice that:

$$|f|^p = \chi_{|f|\geq 1}|f|^p + \chi_{|f|<1}|f|^p \qquad |f|^q = \chi_{|f|\geq 1}|f|^q + \chi_{|f|<1}|f|^q$$

Thus, by the linearity of the integral, we get that:

$$\int |f|^p = \int \chi_{|f|\geq 1}|f|^p + \int \chi_{|f|<1}|f|^p \qquad \int |f|^q = \int \chi_{|f|\geq 1}|f|^q + \int \chi_{|f|<1}|f|^q$$

It thus suffice show that $\int \chi_{|f|\geq 1}|f|^r < \infty$ and $\int \chi_{|f|<1}|f|^r < \infty$. This is essentially immediate:

$$\chi_{|f|\geq 1}|f|^r \leq \chi_{|f|\geq 1}|f|^q < \infty \qquad \chi_{|f|<1}|f|^r \leq \chi_{|f|<1}|f|^p < \infty$$

Therefore

$$\int |f|^r < \infty$$

showing that $r \in E$, as we sought to show. □

(b) Show that we can pick a measure and an f such that $E = (a, b), [a, b), [a, b]$, or $\{a\}$ for $1 \leq a \leq b \leq \infty$. In other words, all connected subsets of $[1, \infty)$ are possible for E (You may freely use the fact that any Riemann integrable function is also integrable with respect to the Lebesgue measure and its Riemann integral is equal to the Lebesgue integral)
Hint: Think about integrability properties of functions like x^{-a} and $\frac{-1}{\log^2(x)x} = \frac{d}{dx}\frac{1}{\log(x)}$

Proof. Throughout this hole proof we'll be using Riemann integration as a substitute for Lebesgue integration (we showed earlier that this is allowed). We'll be using these function's to construct our intervals:

i. $f(x) = x^{-1/a}$ on $(1, \infty)$: Notice that

$$\int_1^\infty x^{-1/a}dx = \lim_{x\to\infty} \frac{x^{-1/a+1}}{-1/a+1} - \frac{1}{-1/a+1}$$

so $\|f\|_p < \infty$ if and only if $p \in (a, \infty]$.

ii. $f(x) = x^{-1/a}$ on $(0, 1)$: Notice that

$$\int_0^1 x^{-1/a}dx = \frac{1}{-1/a+1} - \lim_{x\to 0} \frac{x^{-1/a+1}}{-1/a+1}$$

so $\|f\|_p < \infty$ if and only if $p \in [1, a)$.

iii. $f(x) = \frac{1}{(\ln^2(x)x)^{1/a}}$ on $(1, \infty)$: This integral doesn't have an elementary solution, however we can still ascertain some basic convergence properties. In particular, if $p = a$, then:

$$\int_1^\infty |f(x)|^a dx = \lim_{x\to\infty}\left[-\frac{1}{\ln(x)}\right] - \lim_{x\to 1}\left[-\frac{1}{\ln(x)}\right]$$

so $\|f\|_p < \infty$ if and only if $p \in [a, \infty]$

iv. $f(x) = \frac{1}{(\ln^2(x)x)^{1/a}}$ on $(0, 1)$: This result is similar to the other 3 cases we've proven. The result for this integral shows that $\|f\|_p < \infty$ if and only if $p \in [1, a]$.

Thus, we have proven all possible "sub-rays" of $[1, \infty]$ can exist. To show all possible sub-intervals (closed, open, or half), we simply add the appropriate function's to produce the result. For example, if we want to have the interval $[a, b)$, simply take:

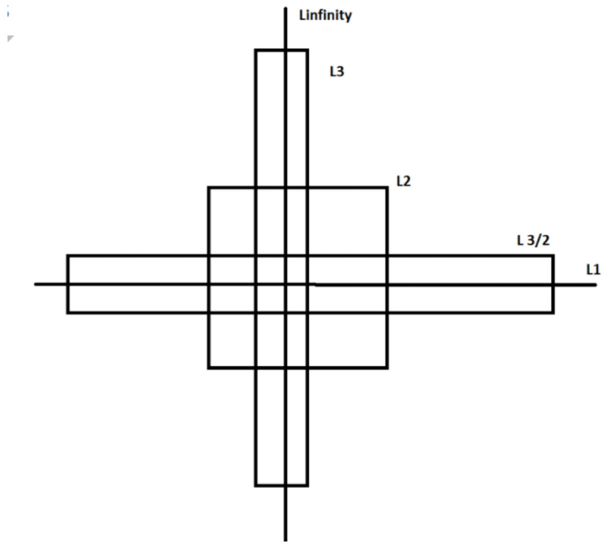
$$f(x) = \frac{1}{(\ln^2(x)x)^{1/a}}\chi_{(1,\infty)} + x^{-1/b}\chi_{(0,1)}$$

which clearly has interval of convergence of $[a, b)$. Similarly, to get a singleton $\{a\}$, take:

$$f(x) = \frac{1}{(\ln^2(x)x)^{1/a}}\chi_{(1,\infty)} + \frac{1}{(\ln^2(x)x)^{1/a}}\chi_{(0,1)}$$

Thus, we have all possible connected subintervals of $[1, \infty]$. $\square$

So if $f \in L^p$, it does not imply that $f \in L^q$ for $p > q$ or $p < q$! Later on, we will show some of the possible relations between L^p spaces. A more accurate intuition is the following [a]: If $f \in L^p$ for any p, then f must decays fast enough at infinity, and f must not blow up too fast at any finite point. As p increases, the first requirement gets less strict, and the second gets more strict. If X is bounded, the first requirement is vacuous, and if X is discrete, the second is vacuous. Sometimes this visual is used to represent these relations:



2. A particular case that will be common for analyzing sequences is when μ is the counting measure on A. In that case, we usually denote $L^p(\mu)$ by $\ell^p(A)$. If $A = \mathbb{N}$, we often simply

abbreviate notation to ℓ^p. If we write $f : \mathbb{N} \to \mathbb{R},\ f(k) = a_k$, then the sequence f (i.e. $\{a_k\}$ is in ℓ^p if:

$$\sum_{k=1}^{\infty} |a_k|^p < \infty$$

Thus, ℓ^1 is the set of convergence series.

[a] credit to this post on stack-exchange: https://mathoverflow.net/questions/314050/intuition-about-lp-spaces

Notice that we are "normalizing" each integral by a factor of $1/p$. This normalization is done to make sure that the constant can be brought out of the norm. It might seem tempted to say that:

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

However, this is most certainly not true (simply take $f(x) = x$). As mentioned before, this makes it non-trivial to show that $\|\cdot\|_p$ is a norm on L^p. However, it is easy to see that L^p is a vector-space. If $f, g \in L^p$, then:

$$|f+g|^p \leq [2\max(|f|,|g|)]^p \leq 2^p(|f|^p + |g|^p)$$

which is finite by assumption, and so $f + g \in L^p$. It is also immediate that $cf \in L^p$. Thus, we may freely add and multiply by L^p functions without worrying about leaving L^p. Because of this, it makes it meaningful to ask whether

$$\|f_1 + f_2\|_p \leq \|f_1\|_p + \|f_2\|_p$$

What we will now do is show that $\|\cdot\|_p$ is does indeed satisfy the triangle inequality for $p \geq 1$. If $p < 1$, then the triangle inequality fails in general:

Example 5.2: $p < 1$ then no Triangle Inequality

Take $a, b > 0$ and $0 < p < 1$. Then for any $t \in \mathbb{R}$, we have that:

$$t^{p-1} > (a+t)^{p-1}$$

Integrating both sides from 0 to b, we get:

$$\begin{aligned}
\int_0^b t^{p-1} &> \int_0^b (a+t)^{p-1} \\
\frac{b^p}{p} &> \frac{(a+b)^p}{p} \\
b^p &> (a+b)^p \\
a^p + b^p &> (a+b)^p \\
(a^p + b^p)^{1/p} &> a + b
\end{aligned}$$

Now, if E and F are disjoint union with positive finite measures in X such that $\mu(E)^{1/p} = a$ and $\mu(E)^{1/p} = b$, then $\chi_E, \chi_F \in L^p$. Taking advantage of the fact that $\chi_E\chi_F = 0$ (i.e. the zero function) since $E \cap F = \emptyset$, then:

$$\|\chi_E + \chi_F\|_p = (a^p + b^p)^{1/p} > a + b = \|\chi_E\|_p + \|\chi_F\|_p$$

Thus, the $\|\cdot\|_p$ for $0 < p < 1$ cannot be a norm.

Thus, we stick to showing the case where $p \geq 1$ First, we will state a simple inequality fact similar Jensen's inequality. This is in fact the key idea behind the entire proof:

Lemma 5.1.1: Young's Inequality

If $a, b \geq 0$ and $0 < \lambda < 1$, then:

$$a^{\lambda} b^{1-\lambda} \leq \lambda a + (1-\lambda) b$$

with equality if and only if $a = b$. In particular, notice that if $\frac{1}{p} + \frac{1}{q} = 1$, then if $\lambda = p^{-1}$, then $(1-\lambda) = q^{-1}$ and so:

$$a^{1/p} b^{1/q} \leq \frac{1}{p} a + \frac{1}{q} b$$

Proof **:**
This proof is trying to capture what this image is doing in formal language:

`https://www.desmos.com/calculator/jdznjzqcju`

If $a = b$, equality clearly holds. If $a = 0$ or $b = 0$, inequality clearly holds. If $a, b \neq 0$, then dividing by b and letting $t = a/b$, we get

$$a^{\lambda} b^{-\lambda} \leq \lambda(a/b) + (1-\lambda) \quad \Leftrightarrow \quad t^{\lambda} \leq \lambda t + (1-\lambda) \quad \Leftrightarrow \quad t^{\lambda} - \lambda t \leq (1-\lambda) \tag{5.1}$$

From here, we use some calculus: we see that $t^{\lambda} - \lambda t$ is strictly increasing for $0 < t < 1$ (since $a, b \geq 0$, $t \geq 0$) and strictly decreasing for $t > 1$, so the maximal value occurs at $t = 1$, and in fact we get that the maximal value is $1 - \lambda$:

$$1^{\lambda} - \lambda 1 = 1 - \lambda \leq 1 - \lambda$$

and so the inequalities in equation (5.1) hold! Equality happens when $t = a/b = 1$ which implies $a = b$, completing the proof.

Usually, Young's inequality is stated as

$$a^{\alpha} b^{\beta} \leq \alpha a + \beta b \qquad 0 \leq \alpha, \beta \leq 1, \quad \alpha + \beta = 1$$

which is exactly what we have, since if we fix some α, then it must be that $\beta = 1 - \alpha$. Another formulation of Young's inequality is:

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}$$

with $p^{-1} + q^{-1} = 1$. This proof is a "Jensen equality" like proof that more clearly shows the role of concavity and so I'll dedicate a moment to prove this one too:

Proof **:**
If $a = 0$ or $b = 0$, the result is immediate, so assume $a, b > 0$. Then notice that since log is monotone:

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q} \quad \Leftrightarrow \quad \log(ab) \leq \log\left(\frac{a^p}{p} + \frac{b^q}{q}\right)$$

which is great, since we know log is concave and so Jensen's inequality applies. Since $p^{-1}+q^{-1}=1$, we get $t=1/p$ and $(1-t)=1/q$, and so by Jensen's inequality:

$$\log(ta^p+(1-t)b^q)\geq t\log(a^p)+(1-t)\log(b^q)=\frac{p}{p}\log(a)+\frac{q}{q}\log(b)=\log(ab)$$

completing the proof.

The second interpretation of $p^{-1}+q^{-1}=1$ in Young's inequality is a nice way to write the straight line (or sub-straight line) equation without the $1-\lambda$ term, but just two numbers. It furthermore let's us relate two numbers that can be much larger. For example, if I choose $p=100$, then it must be that $q=100/99$ so that

$$p^{-1}+q^{-1}=1/100+99/100=1$$

which, if we consider p and q to be the exponents associated to L^p and L^q, we see that this gives us a way to relate different L^p space!

To see this more succinctly, we use this convexity property given by Young's inequality in the "build-up" lemma towards the triangle inequality. This lemma is important later on in functional analysis since it let's us bound the operation of multiplication of tow functions fg, (namely, $\|fg\|_1$ will be less than a value on the right hand side that seperates f and g). For our current purposess, it is important for the triangle inequality:

Theorem 5.1.1: Hölder's Inequality

Let $1\leq p,q\leq\infty$ satisfy:

$$\frac{1}{p}+\frac{1}{q}=1$$

or equivalently $q=p/(p-1)$, where we will interpret $\frac{1}{\infty}$ as 0. Then if f and g are measurable functions on X then:

$$\|fg\|_1\leq\|f\|_p\,\|g\|_q$$

In particular, if $f\in L^p$ and $g\in L^q$ then $fg\in L^1$. Equality holds if and only if $\alpha|f|^p=\beta|g|^q$ a.e. for some constants α,β with $\alpha\beta\neq 0$

Notice that when $p=q=2$ (if $p=2$, it must be that $q=2$ to satisfy the equation), then this inequality is called the *Cauchy-Schwartz inequality*[1]. Notice too that this is the only time when $p=q$; this will become important when we analyze the $p=2$ case later.

I also found this intuition on a stack-exchange: the slower f decays, the faster g needs to decay, and the faster f blows up, the more we need g to not blow up. The previous sentence is equally true if we switch f and g, which convinces us that L^p spaces should be reflexive, and by considering $f=g$, we see that the restrictions on f and g should balance at exactly $p=2$, so L^2 should be self-dual.

Notice too that since we are taking the absolute value of the function on the inside, the integral can be thought of as a sort of "concave function", where the input would be, say in the special case of $\mathbb{R}$ as the domain, some interval. This makes it easier to see why this inequality holds (notice that convex functions must preserve monotonicity, and so does the integral).

[1]Technically we need to define an inner product on L^2 to satisfy this inequality. However, as well see in section 4.5, L^2 is indeed a Hilbert space

Proof **:**
If $\|f\|_p = 0$ or $\|g\|_p = 0$, then $f = 0$ or $g = 0$ a.e., and so equality holds. Similarly if $\|f\|_p = \infty$ or $\|g\|_p = \infty$, inequality immediately holds. Furthermore, notice that if we find a particular f and g for which the result holds, then the result holds for af and bg since

$$\begin{aligned} &\Leftrightarrow \|afbg\|_1 \le |ab| \|af\|_p \|bg\|_q \\ &\Leftrightarrow |ab| \|fg\|_1 \le |ab| \|f\|_p \|g\|_q \\ &\Leftrightarrow \|fg\|_1 \le \|f\|_p \|g\|_q \end{aligned}$$

So, it suffices to show the result when $\|f\|_p = \|g\|_q = 1$, and equality holds of $|f|^p = |g|^q = 1$. To this end, we use the previous lemma and set $a = |f(x)|^p$ and $b = |g(x)|^q$ and $\lambda = p^{-1}$ (so that $1 - \lambda = 1 - p^{-1} = q^{-1}$). Then by the previous lemma we have:

$$|f(x)|^{p \cdot p^{-1}} |g(x)|^{q \cdot q^{-1}} \le p^{-1} |f(x)|^p + q^{-1} |g(x)|^q \quad \Leftrightarrow \quad |f(x)g(x)| \le p^{-1}|f(x)|^p + q^{-1}|g(x)|^q \tag{5.2}$$

Now, integrating both sides, we get:

$$\|fg\|_1 \le p^{-1} \int |f|^p + q^{-1} \int |f|^q = p^{-1} + q^{-1} = 1 = \|f\|_p \|g\|_q$$

where equality holds when equality holds in equation (5.2) a.e., which holds only when $|f|^p = |g|^q$ a.e., as we sought to show.

For a visual conformation, you can check out:

`https://www.desmos.com/calculator/gwqvfv4tie`

When p and q satisfy the equality in the equation, we call them *conjugate exponents*. As mentioned before, the conjugate exponents is simply another way or writing the equation of a straight line to apply Young's inequality. From this, we can take advantage of this convexity property yet again to prove the triangle inequality of $\|\cdot\|_p$, which has a different name since for the special case of the p-norm due to the fact it is used on other places and needs to be referenced:

Theorem 5.1.2: Minkowski's Inequality

If $1 \le p < \infty$ and $f, g \in L^p$. Then:

$$\|f + g\|_p \le \|f\|_p + \|g\|_p$$

Before proceeding with the proof, a quick word on notation to avoid confusion:

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

We often work with $\|f\|_p^p = \int |f|^p$ to get rid of the $1/p$ exponent to be able to work with the linearity of the integral, and re-introduce it at the end to get the desried result. We did not have to do this for Hölder's inequality since we reduced to a very nice special case.

Proof :
If $p = 1$, then the result immediately follows by the regular triangle inequality and linearity of the integral. If $f + g = 0$ a.e., then the result also follows, so assume that $f + g \neq 0$ a.e., which using the triangle inequality:

$$|f + g|^p \leq (|f| + |g|)|f + g|^{p-1} = |f||f + g|^{p-1} + |g||f + g|^{p-1}$$

So, we will apply Hölder's inequality using the fact that $q = \frac{p}{p-1}$ and integrating:

$$\begin{aligned}
\|f + g\|_p^p &= \int |f + g|^p d\mu \\
&= \int |f + g| \cdot |f + g|^{p-1} d\mu \\
&\leq \int (|f| + |g|)|f + g|^{p-1} d\mu \\
&= \int |f||f + g|^{p-1} d\mu + \int |g||f + g|^{p-1} d\mu \\
&= \left\| |f|(|f + g|^{p-1}) \right\|_1 + \left\| |g|(|f + g|^{p-1}) \right\|_1 \\
&\leq \||f|\|_p \|f + g\|_q^{p-1} + \|g\|_p \|f + g\|_q^{p-1} && \text{Hölder's Inequality} \\
&= \left(\left(\int |f|^p d\mu \right)^{\frac{1}{p}} + \left(\int |g|^p d\mu \right)^{\frac{1}{p}} \right) \left(\int |f + g|^{(p-1)\left(\frac{p}{p-1}\right)} d\mu \right)^{\frac{p-1}{p}} && \text{recall } q = \frac{p}{p-1} \\
&= \left(\left(\int |f|^p d\mu \right)^{\frac{1}{p}} + \left(\int |g|^p d\mu \right)^{\frac{1}{p}} \right) \left(\int |f + g|^p d\mu \right)^{\frac{p-1}{p}} \\
&= (\|f\|_p + \|g\|_p) \frac{\|f + g\|_p^p}{\|f + g\|_p}
\end{aligned}$$

Thus, multiplying both sides by $\frac{\|f+g\|_p}{\|f+g\|_p^p}$, we get

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p$$

As we sought to show

Thus, we get that $\|\cdot\|_p$ is indeed a norm! Thus L^p is a normed vector space. In fact, just like L^1, it is a complete normed vector space, i.e., a Banach space. We first recall a lemma to simply our task. To understand the lemma, notice that every norm defines a metric:

$$\rho(x, y) = \|x - y\|$$

and so defines a topology, in particular a normal space, that has the notion of sequences well-defined. Next, recall that if $\{x_n\}$ is a sequence in V, it's series is said to converge if there exists an $x \in V$ such that $\sum_{n=1}^{\infty} x_n = x$, and is said to be absolutely convergence if $\sum_{n=1}^{\infty} \|x_n\| < \infty$ (this is lemma 4.1.1). If you did not cover this result yet, it is important to go see it now to understand the following result:

Theorem 5.1.3: L^p is a Banach Space

For $1 \leq p < \infty$, L^p is a complete normed vector-space, i.e., a Banach space. Moreover, if $f_n \to f$ in L^p, then there exists a subsequence that converges pointwise to f a.e.

***Proof* :**
Let $1 \leq p < \infty$. By the previous lemma, it suffices to show that every absolutely convergent series converges:

$$\sum_{k=1}^{\infty} \|f\|_p = B < \infty \quad \Rightarrow \quad \exists f \text{ s.t. } \sum_{k=1}^{\infty} f_k = f \in L^p$$

Let $G_n = \sum_{k=1}^{n} |f_k|$ be the partial sums of the series and $G = \sum_{k=1}^{\infty} |f_k|$ be the series. Then by Minkowski's inequality, we get:

$$\|G_n\|_p \leq \sum_{k=1}^{n} \|f_k\|_p \leq B \quad \forall n \in \mathbb{N}$$

and so $\|G\|_p = \left(\int |G|^p\right)^{1/p} < \infty$ and so $\left(\int |G|^p\right) < \infty$. Since $G_n \leq G_{n+1}$, and $G_n : X \to [0, \infty)$, by the Monotonce Convergence Theorem:

$$\int \lim_{n\to\infty} G_n^p = \lim_{n\to\infty} \int G_n^p \leq B^p < \infty$$

and so $G \in L^p$. By proposition 2.3.2, $G(x) < \infty$ a.e., so the series $\sum_{k=1}^{\infty} f_k$ converges a.e. If we let $F = \sum_{k=1}^{\infty} f_k$, we have that $|F| \leq G$ a.e., and so $F \in L^p$. From this, we can see that:

$$\left\| F - \sum_{k=1}^{\infty} f_k \right\|_p^p = \int \left| F - \sum_{k=1}^{\infty} f_k \right|^p \to 0$$

Thus, the series $\sum_{k=1}^{\infty} f$ converges (to F) in the p-norm, and so by lemma 4.1.1, L^p is complete, as we sought to show.

This is the proof the prof was doing in class, but I got stuck so found a different proof (the above proof) (Commented it out for now)

This also finally showed that L^1 is a complete space under $\|\cdot\|_1$! Note the importance of taking a subsequence: since $f_n \to f$ in the p-norm, i.e. the "sum-averaging" of the function converges, then there can be lots of wiggle room for hwo the functions f_n approach f:

Example 5.3: Taking A Subsequence
Construct a sequence of functions such that all the following hold:

1. $f_n : [0, 1] \to \mathbb{R}$ is continuous
2. $0 \leq f_n(x) \leq 1$ for every x
3. $\lim_{n\to\infty} \int_0^1 f_n dm = 0$

4. For every $x \in [0,1]$ we have that $\lim_{n\to\infty} f_n(x)$ does not exist

(that is, it is definitely important to take a subsequence in order to deduce pointwise convergence from converge in L^1)

Proof. Essentially, we'll be taking advantage that $\sum_n \frac{1}{n}$ does not converge. Let $f_1(x) = 1$, that is, that constant function with constant 1.

For g_2, we split the interval into 2 equal parts. Define

$$g_2(x) = \begin{cases} 1 & x \in [0,1/2] \\ 0 & x \in (1/2,1] \end{cases}$$

This function is not continuous. We will make it continuous by simply climbing to it (and for other values, we'll be descending back down too). At the point $1/2$, slope down at a speed of 2^2 until we reach 0:

$$f_2(x) = \begin{cases} 1 & x \in [0,1/2] \\ -2^2\left(x - \left(1 - \frac{1}{4}\right)\right) & x \in [1/2,3/4] \\ 0 & x \in (3/4,1] \end{cases}$$

From here, we will define our function in a sort of "recursive" sense. Starting form the inteval $[0,1/2)$, we keep the right end point, and add an interval of length $1/3$ to it: $[1/2,5/6)$. Now, define f_3 to be 1 on $[1/2,5/6]$, and zero everywhere else. Next, to define f_4, repeat the process by adding $1/4$ to $5/6$, giving us the interveal $[5/6,13/12]$. Note that $13/12$ goes beyond 1. We thus wrap it around back to 0: $[0,1/12] \cup [5/6,1]$. Now, define a function f_3 to be 1 on this set, and zero everywhere else.

In each of these functions, we extend them to be continuous by attaching them a slope which has 2^n as a factor. In this way, the slope shrinks in size quicker than the size of the rectangles, and so we can morally imagine the problem to be one of indicator functions.

We can continue doing this, defining every f_k by appending $1/k$ and wrapping around when we go past 1 and then making the function continuous. Clearly, $\int f_k \to 0$ as $k \to \infty$ since each inteval will be of total length $1/k + 1/2^k$. However, f_k does not converge anywhere: for any $x \in [0,1]$, we can choose a subsequence from $f_k(x)$ (say $f_{k_l}(x)$) such that $f_{k_l}(x) = 1$. or $f_{k_l}(x) = 0$

To construct such a sequence, it is easier to imagine our construction without the wrapping around, that is, we keep adding intervals of length $1/k$ as though our co-domain was $[0,\infty)$. Then, our first value of l will be the value for which $x+1$ is in some interval of length $1/k$ for appropriate k. This interval will exist, since $\sum_k 1/k$ diverges. Similarly, our next value of l for which $x+2$ will be in some interval of length $1/k$ for appropraite k. We can continue doing this indefinitely, showing that there exists a subsequence of f_k that doesn't converge to 0, meaning f_k cannot converge to 0. On the other hand, it is easy to see that we can choose a subsequence that converges to 0 (since f_k is zero after the trapezoid passes by our point).

Since x was arbitrary, this will work any $x \in [0,1]$. Thus, f does not converge anywhere, as we sought to show. □

Retruning to analyzing L^p spaces, we can now extend theorem 2.3.3 to show that simple function are dense in every L^p!

Theorem 5.1.4: Simple Functions Dense in L^p

For $1 \leq p < \infty$, the set of simple functions $f = \sum_{i=1}^n a_i \chi_{E_i}$ where $\mu(E_i) < \infty$ for all i is dense in L^p

Proof **:**
First of all, the set of simple functions is in L^p for every $1 \leq p < \infty$. Next, choosing any $f \in L^p$, choose some standard representation $\{f_n\}$ where $f_n \to f$ a.e[a]. Then since $|f_n| \leq |f|$, we get that $|f_n| \in L^p$. Furthermore

$$|f - f_n|^p \leq |2f|^p = 2^p |f|^p \in L^1$$

and so by the dominated convergence theorem, we get:

$$\int |f - f_n|^p \to 0 \quad \Rightarrow \quad \left(\int |f - f_n|^p \right)^{1/p} = \|f - f_n\|_p \to 0$$

Moreover, if $f_n = \sum_{k=1}^\infty a_k \chi_{E_k}$ was a simple function where all E_i are pairwise disjoint and a_i are nonzero, it must be that $\mu(E_i) < \infty$ since

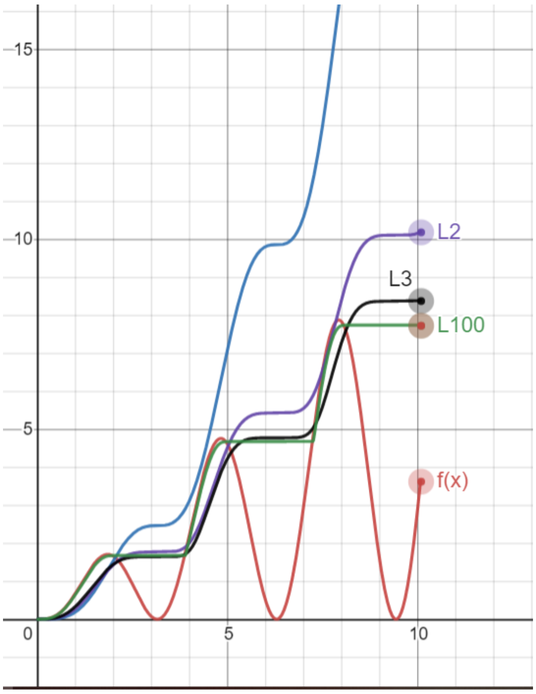
$$\sum_{k=1}^\infty |a_k|^p \mu(E_k) = \int |f_n|^p < \infty$$

completing the proof

[a] Recall that the any measurable function has a standard representation, not just integrable functions

Some further remarks about which sets are dense can be enlightening. For $1 \leq p, q < \infty$, L^p is dense in L^q, since each simple set of the form we descriped are subsets of L^p for all $p \geq 1$. If we modify the characteristic function to make them continuous (simply draw a connecting line as we've done in example 5.3), then we see that $C_0(\mathbb{R})$ (continuous functions that vanish at infinity) are dense in L^p for every $1 \leq p < \infty$. If we restrict to closed intervals, then $C([a, b])$ is dense in $L^p([a, b])$ since the standard representation becomes uniformaly convergent on bounded domains (see theorem 2.1.1).

So far, we have defined L^p spaces where $0 < p < \infty$ and showed that for $1 \leq p < \infty$, L^p is a Banach space. Interestingly, if we take a look at a graph of $\|f\|_p$ as $p \to \infty$, we'll find that there is a striking pattern that appears:



Notice that as $p \to \infty$, the resulting $\|f\|_p$ value got closer to a max function. In fact, this is almost the case! We might be tempted to then say that $\|f\|_\infty = \sup f$. However, the problem with this is that we might have a measure 0 set that diverges, while the supremum of the rest of the function is attained by a value $k < \infty$. The idea of $\sup f$ on all but a zero measure set will be called the *essential supremum*, and all functions that have a finite essential supremum will be denoted as L^∞. In this way, L^∞ is simply a natural extension of our definition of L^p:

Definition 5.1.2: L^∞ Space

Let $(X, \mathcal{M}, \mu)$ be a measure space and let $f : X \to [0, \infty]$ be a measurable function. Let

$$S = \left\{ r \in \mathbb{R} \;\middle|\; \mu(f^{-1}((r, \infty])) = 0 \right\}$$

Then the *essential supremum* of f, denoted $\|f\|_\infty$ or $\operatorname{ess\,sup} f$ is :

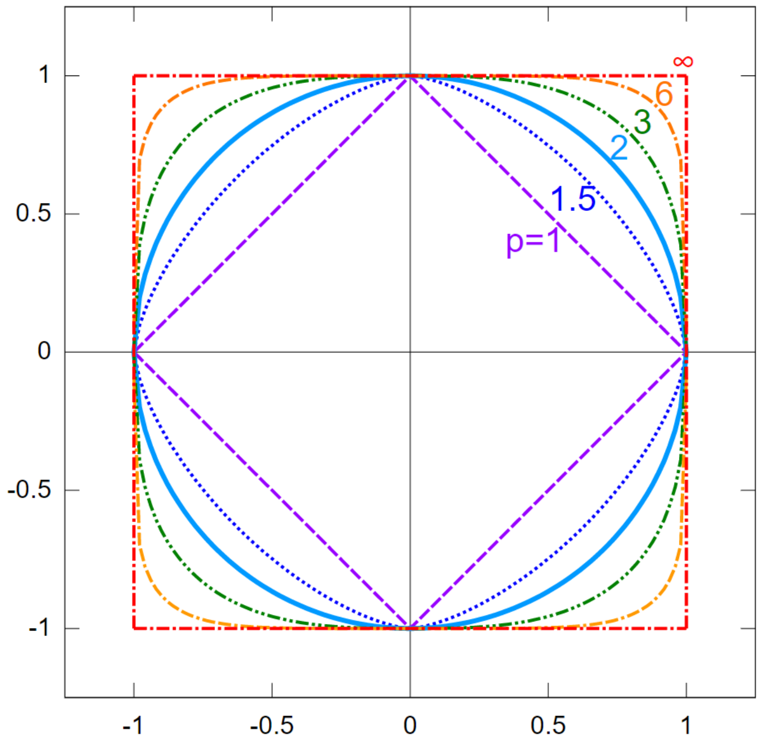
$$\|f\|_\infty := \begin{cases} \infty & \text{if } S = \emptyset \\ \inf S & \text{if } S \neq \emptyset \end{cases}$$

or, equivalently:

$$\|f\|_\infty = \inf_{\substack{E \in \mathcal{M} \\ \mu(X \setminus E) = 0}} \left(\sup_{x \in E} |f(x)| \right)$$

If f is a complex measurable, function, then $\|f\|_\infty = \|\,|f|\,\|_\infty$

Essentially, $\|f\|_\infty$ is the least upper bound of f, ignoring measure zero sets worth of points. Using this definition, we can define L^∞ to be the equivalence classes of functions f such that $\|f\|_\infty < \infty$. Another quick way of visualizing the relation between all p-norms, $1 \le p \le \infty$ is by taking the ℓ^p norm on $\mathbb{R}^2$, taking the function $x \mapsto \|x\|_p = \sqrt[p]{x^p + y^p}$, and considering when $\|x\| = 1$, or in terms of set notation $S_p = \left\{ x \in \mathbb{R}^2 \;\middle|\; \|x\|_p = 1 \right\}$. We'd get:



A couple more quick remarks: $L^\infty(\mu)$ depends on μ only insofar as μ determines which sets have measure zero. If μ and ν are mutually absolutely continuous ($\mu \ll \nu$ and $\nu \ll \mu^2$) then $L^\infty(\mu) = L^\infty(\nu)$. Secondly, if μ is not semi-finite, it is useful to adpot a slightly useful definition of L^∞ (Folland left this as exercsie 23-25 in chatper 6).

Most properties of L^p for $1 \le p < \infty$ generalises naturally to the L^∞ case:

[2] see definition 3.2.1

Proposition 5.1.1: Properties of L^∞

Let $(X, \mathcal{M}, \mu)$ be a measure and L^∞ the set of measurable function satisfying $\|\cdot\|_\infty < \infty$. Then:

1. If f, g are measurable functions on X, then
$$\|fg\|_1 \leq \|f\|_1 \|g\|_\infty$$
If $f \in L^1$ and $g \in L^\infty$, then $\|fg\|_1 = \|f\|_1 \|g\|_\infty$ if and only if $|g(x)| = \|g\|_\infty$ a.e. on the set where $f(x) \neq 0$
2. $\|\cdot\|_\infty$ is a norm on L^∞
3. $\|f_n - f\|_\infty \to 0$ if and only if there exists $E \in \mathcal{M}$ such that $\mu(E^c) = 0$ and $f_n \to f$ uniformally on E
4. L^∞ is a Banach space
5. The simple functions are dense in L^∞

As a consequence of the well-definedness of L^∞, $1^{-1} + \infty^{-1} = 1$ is indeed a well-defined notion (so $\infty^{-1} = 1/\infty = 0$). Thus, 1 and ∞ are conjugate exponents!

Proof **:**
We will prove (1), from which the rest of the results follow.

1. We'll be using the $\|f\|_\infty = \inf\limits_{\substack{E\in\mathcal{M} \\ \mu(X\setminus E)=0}} (\sup_{x\in E} |f(x)|)$ definition of the essential supremum to take advantage of the supremum within the definition. Our goal is to show that
$$\int |fg| = \|fg\|_1 \leq \|f\|_1 \|g\|_\infty$$
First, we will limit our attention of $\int |fg|$ to some $E \in \mathcal{M}$ such that $\mu(X \setminus E) = \mu(E^c) = 0$ (i.e. ignoring a zero-measure set). Then:
$$\int |fg| = \int_E |fg| + \int_{E^c} |fg| = \int_E |fg|$$
Next, we take advantage the fact that we are working with an infimum to see that:
$$\|f\|_\infty \leq \sup_{x\in A} |f(x)| < \|f\|_\infty + \epsilon$$
Therefore, we can take:
$$\int_E |fg| \leq \int_E \left| f \cdot \sup_{x\in E} |g(x)| \right| = \int_E |f| \cdot \left| \sup_{x\in E} |g(x)| \right| = \left| \sup_{x\in E} |g(x)| \right| \int_E |f|$$
where the last equality comes from the fact that the supremum is a constant. Finally, by the property of the infimum, we have:
$$\left| \sup_{x\in E} |g(x)| \right| \int_E |f| \leq (\|g\|_\infty + \epsilon) \int |f|$$

Notice that this inequality does not depend on ϵ, and so:

$$\|fg\|_1 = \int_E |fg| \leq \|f\|_1 \|g\|_\infty$$

as we sought to show

Notice how similar $\|\cdot\|_\infty$ is to the uniforma metric $\|f\|_u = \sup_{x \in X} f(x)$. If we are dealing with Borel measures that assign a positive value to all open sets, then if f is continuous we in fact have $\|f\|_\infty = \|f\|_u$ (since $\{x \mid |f(x)| > a\}$ is open). In this way, we can consider the space of bounded continuous functions as a (closed!) subspace of L^∞.

We conclude with some final remarks about the density of simple functions in L^∞. Notice that for L^p where $p < \infty$, simple functions where $\mu(E_i) < \infty$ where dense. This condition is no longer present for density in L^∞; we in fact require simple where $\mu(E) = \infty$ now since the function $1(x) = 1$ is in $L^\infty(\mathbb{R})$. This also means that $C_0(\mathbb{R})$ cannot be dense in L^∞. In the case of $L^\infty([a,b])$ $C([a,b])$, not all funcions can be converged to uniformally dense: consider $f : [0,1] \to \mathbb{R}$, $f(x) = 0$ for $x \in [0, 1/2]$ and $f(x) = 1$ for $x \in [1/2, 1]$. The no continuous function uniformally converges to f in the ∞-norm.

5.1.1 Relating L^p spaces: Interpolation

In this section, we will show how we can relate L^p spaces with each other. Fundamentally, the way we get any order with L^p spaces is through Hölder's inequality. We take full advantage of it in the following proofs. The end-goal is to show that if $p < q < r$ where we allow $r = \infty$, then:

$$L^p \cap L^r \subseteq L^q \subseteq L^p + L^r$$

This should reflect on the example that we have demonstrated in example 5.1. In particular, the larger p is, the less f has decay at infinity, but the more f must not explode at any point. In other words p gets bigger, we can let p decay slower, but it must behave better at points where it can explod3. If our space X is bounded, then we don't have to care about decay (this will give us that $L^p \supseteq L^q$ where $p < q$). If X is discrete, we don't have to care about exploding at any point (this will give us that $\ell^p \subseteq \ell^q$ if $p < q$). We thus explore these relation's in this section. These types of relations bewteen space via addition of functions or intersection of spaces is called *interpolation*.

Proposition 5.1.2: L^p subset

Let $0 < p < q < r \leq \infty$. Then

$$L^q \subseteq L^p + L^r$$

In particular, each $f \in L^q$ is the sum of a function in L^p and a function in L^q

Proof **:**
This proof is really similar to that of example 5.1[1]. Let $f \in L^q$. Then consider:

$$|f| = \chi_{|f| \geq 1} |f| + \chi_{|f| < 1} |f|$$

that is, we're splitting $|f|$ to where f is greater than or equal to 1 and strictly less than 1. The equality holds since $A = \{x \mid |f(x)| \geq 1\}$ and $B = \{x \mid |f(x)| < 1\}$ are disjoint. Importantly, notice that:

$$|f|^p = \chi_{|f|\geq 1}|f|^p + \chi_{|f|<1}|f|^p \qquad |f|^r = \chi_{|f|\geq 1}|f|^r + \chi_{|f|<1}|f|^r$$

Thus, by the linearity of the integral, we get that:

$$\int |f|^p = \int \chi_{|f|\geq 1}|f|^p + \int \chi_{|f|<1}|f|^p \qquad \int |f|^r = \int \chi_{|f|\geq 1}|f|^r + \int \chi_{|f|<1}|f|^r$$

Then we get:

$$\chi_{|f|\geq 1}|f|^p \leq \chi_{|f|\geq 1}|f|^q < \infty \qquad \chi_{|f|<1}|f|^r \leq \chi_{|f|<1}|f|^q < \infty$$

Therefore, $\chi_{|f|\geq 1} f \in L^p$ and $\chi_{|f|<1} f \in L^r$. If $r = \infty$, then clearly $\|\chi_{|f|\leq 1} f\|_\infty \leq 1$, showing the result still holds, as we sought to show.

Proposition 5.1.3: L^p superset

Let $0 < p < q < r \leq \infty$. Then

$$L^p \cap L^r \subseteq L^q$$

In particular, if if we find that f is bounded in growth from above and bellow by p and r, $\|f\|_p < \infty$ and $\|f\|_r < \infty$, then f is bounded in growth by q.

Proof **:**

Let $\|f\|_r < \infty$ and $\|f\|_q < \infty$. We need to show that $\|f\|_p < \infty$. Notice that it is equivalent to show that $\|f\|_r^n < \infty$ and $\|f\|_q^m < \infty$ then $\|f\|_p^\ell < \infty$ for $n, m, \ell \in \mathbb{R}_{\geq 0}$ (since if it's finite, then all powers will be finite). We'll prove it in the case where $r = \infty$ and $r < \infty$. In the infinite case, we simply take advantage of bounding properties of the essential supremum, and in the other case we use Hölder's inequality.

Starting with the $r = \infty$ case, notice that

$$|f|^q = |f|^{q-p+p} = |f|^q |f|^{q-p} \leq \|f\|_\infty^{q-p} |f|^p$$

and so, taking the integral and the qth root, we get:

$$\begin{aligned}
\|f\|_q &= \left(\int |f|^q\right)^{1/q} \\
&\leq \left(\int \|f\|_\infty^{q-p} |f|^p\right)^{1/q} \\
&= \|f\|_\infty^{\frac{q-p}{q}} \left(\int |f|^p\right)^{1/q} \\
&= \|f\|_\infty^{\frac{q-p}{q}} \left(\int |f|^p\right)^{p/pq} \\
&= \|f\|_\infty^{1-\frac{p}{q}} \|f\|_p^{p/q} < \infty
\end{aligned}$$

The fact that the last inequality is less than infinity comes from the that $\|f\|_p < \infty$ and $\|f\|_\infty < \infty$, so any power is less than infinity, and so $\|f\|_q < \infty$. Furthermore, notice that the sum of the power's is 1; this fact will come back for another intuition of $\|f\|_\infty$.

For the $r < \infty$ case the key is to find a way to integrate p and r using conjugate exponents. Since

$$p < q < r$$

then

$$\frac{1}{p} > \frac{1}{q} > \frac{1}{r}$$

therefore, define the continuous function $\lambda \mapsto \frac{1-\lambda}{p} + \frac{\lambda}{r}$. Since $(0,1)$ is open, and $1/p > 1/r$, there exists a λ_0 such that

$$\frac{1-\lambda_0}{p} + \frac{\lambda_0}{r} = \frac{1}{q}$$

manipulating the equation around to get conjugate exponents, notice that:

$$\frac{(1-\lambda_0)q}{p} + \frac{\lambda_0 q}{r} = 1 \quad \Leftrightarrow \quad \frac{1}{\frac{p}{(1-\lambda_0)q}} + \frac{1}{\frac{r}{\lambda_0 q}} = 1$$

Thus, we have that $\frac{p}{(1-\lambda_0)q}$ and $\frac{r}{\lambda_0 p}$. With conjugate exponents established, we proceed with trying to bound $\|f\|_q$. As we commented, it is equivalent to bound $\|f\|_q^m$, so let $m = q$ (this will be a common trick to apply Hölder's inequality). Also, since $1 - \lambda_0 + \lambda_0 = 1$, we have that $q = \lambda_0 q + (1-\lambda_0)q$, giving us the ability to split up our integral:

$$\begin{aligned}
\|f\|_q^q &= \int |f|^q \\
&= \int |f|^{\lambda_0 q + (1-\lambda_0)q} \\
&= \int |f|^{\lambda_0 q} |f|^{(1-\lambda_0)q} \\
&= \left\| |f|^{\lambda_0 q} |f|^{(1-\lambda_0)q} \right\|_1 \\
&\leq \left\| |f|^{\lambda_0 q} \right\|_{\frac{p}{(1-\lambda_0)q}} \left\| |f|^{(1-\lambda_0)q} \right\|_{\frac{r}{\lambda_0 q}} && \text{Hölder's inequality} \\
&= \left(\int |f|^p \right)^{\frac{(1-\lambda_0)q}{p}} \left(\int |f|^r \right)^{\frac{\lambda_0 q}{r}} \\
&= \|f\|_p^{(1-\lambda_0)q} \|f\|_r^{\lambda_0 q} < \infty
\end{aligned}$$

Since the last two terms are finite, any power of them are also finite, and so we have bounded $\|f\|_q^q < \infty$, and so $\|f\|_q < \infty$, as we sought to show.

Proposition 5.1.4: ℓ^p Covariant Inclusion

If $0 < p < q \leq \infty$, then $\ell^p(A) \subseteq \ell^q(A)$

Proof **:**
They key part is to take advantage of the properties of the counting measure. Since the only set of measure 0 in the counting measure is the empty-set, we have that the essential supremum is actually the supremum!

Let's say $f \in \ell^p$, so that $\sum_{a\in A} |f(a)|^p = \|f\|_p < \infty$. We'll start by showing that $\|f\|_\infty < \infty$. Since the $\|f\|_\infty$ is in fact the supremum, we have that:

$$\|f\|_\infty^p = \sup_{a\in A} |f(a)|^p \leq \sum_{a\in A} |f(a)|^p = \|f\|_p < \infty$$

Therefore, $\|f\|_\infty < \infty$.

For the $q < \infty$ case, using proposition 5.1.3 with $\lambda = p/q$, we get:

$$\|f\|_q \leq \|f\|_p^\lambda \|f\|_\infty^{1-\lambda} \leq \|f\|_p$$

therefore, $\|f\|_q < \infty$, completing the proof.

Proposition 5.1.5: L^p Contravariant Inclusion in finite measure

Let $(X, \mathcal{M}, \mu)$ be a measure space where $\mu(X) < \infty$. If $0 < p < q \leq \infty$, then $L^p \supseteq L^q$

Proof **:**
We need to show that if $f \in L^q$ then $f \in L^p$. Since $f \in L^q$, $\|f\|_q < \infty$, so $\|f\|_q^n < \infty$ for all $n \in \mathbb{N}$. If we can show that $\|f\|_p^m \leq \|f\|_q^n$ for some $m, n \in \mathbb{N}$, then the proof is complete.

We'll apply Hölder's inequality in the caes of $q = \infty$ and $q < \infty$. If $q = \infty$, then:

$$\|f\|_p^p = \int |f|^p = \int |f|^p \cdot 1 \overset{\text{H.}}{\leq} \|f\|_\infty^p \int 1 = \|f\|_\infty^p \mu(X) < \infty$$

since $\int 1 = \mu(X)$. If $q < \infty$, then we do a similar argument. First, since $p < q$, notice that:

$$\frac{1}{\frac{q}{p}} + \frac{1}{\frac{q}{q-p}} = \frac{p}{q} + \frac{q-p}{q} = \frac{q}{q} = 1$$

and so $\frac{q}{p}$ and $\frac{q}{p-q}$ are conjugate exponents. Therefore, by Hölder's inequality:

$$\begin{aligned}
\|f\|_p^p &= \int |f|^p \\
&= \int |f|^p \cdot 1 \\
&= \||f|^p \cdot 1\|_1 \\
&\leq \||f|^p\|_{\frac{q}{p}} \|1\|_{\frac{q}{p-q}} && \text{Hölder's Inequality} \\
&= \left(\int |f|^{p\frac{q}{p}}\right)^{\frac{p}{q}} \left(\int 1\right)^{\frac{q-p}{p}} \\
&= \|f\|_q^p \, \mu(X)^{\frac{q-p}{p}} < \infty
\end{aligned}$$

therefore, $\|f\|_p^p < \infty$, so $\|f\|_p < \infty$, showing that $f \in L^p$, as we sought to show.

Proposition 5.1.6: Relating L^p and L^∞

Let $p \in [1, \infty)$ and assume $\|f\|_p < \infty$. Show that:

$$\lim_{p\to\infty} \|f\|_p = \|f\|_\infty$$

***Proof* :**

For ease of typing, we simplify notation: $\|f\|_p := \|f\|_{L^p(\mu)}$

We'll show that $\lim_{p\to\infty} \|f\|_p \leq \|f\|_\infty$ and $\lim_{p\to\infty} \|f\|_p \geq \|f\|_\infty$. For the $\leq$ direction, note that $|f| \leq \|f\|_\infty$ a.e., and since x^n is a monotone function on $[0, \infty)$, we have $|f|^n \leq \|f\|_\infty^n$. Then for some fixed λ_0:

$$\begin{aligned}
\|f\|_p &= \left(\int |f|^p\right)^{1/p} \\
&= \left(\int |f|^{p+\lambda_0-\lambda_0}\right)^{1/p} \\
&= \left(\int |f|^{p-\lambda_0}|f|^{\lambda_0}\right)^{1/p} \\
&\leq \|f\|_\infty^{\frac{p-\lambda_0}{p}} \|f\|_{\lambda_0}^{\frac{\lambda_0}{p}}
\end{aligned}$$

where the last inequality is the same trick as part (a) of this question. Since λ_0 is fixed and this inequality holds true for all p, as $p \to \infty$:

$$\lim_{p\to\infty} \|f\|_p \leq \lim_{p\to\infty} \|f\|_\infty^{\frac{p-\lambda_0}{p}} \|f\|_{\lambda_0}^{\frac{\lambda_0}{p}} = \|f\|_\infty$$

For the $\geq$ direction, we have to bound $\|f\|_p$ from bellow by something that includes $\|f\|_\infty$, but ultimately the inequality will be independent of the choices we make. In particular, let $\epsilon > 0$. Take all points which are epsilon close to the essential supremum (which exists by the infimum property). Let M_ϵ be our set, in particular $M_\epsilon = \{x \mid |f(x)| \geq \|f\|_\infty - \epsilon\}$. Since $\|f\|_\infty = k$, then $\mu(M_\epsilon) > 0$. Thus:

$$\|f\|_p = \left(\int |f|^p\right)^{1/p} \geq \left(\int_{M_\epsilon} (\|f\|_\infty - \epsilon)^p\right)^{1/p} \stackrel{!}{=} (\|f\|_\infty - \epsilon)\mu(M_\epsilon)^{1/p}$$

where $\stackrel{!}{=}$ comes from the fact that we are integrating over a constant region. Since $\mu(M_\epsilon) > 0$, we have that $\mu(M_\epsilon)^{1/p} > 0$ for all p, and $\mu(M_\epsilon)^{1/p} \xrightarrow{p\to\infty} 1$. Furthermore, notice our inequalities are independent of choice of ϵ, and so since we can let ϵ be as small as we want:

$$\lim_{p\to\infty} \|f\|_p \geq \|f\|_\infty$$

and since we've shown the $\leq$ direction, we have:

$$\lim_{p\to\infty} \|f\|_p = \|f\|_\infty$$

as we sought to show

Now that we have examined L^p spaces in more detail, we finish off this seciton with some comments on the usefulness these spaces:

1. L^1 is natural for integration; we will essentially work with variation's of L^1 when we integrate. From an algebraic perspective, it can be thought of as a generalizatgion of the notion of "counting" all points in a space, and working in spaces where the count is finite.
2. L^∞ is useful as a broadening of the space of uniform-bound functions. It is a generlization of the idea of working with functions that are bounded.
3. L^p interpolates in between L^1 and L^∞. Many times, L^1 or L^∞ spaces have pathological behaviors (examples of which I yet do not know). L^p spaces behave more nicely, and can sometimes be used to go from L^1 to L^∞ step by step using the inequalities we constructed (ex. we got a function is L^1, we want to show it's L^∞, and we use the proceeding inequalities to build our way up)
4. L^2 is particularly nice since it is a Hilbert space, and so many euclidean intuitions work on L^2 and help us construct many nice properties. A lot of Fourier analysis happens on L^2, though sometimes we do use L^p.

   Another way of thinking of L^2 is that it is a generlization of the notion of a space that *can* have an inner product (this will be made rigorous with a particular universal property, see theorem 4.5.10). In ℓ^2, we would simply get a natural generalization of an inner product on a vector-space, while in L^2, we get the continuous equivalent.

5.2 The dual of Lp

By Hölder's inequality, we have that if $f \in L^p$, then for all $g \in L^q$, $fg \in L^1$. We can thus define a function:

$$\phi_g : L^p \to \mathbb{R} \qquad \phi_g(f) = \int fg$$

and the operator norm of ϕ_g is always at most $\|g\|_q$[3]. With the opeartor norm, the mapping $g \mapsto \phi_g$ will almost always be an isometry:

Proposition 5.2.1: Operator Norm And Isometry

Let p and q be conjugate exponents, $1 \leq q < \infty$. If $g \in l^q$, then:

$$\|g\|_q = \|\phi_g\| = \left\{ \left| \int fg \right| \;\middle|\; \|f\|_p = 1 \right\}$$

if μ is semifinite, this results alos holds for $q = \infty$

***Proof* :**
Folland p. 188

[3] if $p = 2$, it is usually defined as mapping to $\int f\bar{g}$. This can also be done to $p \neq 2$, there it is just easier to work by conjugating for complex numbers

Conversely, if $f \mapsto \int fg$ is a bounded linaer functional on L^p, then $q \in L^q$ in almost all cases:

Theorem 5.2.1: L^p Linear Functional Implies G In L^q

Let p and q be conjugate exponents. Suppose that g is a measurable function on X such that $fg \in l^1$ for all f inteh space Σ of simple functions that vanish outside a set of finite measure, and the quanitty

$$M_q(g) = \{|fg| \mid f \in \Sigma,\ \|f\|_p = 1\}$$

is finite. Also, suppose either that $S_g = \{x \mid g(x) \neq 0\}$ is σ-finite or tha tμ is semi-finite. Then $g \in L^q$ and $M_q(g) = \|g\|_q$.

Proof **:**
Folland p. 189

Finally, perhaps the most stricking thing is that the maps $g \mapsto \int fg$ are almost always a surjection of $(L^p)^*$!

Theorem 5.2.2: Description Of Dual Of Lp

Let p and q be conjugate exponents. If $1 < p < \infty$, for eac h$\phi \in (L^p)^*$ there exits $g \in L^q$ such that $\phi(f) = \int fg$ for all $f \in L^p$, and ehcne L^q is isometrically isomorphic to $(L^p)^*$. The same conclusion holds for $p = 1$ provided μ is $1s$-finite

Proof **:**
Folland p. 190

Corollary 5.2.1: Lp Is Reflective

If $1 < p < \infty$, L^p is reflective, that is $(L^p)^{**} \cong L^p$

(Some comment in the case of $p = 1$ and $p = \infty$)

5.3 Some Useful Inequalities

In this section, we explore more inequalties, since they are at the heart of working with L^p spaces. We first present a common inequality that shows up often in statistics:

Theorem 5.3.1: Chebychev's Inequality

Let $f \in L^p$ $(0 < p < \infty)$. Then for any $\alpha > 0$

$$\mu(\{x \mid |f(x)| > \alpha\}) \leq \left(\frac{\|f\|_p}{\alpha}\right)^p$$

Proof **:**
Let $E_\alpha = m(\{x \mid |f(x)| > \alpha\})$. Then:

$$\|f\|_p^p = \int_\Omega |f|^p \geq \int_{E_\alpha} |f|^p \geq \int_{E_\alpha} \alpha^p = \alpha^p m(E_\alpha)$$

re-arranging gives us:

$$\mu(\{x \mid |f(x)| > \alpha\}) \leq \left(\frac{\|f\|_p}{\alpha}\right)^p$$

as we sought to show

The next inequality gives us a general statement on being abel to get a function from $L^p(\mu)$ "transformed" into a function of $L^p(\nu)$:

Theorem 5.3.2: Transitioning L^p Spaces

Let$(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be σ-finite measures spaces, K be a $M \otimes N$-measurable function on $X \times Y$, and let's say there exists a $C > 0$ such that

$$\int |K(x,y)| d\mu(x) \leq C \qquad \int |K(x,y)| d\nu(y) \leq C$$

for a.e. x and y. Then if $f \in L^p(\nu)$ $(1 \leq p \leq \infty)$, the integral:

$$Tf(x) = \int K(x,y) f(y) d\nu(y)$$

converges absolutely for a.e. $x \in X$, $Tf \in L^p(\mu)$, and $\|Tf\|_p \leq C\|f\|_p$

Proof **:**
Let $p \in [1, \infty]$, and let q be it's conjugate exponent. Then:

$$|K(x,y)f(y)| = |K(x,y)||f(y)| = |K(x,y)|^{\frac{1}{p}+\frac{1}{q}}|f(y)| = (|K(x,y)|^{1/p})(|K(x,y)|^{1/q} f(y)|)$$

and applying Hölder to the functions in parenthesis, we get:

$$\begin{aligned}\int |K(x,y)f(y)| d\nu(y) &\leq \left(\int |K(x,y)| d\nu(y)\right)^{1/q} \left(\int |K(x,y)||f(y)|^p d\nu(y)\right)^{1/p} \\ &\leq C^{1/q} \left(\int |K(x,y)| f(y)|^p d\nu(y)\right)^{1/p}\end{aligned}$$

for a.e. $x \in X$. Hence, taking the integral over $X \times Y$, using Tonelli's theorem, and substituting the result we just found, we get:

$$\begin{aligned}\int\left[\int |K(x,y)f(y)d\nu(y)\right]^p d\mu(x) &\le \int\int C^{p/q}|K(x,y)||f(y)|^p d\nu(y)d\mu(x) \\ &= C^{p/q}\int\int |K(x,y)||f(y)|^p d\mu(x)d\nu(y) && \text{Tonelli} \\ &\le C^{p/q}\int |f(y)|^p \int K(x,u)d\mu(x)d\nu(y) \\ &= C^{p/q+1}\int |f(y)|^p d\nu(y)\end{aligned}$$

Since $f \in L^p(\nu)$, the last integral is finite. Since this is true for a.e. x, $K(x,\cdot)f \in L^1(\nu)$ for a.e. x. Thus, Tf is well-defined a.e., meaning we can integrate it, and by the inequality we just showed:

$$\int |Tf(x)|^p d\mu(x) = \|Tf\|_p^p \le C^{p/q+1}\|f\|_p^p$$

taking the pth root of each side get's us the desired inequality

The next result we work towards is a generalization of Minkowski: instead of summing we will be taking integrals. You can also interpret part (a) as you can move the $(|\cdot|^p)^{1/p}$ "down" a layer of integrals, and part (b) as the generalization of monotonicity ($|\int f| \le \int |f|$).

Theorem 5.3.3: Minkowski's Inequality For Integrals

Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be σ-finite meaure spaces and let f be a [real?] $M \otimes N$-measurable function on $X \times Y$. Then:

1. if $f \ge 0$ and $1 \le p \le \infty$, then:

$$\left[\int\left(\int f(x,y)d\nu(y)\right)^p d\nu(x)\right]^{1/p} \le \int\left[\int f(x,y)^p d\mu(x)\right]^{1/p} d\nu(y)$$

2. If $1 \le p \le \infty$, $f(\cdot, y) \in L^p(\mu)$ for a.e. y and the function $y \mapsto \|f(\cdot,y)\|_p$ is in $L^1(\nu)$, then $f(x,\cdot) \in L^1(\nu)$ for a.e. x, the function $x \mapsto \int f(x,y)d\nu(y)$ is in $L^p(\mu)$ and

$$\left\|\int f(\cdot,y)d\nu(y)\right\|_p \le \int \|f(\cdot,y)\|_p \, d\nu(y)$$

Proof **:**
If $p = 1$, then this becomes Tonelli's Theorem, so consider $1 < p < \infty$ (the case for $p = \infty$ comes later). Let q be it's conjugate exponent so that $1/p + 1/q = 1$. Suppose $\mu, \nu < \infty$ and $\|f\|_\infty < \infty$ (i.e. it's essential supremum is finite). For notational simplicity, define $H(x) = \int_Y f(x,y)d\nu(y)$.

Since the measure is finite and f is esentially bounded, we have that $\|H\|_p < \infty$. Then:

$$\begin{aligned}
\|H\|_p^p &= \int_X \left(\int_Y f(x,y) d\nu(y) \right)^p d\nu(x) \\
&= \int_X \left(\int_Y f(x,y) d\nu(y) \right) (H(x))^{p-1} d\mu(x) && x \text{ indep. of } y \\
&= \int_X \int_Y f(x,y) H(x)^{p-1} d\mu(x) d\nu(y) && \text{Fubini} \\
&\overset{!}{\leq} \int_X \|f(x,\cdot)\|_p \|H\|_p^{p-1} d\nu(y) && \text{Hölder, } q = p/(p-1) \\
&\leq \|H\|_p^{p-1} \int_X \|f(x,\cdot)\|_p d\nu(y)
\end{aligned}$$

where $\overset{!}{\leq}$ I'm not sure how Hölder's inequality was used. Finally, dividng by $\|H(x)\|_p^{p-1}$ get's us:

$$\|H\|_p \leq \int_X \|f(x,\cdot)\|_p d\nu(y)$$

For the general case, Yakov left it as an exercise to do $X = \bigcup_n X_n$, $Y = \bigcup Y_n$, $E_n = \{x \mid f(x) < n\}$, and

$$\tilde{X}_n = X_n \cap E_n \qquad \tilde{Y}_n = Y_n \cap E_n$$

and apply the result to each of these, and the monotone convergence theorem at the end (feels similar to how it was done for Fubini Tonnelli).

For $p = \infty$, This comes down to monotonicity of integrals:

$$\left\| \int f(x,y) d\nu(y) \right\|_\infty \leq \left| \int f(x,y) d\nu(y) \right| \leq \int |f(x,y) d\nu(y) = \int \|f(x,y)\|_\infty d\nu(y)$$

where the last inequality comes from the fact that the integral ignores zero-measure sets.

There are a few more results in the book, but Yakov decided to prove the generalized Hölder's inequality instead for now:

Theorem 5.3.4: Generalized Hölder's Inequality

Let $1 \leq p_i \leq \infty$ for $1 \leq i \leq n$ and $\sum_i^n p_i^{-1} = r^{-1} \leq 1$. If $f_i \in L^{p_i}$, then $\prod_i^n f_i \in L^r$ and

$$\left\| \prod_i^n f_i \right\|_r \leq \prod_i^n \|f_i\|_{p_i}$$

***Proof* :**

If $r = \infty$, then $r^{-1} = 0$, so all the $p_i^{-1} = 0$, so all the $p_i = \infty$, in which case we simply take:

$$\left| \prod_i^n f_i \right| \leq \prod_i^n |f_i|$$

For $r < \infty$, just like for Hölder's inequality, it suffice to assume $r = 1$, since we can do:

$$\sum \frac{1}{\frac{p_i}{r}} = 1$$

In that case, notice that:

$$\begin{aligned}
\left\| \prod_i^n f_i \right\|_r^r &= \int |f_1 f_2 \cdots f_n|^r \\
&= \int f_1^r f_2^r \cdots f_n^r \\
&\leq \|f_1\|_{p_1}^r \|f_2\|_{p_2}^r \cdots \|f_n\|_{p_n}^r
\end{aligned}$$

<++> Notice that each f_i is raised to the power of p_i/r. (TBD)

Now, we do induction on i. Let the induction hypothesis be:

$$\left(\sum_i^{n-1} p_i^{-1} \right) + p_n^{-1} = r^{-1} + p_n^{-1} = 1$$

Then by Hölder's inequality:

$$\int f_1 \cdots f_n| \leq \|f_1 \cdots f_{n-1}\|_r \|f\|_{p_n} \overset{\text{I.H.}}{\leq} \|f_1\|_{p_1} \cdots \|f_n\|_{p_n}$$

giving our desired result

The final property we'll present a way to represent functions in terms of the integral:

Proposition 5.3.1: Layer Cake Representation

Let $\phi(t) : [0, \infty] \to [0, \infty]$ be an increasing function which is C^1 on $(0, \infty)$, satisfies $\phi(0) = 0$ and satisfies $\lim_{t\to\infty} \phi(t) = \phi(\infty)$. Let (X, M, μ) be a σ-finite measure space and $f : X \to [0, \infty]$ be a non-negative measurable function. Then:

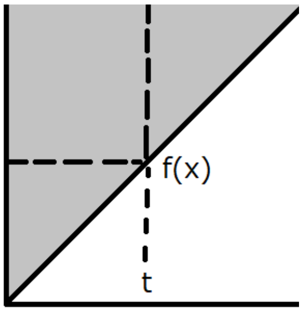
$$\int_X \phi(f(x))\, d\mu(x) = \int_0^\infty \mu(\{x \mid f(x) > t\})\phi'(t)dt$$

One consequence of this equation is that with $\phi(t) = t^p$ for $1 < p < \infty$ is the following formula for the L^p norm of f:

$$\|f\|_p = \left(p \int_0^\infty \mu(\{x \mid f(x) > t)\}\, t^{p-1} \right)^{1/p} dt$$

Proof **:**

The proof can be captured pictorially as integrating over the following area:



Let $S \subseteq X \times [0,1]$, $S = \{(x,t) \mid f(x) \geq t\}$. With this set, define $F : X \times [0,\infty] \to [0,\infty]$, $F(x,t) = \phi'(t)\chi_S$. That is, we are going to integrate the "upper triangle". Since ϕ is an increasing C^1 function and $\phi'(t) > 0$, we have $F \in L^+(X, [0,\infty])$. Therefore, we can apply Tonneli's Theorem and see that it's equivalent to integrate over the iterated integrals to compute integrals.

To compute it, we need to be clever with the bounds we choose which depend on which iterated integral we choose. The different direction of the iterated integral will ultimately lead to the desired result. If we integrate with respect to $d\mu(x)dt$, then we take slices across t: $(\phi'(t)\chi_S)_t = \phi'(t)\chi_{S_t}$, where $(\chi_S)_t = \chi_S(x,t) = \chi_{S_t}$, with $S_t = \{x \mid f(x) \geq t\}$. And so we get:

$$\int_0^\infty \int_X \chi_{S_t}\phi'(t) = \int_0^\infty \int_{\{x \ \mid \ f(x) \geq t\}} \phi'(t) d\mu(x) dt = \int_0^\infty \mu(\{x \mid f(x) \geq t\})\phi'(t)dt$$

Now, integrating with respect to $dtd\mu(x)$, we take x slices: $(\phi'(t)\chi_S)_x = \phi'(t)\chi_{S_x}$ where $(\chi_S)_x = \chi_S(x,t) = \chi_{S_x}$, with $S_x = \{x \mid 0 \leq f(x)\}$. And so we get:

$$\int_X \int_0^\infty \chi_{S_x}\phi'(t)dtd\mu = \int_X \int_0^{f(x)} \phi'(t)dtd\mu(x)$$

and since $[0, f(x)]$ is bounded and $\phi(0) = 0$, the Lebesgue integral matches the Riemann integral, and so we get that:

$$\int_X \phi(f(x))d\mu(x)$$

and by Tonnelli's theorem:

$$\int_X \phi(f(x))d\mu(x) = \int_0^\infty \mu(\{x \mid f(x) \geq t\})\phi'(t)dt$$

As we sought to show. Using this, if we let $\phi(t) = t^p$, then we see that ϕ is C^1 on $(0,\infty)$ and $\phi(0) = 0$ for $1 \leq p < \infty$. Thus, the value of $\|f\|_p$ (where we take f^p instead of $|f|^p$ since $f \geq 0$) is:

$$\|f\|_{L^p} = \left(\int |f|^p\right)^{1/p} = \left(\int f^p\right)^{1/p} = \left(\int_X \phi(f(x))d\mu(x)\right)^{1/p} = \left(p\int_0^\infty \mu(\{x \mid f(x) > t\})\, t^{p-1}\right)^{1/p} dt$$

Another use of this formula is by takign $\phi(t) = t$. Then we will get that:

$$\int f(x)d\mu(x) = \int_0^\infty \mu(\{x \mid f(x) \geq t\})dt$$

This can be useful if we have information of what is happening to the growth of f as f get's larger, in particular if we know that somehow, the amount of points such that $\mu(\{x \mid f(x) \geq t\})$ is shrinking rapidly enough say that $\mu(\{x \mid f(x) \geq t\}) \leq \frac{C}{t^r}$ for some constant C).

Chapter 6

Fourier Analysis

In the 19th century, Joseph Fourier was trying to find out how to model the movement of heat through solid bodies, in particular model how heat seems to dissipate and homogenise within its medium. For simplicity, let's say this body was a rode, and we can represent any point on the rode with the interval $[a, b]$. If $u : [0, t] \times [a, b] \to \mathbb{R}$ is a function that at any given point $c \in [a, b]$ and at any given time $t_0 \in [0, t]$ give you the temperature, then Fourier devised that the function u must respect the following differential equation:

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}$$

which can be interpreted that in any movement in time for some point $c \in [a, b]$, the temperature at c will equal to the *average* of its neighbor (see 3Blue1Brown's video on Fourier analysis for an excellent visualization of this), and $u|_{t_0=0}$ represents the initial distribution of heat in the rode $[a, b]$. Working with this equation, Fourier saw that trigonometric identities are a solution for u given the restraint (with suitable scaling[1]). The next insight Fourier had was that a huge amount of functions can be interpreted as a an infinium sum of trigonometric identities, and even better, he found how to find the appropriate coefficients for the trigonometric identites given a function! This is known as the *Fourier Series*, and going form a function to its Fourier series is called the *Fourier Transform*.

This was the beginning of Fourier Analysis, also known as Harmonic Analysis[2]. Since then, it was asked what spaces of functions admit a fourier transform, can we go back and forth between the fourier series and the original function, and can we generalize the domain of the functions (this will be pushed to be compact abelian group, and more genreally locally compact groups). The great advantage of being able to find the space of functions that have a Fourier transform is that (under suitable conditions) derivatives work really well with infinite series, trigonometric functions have some of the simple derivatives (they are periodic of order 4!), and the Fourier Transform allows us to transfer our problem into a potentially nice space (similarly to how in Representation Theory, it is much easier to find the representation of the monster group to study it rather than directly study it).

These days, Harmonic Analysis has found a large range of applications, from solving my different types of PDE's, to having (as of writing this book) mysterious connections to algebriac number

[1] In particular, some care must be given to the boundary conditions, but this is something that would be covered in more detail in a proper class in PDE's

[2] Sometimes, Harmonic Analysis is used as a generalization of Fourier Analysis in the case of non-abelian groups, something which will be made more clear in this chapter

theory (the persuit of this connection is known as *Langlands Program*). In this chapter, we will cover the elementary theory required to understand the theory of fourier analysis.

6.1 Basic Definitions

In this chapter, we will always work with $\mathbb{R}^n$ with the Lebesgue measure, denoted m. If $E \subseteq \mathbb{R}^n$ is measurable, $L^p(E)$ (or $L^p(E,m)$) is the set of all L^p functions from E to $\mathbb{R}$, $C^k(E)$ is the set of all k-times continuously differentiable functions from E to $\mathbb{R}$, $C^\infty(E)$ (the set of all smooth functions form E to $\mathbb{R}$) is equal to $\cap_{k=1}^\infty C^k(E)$, and $C_c^\infty(E)$ is the set of all smooth function whose compact support is in E. If $E = \mathbb{R}^n$, we will often drop the domain of the function space, so:

$$L^p = L(\mathbb{R}^n), \qquad C^\infty = C^\infty(\mathbb{R}^n) \qquad C_c^\infty = C_c^\infty(\mathbb{R}^n)$$

The inner product and norm on $\mathbb{R}^n$ is the usual dot prodcut and euclidean norm:

$$x \cdot y = \sum_i^n x_i y_i \qquad \|x\| = |x| = \sqrt{x \cdot x}$$

Derivative play a prominent role in this chapter, and so we establish important notation. We will assume the standard basis and write:

$$\partial_j = \frac{\partial}{\partial x^j}$$

and for higher orhder mixed derivatives, we will use a multi-index, which is an ordered n-tuple $\alpha = (\alpha_1, ..., \alpha_n)$. As usual:

$$|\alpha| = \sum_1^n \alpha_i \quad \alpha! = \prod_1^n \alpha_i! \quad \partial^\alpha = \left(\frac{\partial}{\partial x_i}\right)^{\alpha_1} \cdots \left(\frac{\partial}{\partial x_n}\right)^{\alpha_n}$$

and for any $x = (x_1, x_2, ..., x_n) \in \mathbb{R}^n$:

$$x^\alpha = \prod_1^n x_i^{\alpha_i}$$

Be aware that $|\alpha|$ and $|x|$ do not means the same thing; the meaning should be clear from context. For reference, the product rule on a multi-index expands to:

$$\partial^\alpha(fg) = \sum_{\beta+\gamma=\alpha} \frac{\alpha!}{\beta!\gamma!} (\partial^\beta f)(\partial^\gamma g)$$

A common function we'll use a lot in C_c^∞ is:

$$\psi(x) = \eta(1 - |x|^2) = \begin{cases} \exp((|x|^2 - 1)^{-1})) & |x| < 1 \\ 0 & |x| \geq 1 \end{cases}$$

where $\eta(t) = e^{1/t}\chi_{(0,\infty)}(t)$. Another common subset of C^∞ that will be of use is the following which insures that a function and all its derivatives decay faster than polynomial time:

Definition 6.1.1: Schwartz Space

The set $\mathcal{S} \subseteq C^\infty$ is the set of all smooth functions and all their derivatives which vanish faster than any power of $|x|$, that is, for any $n \in \mathbb{N}$ and any multi-index α, if we define:

$$\|f\|_{(N,\alpha)} = \sup_{x \in \mathbb{R}^n} (1+|x|)^n |\partial^\alpha f(x)|$$

Then:

$$\mathcal{S} = \left\{ f \in C^\infty \;\middle|\; \|f\|_{(N,\alpha)} < \infty \text{ for all } N, \alpha \right\}$$

Example 6.1: Schwartz Functions

1. Naturally, $C_c^\infty \subseteq \mathcal{S}$
2. $f_\alpha(x) = x^\alpha e^{-|x|^2}$ for any multi-index α is in $\mathcal{S}$

Note that if $f \in \mathcal{S}$, then $\alpha f \in L^p$ for all α and all $p \in [1, \infty]$ since $|\partial^\alpha f(x)| \leq C_N (1+|x|)^{-N}$ for all $N \in \mathbb{N}$, and $(1+|x|)^{-N} \in L^p$, for all $N > n/p$

Proposition 6.1.1: $\mathcal{S}$ is a Fréchet Space

The Schwartz space $\mathcal{S}$ is a Fréchet space with the topology defined by the norms $\|\cdot\|_{(N,\alpha)}$

Proof **:**
All is easy to prove with the possible exception of completeness. Let $\{f_k\}$ be a cauchy sequence in $\mathcal{S}$ so that $\|f_i - f_j\|_{(N,\alpha)} \to 0$ for all N, α. In particular, for each α the sequence $\{\partial^\alpha f_k\}$ converges uniformally to some g_α. Letting e_i denote the standard ith basis vector, we get:

$$f_k(x + te_i) - f_k(x) = \int_0^t \partial_i f_k(x + se_i) ds$$

Letting $k \to \infty$, we get:

$$g_0(x + te_i) - g_0(x) = \int_0^t g_{e_i}(x + se_i) ds$$

and the fundamental theorem of calculus tells us that $g_{e_i} = \partial_i g_0$, and so by induction we get $g_\alpha = \partial^\alpha g_0$ for all α. From this, it is now easy to verify that $\|f_k - g_0\|_{(N,\alpha)} \to 0$ for all α and $N \in \mathbb{N}$.

A useful characterization of $\mathcal{S}$ is also the following:

Proposition 6.1.2: Characterizing Schwartz Functions

If $f \in C^\infty$, Then $f \in \mathcal{S}$ if and only if $x^\beta \partial^\alpha f$ is bounded for all multi-indices α, β, if and only if $\partial^\alpha(x^\beta f)$ is bounded for all multi-indices α, β.

***Proof* :**
If $f \in \mathcal{S}$, then naturally $|x^\beta| \leq (1+|x|)^N$ for $\beta \leq N$. Conversely, notice that $\sum_1^n |x_i|^N$ is strictly positive on the unit sphere $|x| = 1$ (which is compact), and hence has a positive minimum, say δ. It follows that $\sum_1^n |x_i|^N \geq \delta |x|^N$ for all x since both sides are homogeneous of degree N, and hence:

$$(1+|x|)^N \leq 2^N(1+|x|^N) \leq 2^N \left[1 + \delta^{-1} \sum_1^n |x_i^N|\right] \leq 2^N \delta^{-1} \sum_{|\beta| \leq N} |x^\beta|$$

Giving the first equivalence. For the second equivalence, notice that for every α, $\partial^\alpha(x^\beta f)$ is a linear combination of terms of the form $x^\gamma \partial^\delta f$ and vice versa, by the product rool, completing the proof.

Next, it will be important to have compact notation for translating functions. If $f : R^n \to \mathbb{R}$, then

$$\tau_y f(x) = f(x-y)$$

Notice that:

$$\|\tau_y f\|_p = \|f\|_p \qquad \|\tau_y f\|_u = \|f\|_u$$

Using this, we can give another definition of uniformally continuous: f is uniformally continuous if:

$$\|\tau_y f - f\|_u \to 0$$

Lemma 6.1.1: Compact Support Functions And UC

If $f \in C_c(\mathbb{R}^n)$, then f is uniformally continuous

***Proof* :**
should be easy proof (Folland p.238)

Proposition 6.1.3: Translation Is Continuous In Lp

If $1 \leq p < \infty$, then translation is continuous in the L^p norm, in particular if $f \in L^p$ and $z \in \mathbb{R}^n$, then

$$\lim_{y \to 0} \|\tau_{y+z} f - \tau_z f\|_p = 0$$

Note this proposition is fale when $p = \infty$

***Proof* :**
Folland p. 238

Next we define a periodic function to be a function for which $f(x+k) = f(x)$ for all $x \in \mathbb{R}^n$ and $k \in \mathbb{Z}^n$. In other words, the period of the functions will be of size 1. Thus, we can consider the space on which we define periodic functions to be $\mathbb{R}^n/\mathbb{Z}^n \cong (\mathbb{R}/\mathbb{Z})^n \equiv \mathbb{T}^n = S^1 \times S^1 \times \cdots \times S^1$, i.e. we can think of defining period functions on an n-torus (note that $\mathbb{T}^1 = S^1$, but $\mathbb{T}^2 \neq S^2$). Naturally,

$\mathbb{T}^n$ is a compact Hausdorff space. It can be identified with all points $z = (z_1, z_2, ..., z_n) \in \mathbb{C}^n$ with $|z_i| = 1$ via the map

$$(x_1, x_2, ..., x_n) \mapsto (e^{2\pi i x_1}, ..., e^{2\pi i x_n})$$

For measure-theoretic purposes, we will sometimes identify $\mathbb{T}^n$ with the unit cube:

$$Q = \left[\frac{-1}{2}, \frac{1}{2}\right)^n$$

and when we talk about the Lebesgue measure on $\mathbb{T}^n$, we will mean the induced measure from Q. This means that $m(\mathbb{T}^n) = 1$. We will think of functions on $\mathbb{T}^n$ as either periodic functions on $\mathbb{R}^n$ or as functions on Q (the context will make it clear which we pick).

Exercise 2

1. Is the sum of two periodic functions periodic?

6.2 Convolution

In this section, we introduce a new way of multiplying two functions that will be very useful and very natural. The naturality comes from the discrete case, and so we will start with it: recall that if we have two polynomials $p(x)$ and $q(x)$ (which can be thought of as formal sums, bringing in the naturality), then the multiplication formula which groups the terms of the same degree is:

$$\left(\sum_{i=0}^{\deg(p)} p_i x^i\right) \cdot \left(\sum_{i=0}^{\deg(q)} q_j x^j\right) = \sum_{k=0}^{\deg(p)+\deg(q)} \left(\sum_{l=0}^{k} p_l q_{k-l}\right) x^k$$

If we p_k or q_{l-k} is zero if k or $l-k$ is negative, then we can re-write the sume as:

$$\left(\sum_{i=0}^{\deg(p)} p_i x^i\right) \cdot \left(\sum_{i=0}^{\deg(q)} q_j x^j\right) = \sum_{k=0}^{\deg(p)+\deg(q)} \left(\sum_{-\infty}^{\infty} p_l q_{k-l}\right) x^k$$

If we let $p[n]$ and $q[n]$ represent the coefficient of the nth degree term, and $h = pq$, then:

$$h[n] = \sum_{k=-\infty}^{\infty} p_n q_{n-k}$$

We can write this as:

$$h[n] = [p * q][n] = \sum_{k=-\infty}^{\infty} p_n q_{n-k}$$

where $p * q$ is a compact notation the right hand side. This, generlized to the continuous case, is called *convolution*:

Definition 6.2.1: Convolution

Let f, g be measurable on $\mathbb{R}^n$. Then the *convolution* of f and g is a function $f * g : \mathbb{R}^n \to \mathbb{R}$ defined by:

$$f * g(x) := \int f(x-y)g(y)dy$$

for all x such that the integral exists

To see this in action, Let $f = \chi_{[0,1]}$ and $g = \chi_{[0,1]}$, and you'll find that $f * g$ is a a triangle. You can imagine sliding the graph of f and finding the area of the intersection of f and g, and plotting it. There are many ways to make sure that $f * g$ is a.e. defined. For example, if f was bounded and compactly supported, g can be locally integrable. Note that in many cases, we will need that $K(x, y) = f(x-y)$ be measurable on $\mathbb{R}^n \times \mathbb{R}^n$. If we let $s(x, y) = x - y$, then since s is continuous, if f is Borel measurable, so is K. (Folland makes a comment here saying we can "assume" that this is the case, which I'm guessing is f being Borel measurable in the definition of convolution, becasue of his theorem 2.12, which is my proposition 2.1.6, and he says the same can happen for Lebesgue measurability). Usually, we will use convolution to somehow combine the property of two functions f and g. We will soon see that convolution often preserves the "nice" properties the "nicer" function, so that we would convolution, say, a smooth function f with an integrable function g to get a smooth function $f * g$.

The following results all work over $\mathbb{R}^n$ or $\mathbb{T}^n$ (interpreted as the unit cube)

Proposition 6.2.1: Algebraic Properties Of Convolution

Assume all the following integrals are defined [and converge absolutely, Yakov added]. Then:

1. $f * g = g * f$
2. $(f * g) * h = f * (g * h)$
3. $f * (g + h) = f * g + f * h$
4. $a(f * g) = (af) * g$ for all $a \in \mathbb{C}$
5. for all $z \in \mathbb{R}^n$, $\tau_z(f * g) = (\tau_z f) * g = f * (\tau_z g)$
6. If A is the closure of $\{x + y \mid x \in \operatorname{supp}(f),\ y \in \operatorname{supp}(g)\}$, then $\operatorname{supp}(f * g) \subseteq A$

There are more properties, but they are currently beyond the course

Proof **:**

1. $$\begin{aligned} f * g(x) &= \int f(x-y)g(y)dy \\ &= \int f(z)g(x-z)dz && z = x-y,\ y = x-z \\ &= \int g(x-z)f(z)dz \\ &= g * f(x) \end{aligned}$$

2. Follows from (1) from Fubini's theorem
3. Follows from direct computation
4. Follows from direct computation
5. Follows from direct computation and (1)
6. If $x \notin A$, then for any $y \in \text{supp}(g)$, we have $x - y \notin \text{supp}(f)$, thus $f(x-y)g(y) = 0$, i.e. $f * g(x) = 0$

One of the first properties important to establish is the relation between convolution of a function and p-norms:

Theorem 6.2.1: Young's Inequality

Let $1 \leq p, q, r \leq \infty$ be such that $p^{-1} + q^{-1} = r^{-1} + 1$, and let $f \in L^p$ and $g \in L^q$. Then $f * g(x)$ exists for a.e. x, $f * g \in L^r$ and

$$\|f * g\|_r \leq \|f\|_p \|g\|_q$$

Notice in the statement that we have something a little different than the usual conjugate exponents. I'll explain this soon. There is another thing worth pointing out, which is that this condition is fixed under scaling. In particular, if $f_\lambda(x) = f(\lambda x)$ and $g_\lambda(x) = g(\lambda x)$. Then:

$$\begin{aligned} \|f_\lambda\|_p &= \left(\int_{\mathbb{R}^n} f(\lambda x)^p \right)^{1/p} \\ &= \left(\int_{\mathbb{R}^n} \lambda^{-n} f(y)^p \right)^{1/p} && y = \lambda x,\ dy = \lambda^n dx \\ &= \lambda^{-n/p} \|f\|_p \end{aligned}$$

where λ^n is the determinant of the Jacobian matrix. So:

$$\begin{aligned} (f_\lambda * g_\lambda)(x) &= \int_{\mathbb{R}^n} f(\lambda x - \lambda y) g(\lambda y) dy \\ &= \lambda^{-n} (f * g)(\lambda x) \end{aligned}$$

Thus we have:

$$\begin{aligned}\|f_\lambda * g_\lambda\|_r &= \lambda^{-n}\|(f*g)_\lambda\|_r \\ &= \lambda^{-n/r-n}\|f*g\|_r \\ &\leq \lambda^{-n/p-n/q}\|f\|_p\|g\|_q\end{aligned}$$

which will imply

$$\|f*g\|_r \leq \frac{1}{2}\|f\|_p^2 + \frac{1}{2}\|g\|_q^2$$

which somehow implies the condition of $p^{-1}+q^{-1}=r^{-1}+1$

***Proof* :**
First, consider $r=\infty$, meaning $r^{-1}=0$, $p^{-1}+q^{-1}=1$. So, we want to show $\|f*g\|_1 \leq \|f\|_p\|g\|_q$. So:

$$\begin{aligned}\|f*g\|_1 &= \int |f(x-y)g(y)|dy \\ &\leq \|f(x-\cdot)\|_p\|g\|_q \\ &\overset{!}{=} \|f\|_p\|g\|_q\end{aligned}$$

where $\overset{!}{=}$ comes from $\int |h(x-y)|dy = \int |h(y)|dy$ where $z=x-y$, $dz=(-1)^n dz$ so the $(-1)^n$ dissapears in the absolute value.

Next, if p or q is 1 (say $p=1$), then $1+q^{-1}=r^{-1}+1$, so $q^{-1}=r^{-1}$. Then we would have:

$$\|f*g\|_q \leq \|f\|_1\|g\|_q$$

To see that this happens, we use theorem 5.3.2. In particular, let $K(x,y)=f(x-y)$. Then if $\|g\|_q$ exists, $\|K(x,\cdot)\|_1 = \|f\|_1$ and $\|K(\cdot,y)\| = \|f\|_1$, giving us the result.

Now, assume $1<p<r$, $1<q<r$. First, we manipulate it to be able to use Hölder's inequality:

$$|f(y)g(x-y)| = (|f(y)|^p|g(x-y)|q)^{1/r} = |f(y)|^{1-p/r}|g(x-y)|^{1-q/r}$$

Next, notice that:

$$\frac{1}{r}+\frac{r-q}{rq}+\frac{r-p}{rp} = \frac{1}{r}\left(1+\frac{r}{q}-1+\frac{r}{p}-1\right) = 1$$

So, we can apply general Hölder's inequality since $r-p, r-q \neq 0$ which gives us:

$$\begin{aligned}|(f*g)(x)| &\leq \left(\int_{\mathbb{R}^n} |f(y)^p||g(x-y)|^q\right)^{1/r} \cdot \|f\|_p^{(r-p)/r}\|g\|_q^{(r-q)/r} \\ \Rightarrow\ |(f*g)(x)|^r &\leq \left(\int_{\mathbb{R}^n} |f(y)^p||g(x-y)|^q\right) \cdot \|f\|_p^{(r-p}\|g\|_q^{r-q}\end{aligned}$$

And now, we can integrate the left hand side with respect to x and use Tonnelli's theorem:

$$\Rightarrow\ \|(f*g)\|_r^r \leq \|f\|_p^p\|f\|_p^{r-p}\|g\|_q^q\|g\|_q^{r-q}$$

completing the proof

If we focus on the special case of $p = q = r = 1$, then $f * g \in L^1(\mathbb{R}^n)$. Then proposition 6.2.1 turns $L^1(\mathbb{R}^n)$ into a "commutative associative algebra". Note that this algebra has no multiplicative identity (i.e. we are working over *rngs*). It does have a notion of an identity in a limit, which we will shortly explore. We first show that convoluting two functions, one nicer than the other, usually keeps the properties of the nicer function:

Lemma 6.2.1: Convolution Preserves C_c^k

Let $f \in L^1$ and $g \in C_c^k$ (more generally, $\partial^\alpha g$ is bounded for $|\alpha| \leq k$). Then

$$f * g \in C^k$$

and

$$\partial^\alpha (f * g) = (f * \partial^\alpha g)$$

Proof **:**
In Folland textbook, he says it's clear from theorem 2.3.4.

This is a matter of computation. We'll do the case of $k = 1$, since the rest is a matter of induction. First, write what you know about $f * g$:

$$\begin{aligned} f * g(x) &= \int_{\mathbb{R}^n} f(x-y)g(y)dy \\ &= \int_{\mathbb{R}^n} f(y)g(x-y)dy \end{aligned}$$

So consider

$$\lim_{h \to 0} \frac{f * g(x+h) - f * g(x)}{h} = \lim_{h \to 0} \int_{\mathbb{R}^n} f(y) \left(\frac{g(x+h-y) - g(x-y)}{h} \right) dy$$

Notice that:

$$\left| \frac{g(x+h-y) - g(x-y)}{h} \right| \leq \frac{|g'(c)| \cdot h}{h}$$

meaning we have a dominating function, and so we can apply dominated convergence theorem to move the limit in!

$$\begin{aligned} \lim_{h \to 0} \frac{f * g(x+h) - f * g(x)}{h} &= \lim_{h \to 0} \int_{\mathbb{R}^n} f(y) \left(\frac{g(x+h-y) - g(x-y)}{h} \right) dy \\ &= \int_{\mathbb{R}^n} f(y) \lim_{h \to 0} \left(\frac{g(x+h-y) - g(x-y)}{h} \right) dy \end{aligned}$$

and since $g \in C_c^1$ (in particular, g is differnetiable), this limit exists, and since g has comapct support, the integral is still well-defined and is C^1 (see the remarks we made at the beginning).

From here we can apply the induction, and the proof is done.

Proposition 6.2.2: Convolution In Schwartz Space

Let $f, g \in \mathcal{S}$. Then $f * g \in \mathcal{S}$

***Proof* :**
First, $f * g \in C^\infty$ by lemma 6.2.1. Next, we establish the following for the coming chain of inequalities:

$$1 + |x| \overset{\triangle\text{ineq.}}{\leq} 1 + |x - y| + |y| \leq (1 + |x - y|)(1 + |y|)$$

and so:

$$\begin{aligned}(1 + |x|)^N |\partial^\alpha (f * g)(x)| &\leq \int (1 + |x - y|)^N |\partial^\alpha f(x - y)|(1 + |y|)^N |g(y)| dy \\ &= \|f\|_{(N,\alpha)} \|g\|_{(N+n+1,\alpha)} \int (1 + |y|)^{-n-1} dy\end{aligned}$$

which is certainly finite

With this information, we are going to define a notion of "multiplicative idenity"

Definition 6.2.2: Good Kernels

Let $\{\phi_t\}_{t>0}$ be a set of functions of the form $\phi_t : \mathbb{R}^n \to \mathbb{C}$. Then they are called a *good kernel* if $t \in (0, \epsilon_0)$ and:

1. $\sup_{t>0} \|\phi_t\|_1 < \infty$
2. $\int_{\mathbb{R}^n} \phi_t(x) dx = 1$
3. $\lim_{t\to 0} \int_{|x|>\delta} |\phi_t(x)| dx = 0$ for all $\delta > 0$

Theorem 6.2.2: Good Kernels Like Mult. Identity

Let $f \in L^p$ for $1 \leq p < \infty$ (note it is $<$, not $\leq$), then:

$$f * \phi_t \to f$$

in L^p as $t \to \infty$. Furthermore, if f is continuous at some x_0, then:

$$f * \phi_t(x_0) \to f(x_0)$$

in L^p

***Proof* :**
It is equivalent to show $f * \phi_t - f \to 0$, so for all x:

$$
\begin{aligned}
f * \phi_t(x) - f(x) &= \int_{\mathbb{R}^n} f(x-y)\phi_t(y)dy - f(x) \cdot 1 \\
&= \int_{\mathbb{R}^n} f(x-y)\phi_t(y)dy - f(x) \cdot \int_{\mathbb{R}^n} \phi_t(y)dy \\
&= \int_{\mathbb{R}^n} f(x-y)\phi_t(y)dy - \int_{\mathbb{R}^n} f(x)\phi_t(y)dy \\
&= \int_{\mathbb{R}^n} f(x-y)\phi_t(y) - f(x)\phi_t(y)dy \\
&= \int_{\mathbb{R}^n} [f(x-y) - f(x)]\phi_t(y)dy \\
&= \int_{|y|\leq\delta} [f(x-y) - f(x)]\phi_t(y)dy + \int_{|y|>\delta} [f(x-y) - f(x)]\phi_t(y)dy
\end{aligned}
$$

and so, taking the p-norm we get:

$$
\|f * \phi_t - f\|_p \leq \int_{|y|\leq\delta} \|f(\cdot - y) - f(\cdot)\|_p |\phi_t(y)|dy + \int_{|y|>\delta} 2\|f\|_p |\phi_t(y)|dy
$$

at this point, we are almost done. If we let $c = \sup_{t>0} \int_{\mathbb{R}^n} |\phi_t(y)|dy$, then we get:

$$
\limsup_{t\to 0} \|f * \phi_t - f\|_p \leq c \limsup_{\delta\to 0,\ |y|\leq\delta} \|f(\cdot - y) - f(\cdot)\|_p = 0
$$

where the last inequality comes from the srhinking delta.

It was left as an exercise to do the point-convergencee in the continuous csae.

Note that if f is uniformally continuous and bouned, then $f * \phi_t \to f$ uniformally, and if $\int \phi = a$, then $f * \phi_t \to af$.

Example 6.2: Good Kernels
Let $X(x) \in C_c^\infty(\mathbb{R}^n)$ where $\int X = 1$. Then:

$$
\phi_t(x) = t^{-n}X(x/t)
$$

is a good kernel (ex. let $X(x) = \frac{1}{\sqrt{\pi}}e^{-x^2}$). This is just a matter of computation. Note that $\phi_t = 0$ if $\frac{x}{t} \geq c_0$ for some c_0, and so $t \leq \frac{x}{c_0}$. It [somehow] follows that C_c^∞ is *dense* in L^p $(1 \leq p < \infty)$ (note the $<$ and not $\leq$)

For another Another example of a family of good kernel, take:

$$
K_\delta(x) = \delta^{-1/2}e^{\frac{-\pi x^2}{\delta}}
$$

These can be visualized here:

`https://www.desmos.com/calculator/e8xc37r3ck`

Notice that since the function $X \in C_c^\infty(\mathbb{R})$ is smooth on compact supporet then by lemma 6.2.1, $f * X$ is also smooth, and will remain smooth for every ϕ_n. Since $f * \phi_n \to f$, we see that a sequence of smooth functions converges to an L^p function, meaning smooth functions are *dense* in L^p!

I found this interesting stack-exchange post on the importance of good Kernels:

`https://math.stackexchange.com/questions/3294744/good-kernels-that-exist-in-the-real-world`

Using the fact that convolution preserves the "nicer" structure, we strengthen's Urysohn's lemma (or, if you know some differential geometry, generlizes partitions of unity)

Theorem 6.2.3: Smooth Urysohn's Lemma

Let $K \subseteq \mathbb{R}^n$ be compact and $U \subseteq \mathbb{R}^n$ be open and contain K ($U \supseteq K$). Then there exists a $f \in C_c^\infty$ such that $0 \leq f \leq 1$ and $f \equiv 1$, K and $\mathrm{supp}(K) \subseteq f$.

Proof **:**
Since K is compact, then $\delta = \rho(K, U^c)$ (the shortest distance from K and U^c) is non-zero. We are going to do a similar trick to that of lemma 3.4.1: let $V = \left\{x \mid \rho(x,K) < \frac{\delta}{3}\right\}$. Choose some ϕC_c^∞ such that $\int \phi = 1$ and $\phi(x) = 0$ for $|x| \geq \delta/3$ (as an example, $\{\psi_t\}$ to be a good kernel and take $\phi = \left(\int \psi\right)^{-1} \psi_{\delta/3}$).

Now, let $f = \chi_V * \phi$. Then as we've shown before, $f \in C_c^\infty$, and it is easy to see that $0 \leq f \leq 1$, $f \equiv 1$ on K, and $\mathrm{supp}(f) \subseteq \{x \mid \rho(x,K) \leq 2\delta/3\} \subseteq U$

6.3 Fourier Transform

The Fourier Transform has its origins in studying how the heat equation:

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}$$

evovles given some starting temperature on $[a,b]$, that is, $u(0,\cdot) : [a,b] \to \mathbb{R}$ gives a temperature at each point. Normalizing so that $[a,b]$ is $[0,2\pi]$, then we can find a solution to u if $u(0,\cdot) = \sin$ or $\cos$ (in particular $\alpha \sin e^{-\alpha t}$ for some constant α). having found the solution for sin and cos we can also find the solution to $\sin(kx)$, which is simply $k^2 \sin(kx) e^{-kx}$ (similalry for cos). We might now hope that this set of functions is dense and still preserves the fact that they form a solution when taking the limit. More generally,for those who are not too intersted in physics, will this allow a new way of studying functions and finding properties that allow us to characterize them? In other words, can this bring a whole new dimension in how we percieve functions?(!)

Though it is not obvious, this family of equations is *dense* in the space of all L^2 functions on a compact interval (or even better, they form an orthonormal basis), and since the PDE in question is *linaer*, we can use the Dominated convergence theorem to show that the resulting sequence can mean that $u(0,\cdot)$ can be any L^2-integrable function (or really, simply integralbe since the domain is compact). we will show that these functions form an orthonormal basis in the following way: we first generlize from an interval $[a,b]$ to a periodic function (any function on an interval can be suitably made periodic). Since they are now defined over $\mathbb{R}$ (or equivalently S^1) we will insure that we limit ourselves to L^2 functions so that we still have an inner product; later we will figure out how we can drop this requirement. Next, we will take advantage of the geometry of the complex plane and that

$$e^x = \cos(x) + i\sin(x)$$

and so instead look for the the functions $\left\{e^{kx} \mid k \in \mathbb{Z}\right\}$ being dense (which is simpler than before, since we only have one family of functions to deal with). Let $L^2(\mathbb{T}^n)$ represent all periodic functions which are square integrable. These can be thought to include $C_c^\infty(\mathbb{R}^n)$, since anything that happens on a compact interval can be smoothly extended to be periodic, and we can squish it so that the period is always 1. We chose L^2 since it has the inner product:

$$\langle f, g\rangle = \int f\overline{g}$$

There is another way of seeing the space $\mathbb{T}^n$: Take $\mathbb{R}^n$. Then this group acts on itself by (left) translation. Let $f \in L^2(\mathbb{R}^n, \mathbb{C})$ be a function that is invariant to translation up to a constant multiple of 1 (this represents a periodic functions, since any period of a function can be mapped onto S^1). Mathematically, if f was one such such function, then for any $y \in \mathbb{R}^n$ (resp. $y \in \mathbb{T}^n$), there exists a $\phi(y)$ with $|\phi(y)| = 1$ and:

$$f(y + x) = \phi(y)f(x)$$

If you know some character theory, you can think of these as the *class functions* where instead of being constant on conjugacy classes, they are constant on S^1. If ϕ was also translation invariant up to absolute value of 1, then:

$$f(x) = f(x + 0) = \phi(x)f(0)$$

and so f is completely determined by what $f(0)$ is. Furthermore:

$$\phi(x)\phi(y)f(0) = \phi(x)f(y) = f(x + y) = \phi(x + y)f(0)$$

and so $\phi(x + y) = \phi(x)\phi(y)$ (given $f(0) \neq 0$). Thus, to find all f's that are translation invariant in the way described above, it suffices to classify all functions with the above homomorhpism property. Imposing that ϕ be measurable, we can find exactly that:

Theorem 6.3.1: Classifying Translation-Invariant Functions

Let ϕ be a measurable function on $\mathbb{R}^n$ (resp. $\mathbb{T}^n$) such that $\phi(x+y) = \phi(x)\phi(y)$ and $|\phi(x)| = 1$. Then there exists a $\xi \in \mathbb{R}^n$ (resp. $\xi \in \mathbb{T}^n$) such that $\phi(x) = e^{2\pi i\xi\cdot x}$ where $\xi \cdot x$ is the dot product.

***Proof* :**
We prove it for $\mathbb{R}$, the result naturally generalizing for $\mathbb{R}^n$ and automatically working for $\mathbb{T}^n$. Let $a \in \mathbb{R}$ such that $\int_0^a \phi(t)dt \neq 0$. Such an a exists, or else by the Lebesgue Differentiation Theorem $\phi = 0$ a.e. Setting $A = \left(\int_0^a \phi(t)dt\right)^{-1}$, we get:

$$\phi(x) = \phi(x)AA^{-1} = A\int_0^a \phi(x)\phi(t)dx = A\int_0^a \phi(x + t)dt = A\int_x^{x+a} \phi(t)dt$$

Thus, ϕ must be continuous, being the integral of a locally integrable function. But that makes the function inside the integral continuous, and so ϕ is in fact C^1 (in fact C^∞, but we will not need this fact). Now, notice that:

$$\phi'(x) = A[\phi(x+a) - \phi(x)] = B\phi(x)$$

where $B = A(\phi(a) - 1)$. This is not a simple ODE, which can be sovled to show we get an exponential (and is unique by the uniquenss of a solution to an ODE). If you don't know about ODE's, then notice that

$$\frac{d}{dx}(e^{-Bx}\phi(x)) = 0$$

so $e^{-Bx}\phi(x)$ is constant. Since $\phi(0) = 1$, we have $\phi(x) = e^{Bx}$, and since $|\phi(x)| = 1$, B must be purely imaginary meaning $B = 2\pi i\xi$ for some $\xi \in \mathbb{R}$, completing the proof for $\mathbb{R}$. For $\mathbb{T}$, we would consider ϕ to be periodic with period 1 and $e^{2\pi i\xi} = 1$ if and only if $\xi \in \mathbb{Z}$

For the n-dimensional case, let $e_1, e_2, ..., e_n$ be the standard basis for $\mathbb{R}^n$ and define the coordiante functions $\phi_i(t) = \phi(te_i)$. Naturally, each ϕ_i satisifes the property $\phi_i(t+s) = \phi_i(t)\phi_i(s)$ on $\mathbb{R}$, and so $\phi_i(t) = e^{2\pi i\xi_i t}$, and so:

$$\phi(x) = \phi\left(\sum_1^n x_i e_i\right) = \prod_1^n \phi_i(x_i) = e^{2\pi i\xi\cdot x}$$

completing the proof.

Hence, a function satisfying this translation invariance is of the form $f(x) = Ke^{2\pi i\xi\cdot x}$ for some constant $K \in \mathbb{R}$. With the classification of these functions, we will want to use these to decompose reasonably arbitrary functions in $\mathbb{R}^n$ and $\mathbb{T}^n$ (i.e. we will show they form a dense set). In the case of L^2, we get a very precise result:

Theorem 6.3.2: Orthonormality Of Translation Functions

Let $E_k(x) = e^{2\pi ik\cdot x}$. Then the set $\{E_k \mid k \in \mathbb{Z}^n\}$ is an orthonormal basis for $L^2(\mathbb{T}^n)$

Proof **:**
For orthonormality, notice that:

$$\int_0^1 e^{2\pi ikt}dt = \begin{cases} 1 & k = 0 \\ 0 & \text{otherwise} \end{cases}$$

using this, notice that if $n \neq m$, then:

$$\begin{aligned}
\langle E_a, E_b\rangle &= \int_0^1 e^{2\pi iat}\overline{e^{2\pi ibt}}dt \\
&= \vdots \\
&= \frac{1}{2\pi i(a-b)}(e^{2\pi i(a-b)} - e^0) \\
&= \frac{1}{2\pi i(a-b)}(1-1) \\
&= 0
\end{aligned}$$

and if $n = m$, then we get

$$\int_0^1 e^{2\pi ikt}e^{2\pi i-kt}dt = \int_0^1 dt = 1$$

For density, since $E_xE_y = E_{x+y}$, the set of finte linear combinations of the E_k's form an algebra which clearly seperates points on $\mathbb{T}^n$, and $E_0 = 1$, $\overline{E_k} = E_{-k}$. Since $\mathbb{T}^n$ compact, by the Stone-Weiestrass theorem, this algebra is dense in $C(\mathbb{T}^n)$ in the uniform norm and hence in the L^2 norm, and $C(\mathbb{T}^n)$ is dense in $L^2(\mathbb{T}^n)$, and so $\{E_k\}$ is a orthonormal basis for $L^2(\mathbb{T}^n)$, as we sought to show.

Using this, we can now define the fourier transform:

Definition 6.3.1: Fourier Transform And Fourier Series

Let $f \in L^2(\mathbb{T}^n)$. Then define the *Fourier Transform* $\hat{f} : \mathbb{Z}^n \to \mathbb{C}$ to be:

$$\hat{f}(k) = \langle f, E_k\rangle = \int_{\mathbb{T}^n} f(x)e^{-2\pi ik\cdot x}dx$$

and the series:

$$f = \sum_{k\in\mathbb{Z}^n} \hat{f}(k)E_k$$

is called the *Fourier series* of f.

A couple of remarks are in order:

1. the map $f \mapsto \hat{f}$ is sometimes called the fourier transform
2. Theorem 6.3.2 shows that $L^2(\mathbb{T}^n)$ onto $\ell^2(\mathbb{Z}^n)$, and that $\|\hat{f}\|_2 = \|f\|_2$; this is known as *Parsavel's Identity*. Furthermore
$$\hat{f}(0) = \int_{\mathbb{T}^n} f(x)dx$$
3. here is another way of seeing $f = \sum_{k\in\mathbb{Z}^n} \hat{f}(k)E_k$ that is taking the intuition 3blue1brown gave in the video `https://www.youtube.com/watch?v=r6sGWTCMz2k`. Recall that $f(t) = e^{k2\pi it}$ is the function that rotates k times counter clockwise when t ranges from 0 to 1 (and clockwise if k is negative). Integrating this function from 0 to 1 always gives 0. If $k = 0$, then we get $f(t) = 1$. Integrating this from 0 to 1 also gives 1. Recall the average of a function would be found when divding over the measure of area we are integrating over, but since $m([0, 1]) = 1$, we already have our average![3].

   Now, let's say that instead of trying to take the formal appraoch we've done, we think a bit loosely for a moent and say that we are trying to find the approximation:
$$f = \cdots + c_{-1}e^{-1\cdot 2\pi it} + c_0e^{0\cdot 2\pi it} + c_1e^{1\cdot 2\pi it} + c_2e^{2\cdot 2\pi it} + \cdots$$
   with the intuition on why we would come up with this being that we want to break the periodic function f (or equivalently the compactly supported function f) down into the above series.

[3]Interestingly, it seems there is some probably going on here

How do we find the coefficients? First, by the DCT:

$$\begin{aligned}\int f(t)dt &= \int \left(\cdots + c_{-1}e^{-1\cdot 2\pi it} + c_0 e^{0\cdot 2\pi it} + c_1 e^{1\cdot 2\pi it} + c_2 e^{2\cdot 2\pi it}\right) dt \\ &= \cdots + \int c_{-1}e^{-1\cdot 2\pi it}dt + \int c_0 e^{0\cdot 2\pi it}dt + \int c_1 e^{1\cdot 2\pi it}dt + \int c_2 e^{2\cdot 2\pi it}dt + \cdots\end{aligned}$$

Now, notice that all terms except the one containing c_0 are disapear, since the rotation of by the exponential will always guarantee to make the integrals average to zero! Thus, c_0 can be thought of as the average mass point of f. To find the other terms, we simpy have to mutiply f at the begining by $e^{k2\pi it}$

$$\int f(t)e^{-k2\pi it}dt$$

and then we have shifted over all of our terms except the kth term, which we can now find the integral of! In fact, notice we have just recovered:

$$\hat{f}(k) = \int f(t)e^{-k2\pi it}dt$$

giving us a geometric intuition for this formula!

4. By the definition of an orthonormal basis, the Fourier series of f converges to f in the L^2-norm. In the next two sections, we will address point-wise convergence.

5. the definition of the Fourier transform $\hat{f}(k)$ makes sense if $f \in L^1(\mathbb{T}^n)$ and $|\hat{f}(k)| \leq \|f\|_1$, showing us that $\hat{f}$ must be essentially bounded, so the Fourier series would then extend to a norm-decreasing map from $L^1(\mathbb{T}^n)$ to $\ell^\infty(\mathbb{Z}^n)$. We will show how to recover f from $\hat{f}$ in this situation later.

The last point shows us that there is a sort of duality between the speed of decay of a function f and its fourier transform $\hat{f}$. The natur of this duality is made explicit in the following theorem

Theorem 6.3.3: Hausdorff-Young Inequality

Let $1 \leq p \leq 2$ and q be the conjugate exponent of p. If $f \in L^p(\mathbb{T}^n)$, then $\hat{f} \in \ell^q(\mathbb{Z}^n)$ and $\|\hat{f}\|_q \leq \|f\|_p$

***Proof* :**

Since $\|\hat{f}\|_\infty \leq \|f\|_1$ and $\|\hat{f}\|_2 = \|f\|_2$ for $f \in L^1$ and $f \in L^2$, this follows from the Riez-Thorin interpolation theorem (Radon Measure chapter in Folland)

Thus, we see that the Fourier transform is a useful decomposition of periodic functions. This decomposition and way of thinking of a function in isolated "periodic" parts is in fact very powerful (similarly to how power series are very powerful and central to complex analysis among other fields). It is thus fruitful to explore the quesiton of a fourier transform of non-periodic functions. If we simply formally copy the result from theorem 6.3.2, then we get:

$$f(x) = \int_{\mathbb{R}^n} \hat{f}(\xi)e^{2\pi i\xi\cdot x}dx \quad \text{where} \quad \hat{f}(\xi) = \int_{\mathbb{R}^n} f(x)e^{-2\pi i\xi\cdot x}dx$$

so that:

$$f(x) = \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} f(y) e^{-2\pi i \xi \cdot y} dy \right) e^{2\pi i \xi \cdot x} dx$$

This function is defined if we put in some care to make sure our bounds are well-defined: if $f \in L^2$, then it is likely that $\hat{f}(\xi)$ diverges. On the other hand, if $f \in L^1$, $\hat{f}$ will certainly converge. Thus, we can define the Fourier transform for $f \in L^1$ to be:

$$(\mathcal{F}f)(\xi) = \mathcal{F}f(\xi) = \hat{f}(\xi) = \int_{\mathbb{R}^n} f(x) e^{-2\pi i \xi \cdot x} dx$$

Notice that $\|\widehat{f}\|_u \leq \|f\|_1$, and that $\hat{f}$ is continuous, and so:

$$\mathcal{F} : L^1(\mathbb{R}^n) \to \mathrm{BC}(\mathbb{R}^n)$$

Proposition 6.3.1: Properties Of Fourier Transform

Let $f, g \in L^1(\mathbb{R}^n)$. Then:

1. $\widehat{(\tau_y f)}(\xi) = e^{-2\pi i \xi \cdot y} \hat{f}(\xi)$ and $\tau_n(\hat{f}) = \hat{h}$ where $h(x) = e^{2\pi i n \cdot x} f(x)$
2. If T is an invertible linear transformation on $\mathbb{R}^n$ and $S = (T^*)^{-1}$ is its inverse transpose, Then $\widehat{(f \circ T)} = |\det(T)|^{-1} \hat{f} \circ S$. In particular, if T is a rotation, then $\widehat{(f \circ T)} = \hat{f} \circ T$. Furthermore, if $T(x) = t^{-1}x$ (for $t > 0$), then $\widehat{(f \circ T)}(\xi) = t^n \hat{f}(t\xi)$ so that $\widehat{(f_t)}(\xi) = \hat{f}(t\xi)$ in the notation of good kernels
3. Convolution becomes multiplication after the Fourier Transform:
$$\widehat{(f * g)} = \hat{f}\hat{g}$$
4. If $x^\alpha f \in L^1$ for $|\alpha| \leq k$ then $\hat{f} \in C^k$ and $\partial^\alpha \hat{f} = \widehat{[(-2\pi i x)^\alpha f]}$
5. if $f \in C^k$, $\partial^\alpha f \in L^1$ for $|\alpha| \leq k$, and $\partial^\alpha f \in C_0$, for $|\alpha| \leq k - 1$, then
$$\widehat{(\partial^\alpha f)}(\xi) = (2\pi i \xi)^\alpha \hat{f}(\xi)$$
6. (Riemann-Lebesuge Lemma)
$$\mathcal{F}(L^1(\mathbb{R}^n)) \subseteq C_0(\mathbb{R}^n)$$

***Proof* :**

1. computing, we get:
$$\widehat{(\tau_y f)}(\xi) = \int f(x-y) e^{-2\pi i \xi \cdot x} dx = \int f(x) e^{-2\pi i \xi \cdot (x+y)} dx = e^{-2\pi i \xi \cdot y} \hat{f}(\xi)$$

2. Apply theorem 2.6.4:

$$\begin{aligned}\widehat{(f \circ T)}(\xi) &= \int f(Tx)e^{2-\pi i\xi\cdot x}dx \\ &= |\det T|^{-1}\int f(x)e^{-2\pi i\xi\cdot T^{-1}x}dx \\ &= |\det T|^{-1}\int f(x)e^{-2\pi i S\xi\cdot x}dx \\ &= |\det T|^{-1}\hat{f}(S\xi)\end{aligned}$$

3. By Fubini's Theorem:

$$\begin{aligned}\widehat{(f * g)}(\xi) &= \int\int f(x-y)g(y)e^{-2\pi i\xi\cdot x}dydx \\ &= \int\int f(x-y)e^{-2\pi i\xi\cdot(x-y)}g(y)e^{-2\pi i\xi\cdot y}dxdy \\ &= \hat{f}(\xi)\int g(y)e^{-2\pi i\xi\cdot y}dy \\ &= \hat{f}(\xi)\hat{g}(\xi)\end{aligned}$$

4. By theorem ref:HERE and induciton on $|\alpha|$,

$$\partial^\alpha\hat{f}(\xi) = \partial_\xi^\alpha\int f(x)e^{-2\pi i\xi\cdot x}dx = \int f(x)(-2\pi ix)^\alpha e^{-2\pi i\xi\cdot x}dx$$

5. Assuming first that $n = |\alpha| = 1$, since $f \in C_0$, we can integerate by parts to get:

$$\int f'(x)_e^{-2\pi i\xi\cdot x}dx = f(x)e^{-2\pi i\xi\cdot x}\Big|_{-\infty}^{\infty} - \int f(x)(-2\pi i\xi)^{-2\pi i\xi\cdot x}dx = 2\pi i\xi\hat{f}(\xi)$$

   Then the argument for $n > 1$, $|\alpha| = 1$ is the same, to comptue $(\widehat{\partial_j f})$, integrate by parts the jth variable, and the genreal case follows by induciton on $|\alpha|$.

6. By the previous part, if $f \in C^1 \cap C_c$, then $|\xi|\hat{f}(\xi)$ is bounded and hence $\hat{f} \in C_0$. But the set of all such f's is dense in L^1 by proposition ref:HERE (8.17 in folland), and $\hat{f}_n \to \hat{f}$ uniformally whenver $f_n \to f$ in L^1. Since C_0 is closed inteh uniform norm, the result follows, completing the proof.

Notice that (4) and (5) give an interesting relation between the smoothness of f and the decay of $\hat{f}$: the smoother f is, the fast $\hat{f}$ decays (and vice-versa).

Corollary 6.3.1: Fourier Transform On Schwartz Space

The map $\mathcal{F}$ maps the Schwartz space $\mathcal{S}$ into itself

Proof **:**
folland p.250

Proposition 6.3.2: Important Fourier Transform

Let $f(x) = e^{-\pi a|x|^2}$ where $a > 0$. Then $\hat{f}(\xi) = a^{-n/2}e^{-\pi|\xi|^2/a}$

Proof **:**
folland p. 251

Chapter 7

Distribution Theory

Distributions a generlization of functions that allow us to give more general solutiosn than classical solutions, similar to how we found a weak solution last week. locally integrable functions will be distributions (in the appropriate interpretation), but we'll see that there are general solution's that don't even fall under the notion of "function".

(A word how this is a further exploration of "dual theory")

Another use we will soon see

7.1 Intuition

For now, let $f \in L^p$. Then as we know, $(L^p)^* \cong L^q$ is an isometry where $1 \leq p < \infty$. In particular, we have that

$$g \mapsto \int fg$$

is mapped to by f. Using the Lebesgue differentiation theorem, by choosing $\phi_k = m(B_r)^{-1}\chi_{B_r}$, where $B - r$ is a ball of ardius r around x, will allow us to recevoer the value of $f(x)$ for a.e. x, as $\lim_{r \to 0} \int f\phi_r$. In this way, we can think of L^p as being embedded into $(L^q)^*$ (since we really only care what happens a.e.).

Now, let us modify the abvoe by letting f be merely locally integrable (in L^1_{loc}), but $\phi \in C_c^\infty$. Then it is clear that $\phi \mapsto \int f\phi$ is still a well-defined functional on C_c^∞ , and th e po intwise va lues o f f can be r ecoved a.e. from it, by extending theroem ref: HERE (8.15 in Folland). Unlike in the case for L^p spaces, there are in fact *many* linear functionals on C_c^∞ that are *not* of the form $\phi \mapsto \int f\phi$. These, subject ot some mild ocntinuity conditions we will soon specify, will be our "generlized functions".

7.2 Definitions

We will first rapid-fire some important definitions to be up to speed with terminology. Recall that $C_c^\infty(E)$ where $E \subseteq \mathbb{R}^n$ is the set of all smooth functiosn whose support is compact and contained in E. If $U \subseteq \mathbb{R}^n$ is open, $C_c^\infty(U)$ is the union of space $C_c^\infty(K)$ where K ranges oer all compact subsets of U. Each $C_c^\infty(K)$ is a Fréchet space with the topolgoy defined by the norms

$$\phi \mapsto \|\partial^\alpha\|_u \qquad (\alpha \in \{0,1,2,...\}^n)$$

Thus ,the sequence $\{\phi_i\}$ converges if and only if $\partial_i^\phi \to \partial^\alpha \phi$ uniformly for all α. The completness of $C_c^\infty(K)$ is shown through exercise 9 in folland in section 5.1. With these, we take the following definitions:

1. A sequence $\{\phi_i\}$ in $C_c^\infty(U)$ converges in C_c^∞ to ϕ is $\{\phi_i\} \subseteq C_c^\infty(K)$ for some compact set $K \subseteq U$ and $\phi_i \to \phi$ in the topology of $C_c^\infty(K)$, that is, $\partial^\alpha \phi_i \to \partial^\alpha \phi$ uniformally for all α
2. If X is a locally convex topologicla vector space and $T : C_c^\infty(U) \to X$ is a linaer map, T is *continuous* if $T|_{C_c^\infty(K)}$ is continous for each compact $K \subseteq U$, that is, if $T(\phi_i) \to T(\phi)$ whenever $\phi_i \to \phi$ in $C_c^\infty(K)$ and $K \subseteq U$ is compact (i.e. point-wise convergence)
3. A linear map $T : C_c^\infty(U) \to C_c^\infty(U')$ is *continuous* if for each compact $K \subseteq U$, there is a compacty $K' \subseteq U'$ such that $T(C_c^\infty(K)) \subseteq C_c^\infty(K')$ and T is continuous from $C_c^\infty(K)$ to $C_c^\infty(K')$

With these, we will define a distribution:

Definition 7.2.1: Distribution

A *distribution* on U is a continous linear functional on $C_c^\infty(U)$. The space of all distributiosn on U is denoted by $\mathcal{D}'(U)$, and we set $\mathcal{D}' = \mathcal{D}'(\mathbb{R}^n)$. We impose the weak* topology on $\mathcal{D}'(U)$, that is, the topology of pointwise converge on $C_c^\infty(U)$.

A couple of remarks are in order: first, the $\mathcal{D}'$ notation comes from the fact that Schwart'z used the notation of $\mathcal{D}$ for C_c^∞, and so $\mathcal{D}'$ was the functionals on C_c^∞. Next, there is a locally convex topology on C_c^∞ with respect to which sequential convergence in C_c^∞ is given by the convergence defined earlier in C_c^∞, and the continuity of a linear map $T : C_c^\infty \to X$ and $T : C_c^\infty \to C_c^\infty$ is given by the other two points. However, to define it is relatively complicated and doesn't contribute much to the current exposition of distributions , and so will be ommited. Furthermore, when we talk about convergence of distributions $F_i \to F$, we will mean point-wise convergence.

Before giving some examples of distributions, we will remind the reader how of how to check for continuity in a TVS (in particular a Fréchet space)

Proposition 7.2.1: Distribution Continuity Check

Let $u : C_c^\infty(U) \to \mathbb{R}$ (or $\mathbb{C}$) be linear. Then $u \in \mathcal{D}'(U)$ if and only if for every compact set $K \subseteq U$, there exists a $C_K > 0$ and $N_K \in \mathbb{N}$ such that

$$\|\langle u, \phi\rangle\| \le C_k \sum_{\|\alpha\| \le N_K} \|\partial^\alpha \phi\|_L^\infty$$

Proof **:**
Given the above equation, then it is easy to check (as we've done in chapter 4) that if $\phi_i \to \phi$, then $\langle u, \phi_i\rangle \to \langle u, \phi\rangle$.

The way I like to think about the continuity check is as follows: we will soon see that there are many Distribution spaces, for example the Schwartz space $\mathcal{S}$ will have a parallel $\mathcal{S}'$. Then what the continuity condition tells us is that if $f_j \to f$ in $\mathcal{S}$ for some sequence in $\mathcal{S}$, then $f_j \to f$ in $\mathcal{S}'$ for some sequence in $\mathcal{S}'$. Hence, we can transalte convergence information from $\mathcal{S}$ to $\mathcal{S}'$.

Example 7.1: Distributions

1. For every $f \in L^1_{\text{loc}}(U)$, we can define a distribution on U< namely the funcitonal $\phi \to \int f\phi$, and two functions define the same distributiosn precisely when they are equal a.e.
2. Every Radon measure μ on U defines a distributiosn $\phi \mapsto \int \phi d\mu$.
3. If $x_0 \in U$ and α is a multi-index, the map $\phi \mapsto \partial^\alpha(x_0)$ is a distribution that does not arise from a function; it arises from a measure μ prcisely when $\alpha = 0$, in which case μ is the point mass at x_0

As mentioned earlier, we often will associate the a function f with the distribution it has, and even label the distribution by f. For this reason, it might be confusing to write $f(\phi)$ to mean the distribution defined by f taking in ϕ, and so for historical reasons we use the notation $\langle f, \phi \rangle$.

Let's next establish some common notation:

1. First, it is often enough that we will reflect function. Thus, if we put a tilde over a function, we mean:
$$\tilde{\phi}(x) = \phi(-x)$$
2. Second, the point-mass is a very common distribution, which we will denote by δ:
$$\langle \delta, \phi \rangle = \phi(0)$$

As an example of the use of the detla funciton:

Proposition 7.2.2: Convergence To Delta Function

Let $f \in L^1(\mathbb{R}^n)$ and $\int f = a$, and for $t > 0$, let $f_t(x) = t^{-n} f(t^{-1}x)$. Then$f_t \to a\delta$ in $\mathcal{D}'$ as $t \to 0$

***Proof* :**
If $\phi \in C_c^\infty$, by theroem 8.15, we have:
$$\langle f_t, \phi \rangle = \int f_t \phi = f_t * \tilde{\phi}(0) \to a\tilde{\phi}(0) = a\phi(0) = a\langle \delta, \phi \rangle$$
completing the proof.

For equality of distributions, note that it makes to say that $F = G$ on an open set V if and only if $\langle F, \phi \rangle = \langle G, \phi \rangle$ for all $\phi \in C_c^\infty(V)$. If F and G are continuous functions, this will imply theya re pointwise equal, and if F and G are only locally integrable,this means they are equal a.e. on V.

Since a function in $C_c^\infty(V_1 \cup V_2)$ need not be supported in either V_1 or V_2, it is not immediately obvious that if $F = G$ on V_1 and on V_2 then $F = G$, then $F = G$ on $V_1 \cup V_2$. However, this is the case:

Proposition 7.2.3: Equality On Support

Let $\{V_\alpha\}$ be a collectio nof open subsets of U and let $V = \bigcup_\alpha V_\alpha$. If $F, G \in \mathcal{D}'(U)$ and $F = G$ on each V_α, then $F = G$ on V

***Proof* :**
Folland p. 284

As a consequence of proposition 7.2.3, if $F \in \mathcal{D}'(U)$, there is a maximal open subset U on which $F = 0$, namely the union of all open subset on which $F = 0$. IT's compliment in U is called the *support* of F. If it's compliment is compact, it is called the *compact support*.

One of amazing facts about distributions is that even though they are not "functions", we will be able to do many functional operations like diferentation, multiplication by a "constant", convolution, and translation!! TO achieve this, we need some theoretical build-up. Suppose U and V are open sets in $\mathbb{R}^n$ and T is a liner map from some subsapce V of $L^1_{\text{loc}}(U)$ into $L^1_{\text{loc}}(V)$. Suppose that there is antoher linaer map $T' : C_c^\infty(V) \to C_c^\infty(U)$ such that

$$\int (Tf)\phi = \int f(T'\phi) \qquad (f \in V,\ \phi \in C_c^\infty(V))$$

suppose that T' is continuous in the sense defined earlier. Then T can be extended to a map from $\mathcal{D}'(U)$ to $\mathcal{D}'(V)$, still dentoed by T, by:

$$\langle TF, \phi\rangle = \langle F, T'\phi\rangle \qquad (F \in \mathcal{D}'(U),\ \phi \in C_c^\infty(V))$$

The continuity of T' guarantees that the original map T, as well as the extension to distributions, is continuous with respct to the weak* topology on distributions, namely if $F_\alpha \to F \in \mathcal{D}'(U)$, then $Tf_1a \to TF$ in $\mathcal{D}'(V)$. With this, we can define the following on distributions

1. (Differentiation) Let $Tf = \partial^\alpha f$, defined on $C^{|\alpha|}(U)$. If$\phi \in C_c^\infty(U)$, integration by parts gives $\int(\partial^\alpha f)\phi = (-1)^{|\alpha|}\int f(\partial^\alpha \phi)$; there are no boundary terms since ϕ has compact suppor.t. Henc $T' = (-1)^{|\alpha|}T|_{C_c^\infty(U)}$, and we can define the *derivative* $\partial^\alpha F \in \mathcal{D}'(U)$ of any $F \in \mathcal{D}'(U)$ by:

$$\langle \partial^\alpha F, \phi\rangle := (-1)^{|\alpha|}\langle F, \partial^\alpha, \phi\rangle$$

   By this procedure, we can define the derivatives of arbirary locally integrabl funtions, even when they are not differentiable in the classical sense! We will give the derivative of certain locally integrable functions soon!!

2. (multiplication by a smooth function, a "scalar") Let $\psi \in C^\infty(U)$ and define $Tf = \psi f$. Then evidently, $T' = T|_{C_c^\infty(U)}$, and so we can define $\psi F \in \mathcal{D}'(U)$ for any $F \in \mathcal{D}'(U)$ to be:

$$\langle \psi F, \phi\rangle := \langle F, \psi\phi\rangle$$

3. (Translation) Let $y \in \mathbb{R}^n$ and let $V = U + y = \{x + y \mid x \in U\}$. Take $T = \tau_y$ (recall that $\tau_y f(x) = f(x - y)$). Via u-substituion, we get that $\int f(x-y)\phi(x)dx = \int f(x)\phi(x+y)dx$, and so $T' = \tau_{-y}|_{C_c^\infty(U)}$. For a distribution $F \in \mathcal{D}'(U)$, we define:
$$\langle \tau_y F, \phi \rangle := \langle F, \tau_{-y}\phi \rangle$$
4. (convolution with linear maps) Let S be an invertible linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$. Letting $V = S^{-1}(U)$ and $Tf = f \circ S$, Then $T'\phi = |\det(S)|^{-1}\phi \circ S^{-1}$ (by change of variables, see theorem ref:HERE). Thus, for $F \in \mathcal{D}'(U)$, define:
$$\langle F \circ S, \phi \rangle := |\det S|^{-1} \langle F, \phi \circ S^{-1} \rangle$$
5. (convolution I) Let $\psi \in C_c^\infty$, and let
$$V = \{x \mid x - y \in U \text{ for } y \in \operatorname{supp}(\psi)\ \}$$
Note that V is open but may be empty. If $f \in L^1_{\text{loc}}(U)$, we get that the integral:
$$f * \psi(x) = \int f(x-y)\psi(y)dy = \int f(y)\psi(x-y)dy = \int f(\tau_x \tilde{\psi})$$
is well-defined for all $x \in V$. Since we transfered over from f to ψ, we can define for $F \in \mathcal{D}'(U)$:
$$F * \psi(x) = \langle F, \tau_x \tilde{\psi} \rangle$$
Since $\tau_x \tilde{\psi} \to \tau_{x_0}\tilde{\psi}$ as $x \to x_0$, $F * \psi$ is a continuous function (in fact, C^∞, as we'll show in a moment). A pertinent example of this is the following:
$$\delta * \psi(x) = \langle \delta, \tau_x \tilde{\psi} \rangle = \tau_x \tilde{\psi(0)} = \psi(x)$$
so δ is the multiplicative identity of convolution!
6. (convolution II) Let $\psi, \tilde{\psi}$ and V be like the last point. If $f \in L^1_{\text{loc}}(U)$ and $\phi \in C_c^\infty(U)$, then
$$\int (f * \psi)\phi = \int \int f(y)\psi(x-y)\phi(y)dydx = \int f(\phi * \tilde{\psi})$$
Thus $Tf = f * \psi$ maps $L^1_{\text{loc}}(U)$ to $L^1_{\text{loc}}(V)$ and $T'\phi = \phi * \tilde{\psi}$. Thus, we get:
$$\langle F * \psi, \phi \rangle = \langle F, \phi * \tilde{\psi} \rangle$$

(theorem on how the two convolutions are actually the same thing)

Next, we will show that smooth compact functions $C_c^\infty(U)$ are dense in $\mathcal{D}'(U)$

Lemma 7.2.1: Build-Up For Density

Suppose that $\phi \in C_c^\infty, \psi C_c^\infty$ and $\int \psi = 1$, and let $\psi_t(x) = t^{-n}\psi(t^{-1}x)$. Then

1. Given any neighborhood U of $\operatorname{supp}(\phi)$, we have $\operatorname{supp}(\phi * \psi_t) \subseteq U$ for t
2. $\phi * \psi_t \to \phi$ in C_c^∞ as $t \to 0$

Proof **:**
Folland p. 287

Proposition 7.2.4: Density Of Smooth Compact Functions In Distributions

Let $U \subseteq \mathbb{R}^n$ be open. Then $C_c^\infty(U)$ is dense in $\mathcal{D}'U)$ in the topology of $\mathcal{D}'(U)$

Proof **:**
Folland p. 287

7.3 Compactly Supported, Tempered, and Dense Distributions

So far, the distributions we covered need not be compactly supported (ex. $f \in L_{\text{loc}^1}$ is certainly not compactly supported). We thus take a moment to look at compactly supported distributions to take advantage of results pertaining to compactly supported functions. Let $\mathcal{E}'(U) \subseteq \mathcal{D}'(U)$ be the set of all compactly supported distributions of $\mathcal{D}'(U)$, and let $\mathcal{E}' = \mathcal{E}'(\mathbb{R}^n)$.

(Some on $\mathcal{E}'$, $\mathcal{D}'$ on with compact support)

Chapter 8

More Measures and Integrals

This last chapter will cover some additional settings in which notions of measure and integration is defined. We will examine the case of measure being defined on locally compact abelian groups and on low dimensional subsets of $\mathbb{R}^n$.

8.1 Topological Groups and Haar Measure

Definition 8.1.1: Topological Group

Let G be a group. Then G is a *topological group* if there is a topology on G such that the binary operation $(x, y) \mapsto xy$ and inversion $x \mapsto x^{-1}$ is continuous. In other words, topological groups are the group objects of **Top**.

Can you come up with an example to see why it is necessary to add this condition? Since we have a binary operation, we have a few more natural operations we may do on sets:

$$xH = \{xh \mid h \in H\} \qquad Ax = \{hx \mid h \in H\}$$
$$H^{-1} = \{h^{-1} \mid h \in H\} \qquad HK = \{hk \mid h \in H,\ k \in K\}$$

If $A \subseteq G$ such that $A = A^{-1}$, we'll say A is *symetric*

Example 8.1: Topological Group

1. For any group G, if G is equipped with the discrete topology it is a topological group, and all homomorphisms are continuous between any other topological group. In this way, the theory of topological groups can be thought as subsuming group theory. Any group with the indiscrete topology (also known as trivial topology) is also a topological group.
2. If $H \leqslant G$ is also a topological space, then H is a topological group with the subspace topology.
3. The real numbers $\mathbb{R}$ and $\mathbb{R}^\times$ under multiplication is a topological group (can you prove continuity of $+$, $\cdot$, and $(-)^{-1}$?). More generally, $\mathbb{R}^n$ is a topological group, or any *toplogical*

vector-space (a vector-space where the action is continuous as well) is a topological group

4. The circle S^1, and the n-torus $(S^1)^n$ are topological groups. show that $\mathbb{R}/\mathbb{Z}$ is both homeomorphic and isomorphic to S^1. If we take S^1 as a subset of $\mathbb{C}$, then we $S^1 = U(1) = \{x \in \mathbb{C} \mid |x| = 1\}$.

5. The general linear group over the reals $\mathrm{GL}_n(\mathbb{R})$ is a topological group. Similarly, the special linear group,$\mathrm{SL}_n(\mathbb{R}) = \{A \in \mathrm{GL}_n \mid \det(A) = 1\}$, is a topological group.

6. Take $O(n) \subseteq \mathrm{GL}(n)$, where $O(n) = \{A \in \mathrm{GL}_n(\mathbb{R}) \mid AA^t = I\}$. This is a topological group known as the *orthogonal group*, since it consists of all orthonormal basis of $\mathbb{R}^n$. Similarly, we can define $U(n) = \left\{A \in \mathrm{GL}_n(\mathbb{R}) \;\middle|\; A\overline{A}^t = I\right\}$ is a topological group known as the *unitary group*. Similarly to $\mathrm{SL}_n(\mathbb{R})$, we have:

$$\mathrm{SO}(n) = \mathrm{SL}_n(\mathbb{R}) \cap O(n) \qquad \mathrm{SU}(n) = \mathrm{SL}_n(\mathbb{R}) \cap U(n)$$

   Note that all these groups are *compact*.

7. (If you know some Differential Geomtery) Any Lie group is naturally a topological group. An example of a topological group that is not a Lie group is $\mathbb{Q}$ with the subspace topology induced by $\mathbb{R}$.

8. The p-addic integers, $\mathbb{Z}_p$ has the induced inverse-limit topology (in this case, the p-addic topology). This group is compact, homeomorphic to the Cantor set, but is totally disconnected (hence cannot be a Lie Group)

9. Similarly to $\mathbb{Z}_p$, giving $\mathbb{Q}$ a different p-addic metric, we can take a different completion to form $\mathbb{Q}_p$, which is also a Lie Group.

10. In commutative algebra, it is common to define a topology on a group in the following way: Let G be a commutative group and take a descending sequence of subgroups[a]:

$$G = G_0 \supseteq G_1 \supseteq \cdots \supseteq G_n \supseteq \cdots$$

   Then let the collection of (G_n) define the *local neighbhorhood* of 0, and $x + G_n$ be an open neighbhorhood of x. Then this generate's a topology, which we'll call the *subgroup topology*. This topology is particularly important when we talk about the *completion* of groups; more on that at the end of this section.

[a] If G were not commutative, we would take a descending sequence of normal subgroups

Proposition 8.1.1: Properties Of Topological Groups

Let G be a topological group.

1. The topology of G is translation invariatn: If $U \subseteq G$ is open, then Ux and xU is open
2. For every open neighbhorhood of e, say for example U, there is a symmetric neighbhorhood V of e such that $V \subseteq U$
3. For every neighbhorhood U of e, ther is a neighbhorhood V of e with $VV \subseteq U$
4. if H is a subgroup of G, so it $\overline{H}$
5. Every open subgroup of G is also closed
6. If K_1, K_2 are compact subset of G, so is K_1K_2

***Proof* :**

1. exercise (use continuity of binary operation)

2-3. exercise (use continuity of both operations at the identity)

4. Let $x, y \in \overline{H}$ so that there is nets $\langle x_\alpha \rangle_{\alpha \in A}$ and $\langle y_\beta \rangle_{\beta \in B}$ in H that converges to x and y. Then $x_\alpha^{-1} \to x^{-1}$ and $x_\alpha y_\beta \to xy$ (with the usual product ordering on $A \times B$), so x^{-1} and xy belong to $\overline{H}$.
5. Since H is an open subgroup, so are al xH, so $G \setminus H = \bigcup_{x \notin H} xH$ is open, hence H is closed
6. K_1K_2 is the iamge of the comapct set $K_1 \times K_2$ under the binary operation.

Definition 8.1.2: Left And Right Translation

Let f be a continuous fnction on G and $y \in G$. Then a *left* translation (resp. right translation) of f is a function L_y (resp. R_y) that maps f to:

$$L_y f(x) = f(y^{-1}x) \qquad R_y f(x) = f(xy)$$

We put y^{-1} instead of y so that translation is a homomorphism: $L_{yz} = L_y L_z$ and $R_{yz} = R_y R_z$.

Definition 8.1.3: Left And Right Uniform Continuity

Let $f : G \to \mathbb{R}$ (or $\mathbb{C}$). Then f is said to be *left* (resp. *right*) *uniformally continuous* if

$$\forall \epsilon > 0, \ \exists V (e \in V), \ \|L_y f - f\| < \epsilon$$

for $y \in V$ (resp. $\|R_y f - f\|_u < \epsilon$).

Proposition 8.1.2: Compact Domain And Left/Right Uniform Continuity

Let $f \in C_c(G)$. Then f is left and right uniformly continuous.

Proof **:**
Folland p.340

We will mainly bo working with Hausdorff groups. The following shows how this is not much of a restriction

Proposition 8.1.3: Topological Groups And Hausdorfness

Let G be a topological group. TheN

1. If G is T_1, G is Hausdorff
2. If G is not T_1, let H be the closure of $\{e\}$. Then H is a normal subgroup and G/H (with respect to the quotient topology) is hausdorff.

Proof **:**
Folland p.341

1. here
2. We first show H is normal. Evidently, H is the smallest (or minimal) closed subgroup of G (since $\{e\}$ is the smallest subgroup of G). Then H has to be normal, since if it werent then If $H' = gHg^{-1}$ was some conjugate such that $H \neq H'$[a], then $H' \cap H$ would be a closed subgroup smaller than H. Hence, G/H is well-defined, and it is a simple to check that it indeed is a topological group with respect to the quotient topology. Then $\{\overline{e}\}$ is closed since its preimage of H. Finally, by continuity of the binary operation, the pre-image of a closed set is closed so the set in the preimage:

$$(\overline{x}, \overline{x}^{-1}) \mapsto \overline{e}$$

is closed closed in the product topology and so each $\overline{x}$ is closed making G Hausdorff.

[a]Note that H' is closed since left and right multiplication is a homeomorphism

As a consequence of proposition 8.1.3(2), every Borel measurable function on G is constant on the cosets of H, and hence is effectively already a function on G/H, and so for the most parrt we may as well work with the Hausdorff group G/H. For many of the following results, we'll require the additional restraint that G be locally compact.

Definition 8.1.4: Left/Right Invariant Borel Measure

Let G be a locally compact group. Then a Borel measure μ on G is called *left-invariant* (resp. *right-invariant*) if $\mu(xE) = \mu(E)$ (resp. $\mu(Ex) = \mu(E)$) for all $x \in G$ and $E \in \mathcal{B}_G$. Similarly, a linear function I on $C_c(G)$ is called left- or right-invariatn if $I(L_x(f)) = I(f)$ or $I(R_x f) = I(f)$ respectively for all f.

Definition 8.1.5: Haar Measure

Let G be a locally compact topological group. Then a *left* (resp. *right*) *Haar measure* on G is a nonzero left-invariant (resp. right-invariant) Radon measure μ on G.

Example 8.2: Haar Measure

1. Most famously, the Lebesgue measure is a (left- and right-invariant) Haar measure on $\mathbb{R}^n$
2. The counting measure is a (left and right) Haar meausre on any group with the discrete topology
3. Let $\mathbb{T}$ be the circel group, and consider $f : [0, 2\pi] \to \mathbb{T}$ defined by
$$f(t) = (\cos(t), \sin(t))$$
then we may define the Haar measure by:
$$\mu(S) = \frac{1}{2\pi} m(f^{-1}(S))$$
where m is the Lebesgue meaure on $[0, 2\pi]$. We multiplied by $(2\pi)^{-1}$ so that $\mu(T) = 1$
4. Let $G = \mathbb{R}^\times$. Then a Haar measure μ can be given by
$$\mu(S) = \int_S \frac{1}{t} dt$$
For any Borel subset $S \subseteq \mathbb{R}^\times$. For example, if $S = [a, b]$, then $\mu(S) = \log(b/a)$. This measure is indeed translatoin invariant since:
$$\mu(gS) = \mu([ga, gb]) = \log(gb/ga) = \log(b/a) = \mu(S)$$
and similarly for Sg. We may define a similar measure by taking
$$\mu(S) = \int_S \frac{1}{|t|} dt$$
5. Let $G = \mathrm{GL}_n(\mathbb{R})$. Then one may define a (left and right) Harr measure to be
$$\mu(S) = \int_S \frac{1}{|\det(X)|^n} dX$$
where X is the Lebesgue meaure on $\mathbb{R}^{n^2}$ identified with the set of all $n \times n$ matrices (this follows from the change of variables formula).

With the Haar measure, we may naturally define Haar integrals of functions, namely

$$\int f(x) d\mu(x)$$

where μ is the Haar measure.