MEASURE, INTEGRATION, AND BANACH SPACES

Contents

1. HEAT EQUATION

Consider an infinite row of rooms with people in them. Suppose that in each time period and in each room 10% of people move to the room to the left and 10% of people move to the room to the right. Let $f(x,t)$ denote the number of people in room x at time t. Then we have

$$f(x,t+1)-f(x,t)=0.1f(x-1,t)+0.1f(x+1,t)-0.2f(x,t)$$
$$=0.1([f(x+1,t)-f(x,t)]-[f(x,t)-f(x-1,t)]$$

Supposing there are many rooms and each room has small people in it we can approximate this equation by

$$D_t f(x,t)=0.1D_xD_xf(x,t)$$

where D_i stands for the derivative with respect to i. This is called the heat equation.

Let's assume that f is defined on $[0,\pi]$, that we know the initial distribution of people $f(x,0)$ and that $f(0,0)$ and $f(pi,0)$ are 0.

A good function for $f(x,0)$ is $\sin(x)$ or $\sin(2x)$ or perhaps some possibly infinite linear combination of $\sin(nx)$s. Further it turns out that in some sense of convergence any nice function on $[0,\pi]$ can be approximated by a linear combination of $\sin(nx)$s. This suggests that the solution to the heat equation will have the form

$$f(x,t)=a_1(t)\sin(x)+a_2(t)\sin(2x)+\ldots$$

If we put this into the heat equation we get

$$a_1'(t)\sin(x)+a_2'(t)\sin(2x)+\ldots=0.1(-a_1(t)\sin(x)-a_2(t)\sin(2x)2^2-\ldots)$$

One solution then has $a_i'(t) = -0.1i^2 a_i(t)$ which obviously has solution $a_i(t) = C_i e^{-0.1(i^2)t}$ where C_i is a constant.

We can now see that a property of the solution is that as time goes on there the people disperse more and more. We now have to solve for the constants C_i.

The integral of $f(x,t)\sin(ix)$ from 0 to π is the integral of $a_i(t)\sin(ix)^2$ from 0 to π which is $a_i(t)\frac{\pi}{2}$. So C_i is equal to $\frac{2}{pi}e^{0.1(i^2)t}$ times the integral of $f(x,t)\sin(ix)$ from 0 to π. This still depends on t so there's probably more to do here.

2. HEAT EQUATION AND INTEGRATION

You failed to notice two dodgy things about my derivation.

1) I took the derivative of an infinite sum by by differentiating each term of the sum:

$$\frac{d}{dt}(a_1\sin(x) + a_2\sin(2x) + ...) = a_1'\sin(x) + a_2'\sin(2x) + ...$$

2) I took the integral of an infinite sum by integrating each term of the sum

$$\int_0^\pi (a_1\sin(x)\sin(mx) + a_2\sin(2x)\sin(mx) + ...)dx$$
$$= \int_0^\pi a_1\sin(x)\sin(mx)dx + \int_0^\pi a_2\sin(2x)\sin(mx)dx + ...)$$

Are these maneuvers valid?

If yes then I've solved the heat equation.

This is the question posed by Joseph Fourier that led to the development of a new theory of integration.

The more general and basic question is

3) Let $f_1, f_2, ...$ be a sequence of real valued functions defined on $[a,b]$ that converges pointwise to f. Does $\int_a^b f_n(x)dx$ converge to $\int_a^b f(x)dx$? That is, does $\lim_n \int_a^b f_n(x)dx = \int_a^b \lim_n f_n(x)dx$?

A good answer to this question will tell us when 1 and 2 are valid maneuvers.

Let me leave you with two examples that show that 3 can fail.

Define a sequence $f_1, f_2, ...$ of real valued functions on $[0,1]$ each having integral equal to 1. Suppose that $f_n(x)$ is positive if x belongs to $[\frac{n-1}{n}, \frac{n}{n+1}]$ and is otherwise 0.

This sequence of functions converges to $f = 0$ whose integral is 0. So for this sequence the answer to 3 is no. That is, $\int_0^1 f_n(x)dx = 1$ for all n but $\int_0^1 f(x)dx = 0$.

This example works because the functions are getting taller and taller. The area under the graph stays the same but the functions support is shrinking.

Thus for the answer to 3 to be yes we need the functions in the sequence $f_1, f_2, ...$ to be bounded: there is a number N such that $|f_1(x)| \leq N$ for all x in $[a,b]$.

Even if I do this I can still get a "no" answer to 3 as follows. Define the sequence $f_1, f_2, ...$ of real valued functions on $[0, 1]$ by saying that $f_n(x)$ is equal to 1 if x belongs to $\{\frac{1}{n}, \frac{2}{n}, ...\frac{n}{n}\}$ and is otherwise equal to 0.

This sequence of functions converges to a function f that is equal to 1 on a countable set and is otherwise equal to 0.

The answer to 3 is "no" because $\int_0^1 f_n(x)dx = 0$ for all n and this is not equal to $\int_0^1 f(x)dx$ because f is not Riemann integrable (the upper Riemann sum of f is equal to 1 and the lower Riemann sum is equal to 0). Note that we'd like the integral of f to be 0because f is equal to 0 at many more points than it is equal to 1 but the Riemann integral can't deal with a function like f.

Therefore another condition we need for the answer to 3 to be a yes is that the sequence of functions converges to a function we can integrate.

We could stick with the Riemann integral and put more restrictions on our sequence of functions. Or perhaps we could create a new integral, one that agrees with the Riemann integral when a function is Riemann integrable but which can integrate functions like f in my second example.

3. THE LEBESGUE INTEGRAL

Consider a real valued function f defined on $[a, b]$. The Riemann integral of f is constructed in the following way. The interval $[a, b]$ is divided into pieces $a < c_1 < ... < c_n < d$. The lower Riemann sum is

$$(c_1 - a)\inf\{f(x) \mid x \in [a, c_1]\} + ... + (d - c_n)\inf\{f(x) \mid x \in [c_n, d]\}.$$

The upper Riemann sum is

$$(c_1 - a)\sup\{f(x) \mid x \in [a, c_1]\} + ... + (d - c_n)\sup\{f(x) \mid x \in [c_n, d]\}.$$

The function f is Riemann integrable if the difference between the upper and lower Riemann sum can be made arbitrarily small by cutting the domain of the function up into enough pieces. The number the upper and lower Riemann sums converge to is called the Riemann integral of f.

If f is continuous then it is Riemann integrable. If f is Riemann integrable then it must be fairly continuous.

Not all functions are Riemann integrable. In the last note I showed that the function f defined on $[0, 1]$ that has value 1 when $x \in \{\frac{1}{n}, \frac{2}{n}, ..., \frac{n}{n}\}$ for some natural number n and has value 0 otherwise is not Riemann integrable. There we had a sequence of Riemann integrable functions converging to this f. This was a problem for interchanging the limit and integration operators.

Here is another way to compute the area under the graph of a function. Let f be a nonnegative bounded real valued function defined on $[a, b]$. For $0 < t_1 < t_2 <$

$... < t_n$, where t_n is an upper bound for f, let S_{t_i} be the set of points x such that $f(x) \geq t_i$. That is $S_{t_i} = \{x \mid f(x) \geq t_i\}$. For any set S define I_S to be the function that takes value 1 on S and 0 otherwise. Consider the Lebesgue sum

$$t_1 \int_a^b I_{S_{t_1}}(x)dx + (t_2 - t_1) \int_a^b I_{S_{t_2}}(x)dx + ... + (t_n - t_{n-1}) \int_a^b I_{S_{t_n}}(x)dx$$

. We are cutting the area under the graph into horizontal strips. The Riemann integral uses vertical strips. Why does this work. Well, we have that

$$(t_1 I_{S_{t_1}}(x) + (t_2 - t_1) I_{S_{t_2}}(x) + ... + (t_n - t_{n-1}) I_{S_{t_n}}(x)) + \max\{t_i - t_{i-1} \mid i$$
$$= 1, 2, ..., n\} \geq f(x) \geq t_1 I_{S_{t_1}}(x) + (t_2 - t_1) I_{S_{t_2}}(x) + ... + (t_n - t_{n-1}) I_{S_{t_n}}(x).$$

Taking the Riemann integral of each term in this inequality shows that as the image of the function is cut more finely the Lebesgue sum converges to the Riemann integral. (This of course supposes that all the functions in this inequality are Riemann integrable.) One advantage of this is that we don't need as much structure on the domain of the function to compute its integral.

In the Lebesgue sum above the integral $\int_a^b I_{S_{t_i}}(x)dx$ is a Riemann integral. It is the area of the set S_{t_i}. Notice though that the set S_{t_i} may be such that the function $I_{S_{t_i}}$ is not Riemann integral. The way we are going to make the Lebesgue integral more general than the Riemann integral is by finding a way to compute the area of sets S_{t_i} for which $I_{S_{t_i}}$ may not be Riemann integrable.

Given a set S in $\mathbf{R}^n$ let's denote the volume of S by $\mu(S)$ and call it the measure of S. We want μ to satisfy our intuitive notions of volume. Here are some reasonable demands

1) Translation and rotation invariance. The measure of a set S is the same as the measure of a translation or rotation of S.

2) The measure of the union of finitely many disjoint sets is the sum of their measures. In fact because this is analysis we're going to say that the measure of the union of countably many disjoint sets is the infinite sum of their measures. It isn't completely clear to me why this should be true.

So far, assigning each subset of $\mathbf{R}^n$ a measure of 0 satisfies these axioms so we will assume that some sets have positive measure

3) The measure of $[0, 1]^n$ is 1 for any $n \geq 1$. That is, the volume of the n-cube is one.

Here is a surprising result. It turns out that these axioms are inconsistent with one another, even in one dimension. Here is the proof. Define the relation $\sim on[0, 1]$ by saying that $x \sim y$ means that $x - y$ is a rational number. This is an equivalence relation and so it gives a partition of the set $[0, 1]$ into equivalence classes. Let S be a set containing exactly one point from each equivalence class. Now for each rational number q in $(0, 1)$ define S_q the be the set of points x in $[0, 1]$ such that $x - q$ belongs to $S + n$ for some integer n.

First of all, S_r and S_q are disjoint whenever r is not equal to q. To see this let q and r be rational numbers in $(0, 1)$ and let x be a point in S_q and S_r. This implies that $x - r$ belongs to $S - n_1$ for some integer n_1 and so $x - (r + n_1)$ belongs to S. Similarly $x - (q + n2)$ belongs to S for some integer n_2. But $(x - (r + n_1)) - (x - (q + n_2))$ is a rational number so that these two points belong to the same equivalence class. This implies that these points are the same so that $q = r$.

I now claim that the measure of S_q is equal to the measure of S. We can write S_q as the union of the sets $\{x \in S_q \mid x \leq q\}$ and $\{x \in S_q \mid x > q\}$. The first of these sets is equal to $\{x \in S \mid x \geq 1 - q\}$; the second of these sets if equal to $\{x \in S \mid x \leq 1 - q\}$. So by axiom 2 my claim is true.

Now consider the union or the collection of sets $\{S_q \mid q \in \mathbf{Q} \cap (0, 1)\}$. I claim that this union is equal to $[0, 1]$; clearly by the definition of S_q it is a subset of $[0, 1]$. Let x be a point in $[0, 1]$. Then x belongs to an equivalence class. Let y be the point of this equivalence class in S. Then $x - y$ is a rational number so that $x - (x - y) = y$ and y belongs to S. Thus x belongs to S_{x-y} and so is in the union of the collection of sets $\{S_q \mid q \in \mathbf{Q} \cap (0, 1)\}$. So the union of this collection of sets equals $[0, 1]$. Therefore by axiom 3 the measure of this union is 1.

Using these three observations and applying axiom 2 to the countable collection of disjoint sets $\{S_q \mid q \in \mathbf{Q} \cap (0, 1)\}$ gives

$$1 = \mu(\bigcup\{S_q \mid q \in \mathbf{Q} \cap (0, 1)\}) = \sum_{q \in \mathbf{Q} \cap (0,1)} \mu(S_q) = \sum_{q \in \mathbf{Q} \cap (0,1)} \mu(S).$$

But this is completely impossible. Either $\mu(S)$ is positive in which case the sum on the right diverges or it is zero in which case the sum on the right converges to zero. So we've proved that the axioms are inconsistent.

The way we are going to resolve this inconsistency is to restrict the collection of sets that we are allowed to take the measure of. We will still get to keep the three axioms but our function μ will have a smaller domain.

4. OUTER MEASURE

We want our measure μ to de a function from some set of "measurable" subsets of $\mathbf{R}^n$ to the extended real numbers (that is, $\mathbf{R} \cup \{\infty\}$) that satisfies our three axioms:

1) The measure of a measurable set is the same as the measure of a translation or rotation of this set. 2) The measure of a disjoint countable union of measurable sets is equal to the sum of the measure of each of the sets.

For the third axiom we will say that a set B is an open box open box of $\mathbf{R}^n$ if it can be written as the cartesian product of nonempty open intervals. That is, for some $a_1 < b_1, a_2 < b_2, ..., a_n < b_n$ we have $B = (a_1, b_1) \times (a_2, b_2) \times ... \times (a_n b_n)$.

Before, our third axiom was that the measure of the n-cube $[0, 1]^n$ is equal to one. Our new third axiom is a straightforward generalization of this.

3) The measure of an open box $B = (a_1, b_1) \times (a_2, b_2) \times ... \times (a_n b_n) is (b_1 - a_1)(b_2 - a_2)...(b_n - a_n)$ and we will denote this $Vol(B)$.

We want to find out what what subsets we can allow to be measurable so that these axioms do not contradict one another, i.e so there actually exists a measure satisfying the axioms. Hopefully the domain of this measure is large enough that whenever the Riemann integral $\int_a^b I_S(x)dx$ of the indicator function is defined then so is the measure of S and hopefully for many sets where this Riemann integral is not defined the measure will be. Then the Lebesgue integral will be more general than the Riemann integral. So we may be able to show that limits and integration commute in certain cases (It turns out that one can show this much more easily for the Lebesgue integral than directly for the Riemann integral).

This third axiom gives us the power to approximately measure sets. We will call our approximation function the outer measure.

The outer measure μ^* is a function from all subsets of $\mathbf{R}^n$ to the extended real numbers (that is, $\mathbf{R} \cup \{\infty\}$). The outer measure of a set S is the infimum of the measure of any union of open boxes that contain S. That is $\mu^*(S) = \inf\{\mu(B_1 \cup B_2 \cup ...) \mid B_1, B_2, ...$ are open boxes whose union contains $S\}$. We still don't know how to compute the measure of the union of open boxes unless they are disjoint. When they are not disjoint we have the following approximation. Let $B_1, B_2, ...$ be open boxes. Then $B_1, B_1 - B_2, B_3 - (B_2 \cup B_1), ...$ is a countable collection of disjoint open sets with the same union as $B_1, B_2, B_3, ...$ By axiom 2 we have

$$\mu(B_1 \cup B_2 \cup B_3 \cup ...) = \mu(B_1) + \mu(B_2 - B_1) + \mu(B_3 - (B_2 \cup B_1)) + ...$$
$$\leq \mu(B_1) + \mu(B_2) + \mu(B_3) + ...(*)$$

Let's compute the outer measure of some sets.

What is the outer measure of a singleton set $\{v\}$? It's zero because we can put this point in an arbitrarily small open box.

What is the outer measure of a countable set $\{v_1, v_2, ...\}$? It's also zero because we can put each point in an arbitrarily small box and then use the result that we just derived, that is result (*), about the measure of the union of open boxes being no more than the sum of the measures of these open boxes. Incidentally, this shows the rational numbers have outer measure zero.

What about the unit interval $[0,1]$. What is the outer measure of this set? We can compute an upper bound for this is 1 because we can place $[0,1]$ inside close fitting open box. We would hope that the outer measure of this set is 1 but this is difficult to show. The problem is that the outer measure of $[0,1] \cap \mathbf{Q}$ is zero because this is a countable set. This implies that we are going to have to use a property of the real numbers that the rational numbers do not have to show that the outer measure of $[0,1]$ is nonzero.

Finally, let's consider a k-dimensional subspace S of $\mathbf{R}^n$ where $k < n$. What is the outer measure of this set? Just like the volume of a square is zero and the area of a line is zero we would like the outer measure of a lower dimensional subspace to be zero. How can we show this?

First, let's show that $\mathbf{R}^k$ has zero outer measure. To do this first consider the k-dimensional square $[0,1] \times [0,1] \times ... \times [0,1]$. We will show that this has outer measure zero. It's possible to contain this square in the union of $(m+1)^k$ n-dimensional open boxes with edges of length $\frac{1}{m}$. The measure of each of these n dimensional open boxes is $\frac{1}{m^n}$. We can now use the approximation (*) to show that the measure of the k-dimensional square $[0,1] \times [0,1] \times ... \times [0,1]$ is no more than $\frac{(m+1)^k}{m^n}$. Since the m we chose was arbitrary this can be made as close to 0 as we want. So the outer measure of the k-dimensional square $[0,1] \times [0,1] \times ... \times [0,1]$ is zero.

Now tile $\mathbf{R}^k$ with translations of these k-dimensional squares. This can be done with countably many of these squares. Applying the approximation (*) again shows that the outer measure of $\mathbf{R}^k$ is zero.

Our k-dimensional subspace may be a rotation of $\mathbf{R}^k$ but this is fine - we can rotate it to $\mathbf{R}^k$ and then by axiom 1 it follows that its outer measure is also zero.

5. ALGEBRAIC PROPERTIES OF MEASURABLE SETS

A subset F of $\mathbf{R}^n$ is measurable if for all subsets S of $\mathbf{R}^n$ the outer measure of S is equal to the outer measure of the points common to both S and F plus the outer measure of the points that are in S but not in F. That is,

$$\mu^*(S) = \mu^*(S \cap F) + \mu^*(S - F).$$

What are the algebraic properties of the set of measurable sets. For example, is the union of two measurable sets measurable? Is the complement of a measurable set measurable?

Let's first show two things about the outer measure.

(1) The outer measure is countably subadditive: if $E_1, E_2, ...$ are subsets of $\mathbf{R}^n$ then the outer measure of their union is no more than the sum of their outer

measures. That is,

$$\mu^*(E_1 \cup E_2 \cup ...) \leq \mu^*(E_1) + \mu^*(E_2) + ...$$

This is true because if $B_1^1, B_2^1, ...$ are open boxes whose union contains $E_1, B_1^2, B_2^2, ...$ are open boxes whose union contains E_2, and so on, then the union of all these open boxes contains $E_1 \cup E_2 \cup ....$

(2) The outer measure is monotone: if E_1 and E_2 are subsets of $\mathbf{R}^n$ such that E_1 is a subset of E_2 then the outer measure of E_1 is no more than the outer measure of E_2.

This is true because if $B_1, B_2, ...$ are open boxes whose union contains E_2 then they are open boxes whose union contains E_1.

Let us now consider the algebraic properties of measurable sets.

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

Let E be a measurable set. Let S be a subset of $\mathbf{R}^n$. Then by the measurability of E $\mu^*(S \cap E^c) + \mu^*(S - E^c) = \mu^*(S - E) + \mu^*(S \cap E) = \mu^*(S)$ which implies E^c is measurable.

(4) The union of two measurable sets is measurable.

Let E_1 and E_2 be measurable sets. Let S be a subset of $\mathbf{R}^n$. By subadditivity of the outer measure $\mu^*(S \cap (E_1 \cup E_2)) + \mu^*(S - (E_1 \cup E_2)) \geq \mu^*(S)$. We can get the reverse inequality as follows. Not that $E_1 \cup E_2 = E_1 \cup (E_2 \cap E_1^c)$. Therefore by subadditivity and measurability

$$\mu^*(S \cap (E_1 \cup E_2)) + \mu^*(S - (E_1 \cup E_2))$$

$$\leq \mu^*(S \cap E_1) + \mu^*(S \cap (E_2 \cap E_1^c)) + \mu^*(S - (E_1 \cup E_2))$$

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

$$= \mu^*(S \cap E_1) + \mu^*(S - E_1) = \mu^*(S)$$

so we're done.

(5) The empty set is measurable.

Let S be a subset of $\mathbf{R}^n$. Then $\mu^*(S \cap \emptyset) + \mu^*(S - \emptyset) = \mu^*(\emptyset) + \mu^*(S)$ and this equals $\mu^*(S)$ because the outer measure of the empty set is zero.

Note that for free we have that the intersection of two measurable sets is measurable because the complement of the intersection of two sets is the union of their complements.

6. THE COLLECTION OF MEASURABLE SETS IS A SIGMA ALGEBRA

Let X be a set. A σ-algebra of X is a collection of subsets of X containing the empty set and which is closed under complements and countable unions. That is,

$\emptyset$ belongs to it, if E belongs then so does E^c, and if $E_1, E_2, ...$ each belong then so does the union of these sets.

(1) I claim that the measurable subsets of $\mathbf{R}^n$ are a σ-algebra of $\mathbf{R}^n$.

We showed in the previous note that the empty set is measurable. We also showed the complement of a measurable set is measurable.

All we need now do is show the countable union of measurable sets is measurable.

Let $E_1, E_2, ...$ be measurable sets. We may assume these are pairwise disjoint. The reason we may is that we can define the sequence of sets $E_1, E_2 - E_1, E_3 - (E_1 \cup E_2), ...$ and this is a pairwise disjoint sequence of measurable sets (each of these sets is measurable because of the algebraic properties of measurable sets we showed in the last note).

Our goal is to show the union $\bigcup\{E_1, E_2, ...\}$ is a measurable set. That is, we must show that for all subsets S of $\mathbf{R}^n$ the outer measure of S is equal to the outer measure of points common to both S and $\bigcup\{E_1, E_2, ...\}$ plus the outer measure of points that are in S and not in $\bigcup\{E_1, E_2, ...\}$. That is,

$$\mu^*(S) = \mu^*\left(S \cap \bigcup\{E_1, E_2, ...\}\right) + \mu^*\left(S - \bigcup\{E_1, E_2, ...\}\right).$$

We showed in the previous note that the outer measure is countably subadditive. This implies

$$\mu^*(S) \leq \mu^*\left(S \cap \bigcup\{E_1, E_2, ...\}\right) + \mu^*\left(S - \bigcup\{E_1, E_2, ...\}\right).$$

We are left only to show the reverse inequality. Let's first prove a lemma.

(2) Let E and F be measurable and disjoint sets. Let S be a subset of $\mathbf{R}^n$. Then the outer measure of S intersected with the union of E and F is equal to the outer measure of S intersected with E plus the outer measure of S intersected with F. That is,

$$\mu^*(S \cap (E \cup F)) = \mu^*(S \cap E) + \mu^*(S \cap F).$$

This is simple to prove:

$$\begin{aligned}\mu^*(S \cap (E \cup F)) &= \mu^*(S \cap (E \cup F) \cap E) + \mu^*(S \cap (E \cup F) - E) \\ &= \mu^*(S \cap E) + \mu^*(S \cap F).\end{aligned}$$

Let's prove another lemma.

(3) The outer measure is monotone: if E and F are subsets of $\mathbf{R}^n$ and E is a subset of F then the outer measure of E is no more than the outer measure of F. That is, $\mu^*(E) \leq \mu^*(F)$.

This follows from the definition of outer measure: If $B_1, B_2, ...$ are open boxes whose union contains F then the union of these open boxes contains E.

Going back to our proof of (1), for each natural number n we have, by the fact that a finite union of measurable sets is measurable (proved in the last note) and the fact that the outer measure is monotone, that

$$\begin{aligned}\mu^*(S) &= \mu^*\left(S \cap \bigcup\{E_1, E_2, ..., E_n\}\right) + \mu^*\left(S - \bigcup\{E_1, E_2, ..., E_n\}\right) \\ &= \sum_{i=1}^{n} \mu^*(S \cap E_i) + \mu^*\left(S - \bigcup\{E_1, E_2, ..., E_n\}\right) \\ &\geq \sum_{i=1}^{n} \mu^*(S \cap E_i) + \mu^*\left(S - \bigcup\{E_1, E_2, ...,\}\right).\end{aligned}$$

Since this is true for all n we have

$$\begin{aligned}\mu^*(S) &\geq \sum_{i=1}^{\infty} \mu^*(S \cap E_i) + \mu^*\left(S - \bigcup\{E_1, E_2, ...,\}\right) \\ &\geq \mu^*\left(S \cap \bigcup\{E_1, E_2, ...\}\right) + \mu^*\left(S - \bigcup\{E_1, E_2, ...\}\right).\end{aligned}$$

The last step is because the outer measure is countably subadditive.

7. THE LEBESGUE MEASURE

Recall our axioms for measure:

(1) μ is a function from a class of subsets of $\mathbf{R}^n$ to the extended real numbers $[0, \infty]$. (2) Translation invariance: The measure of a set is the same as the measure of a translated version of that set. That is, for all sets Sin the domain and for all v in $\mathbf{R}^n$, $\mu(S) = \mu(S + \{v\})$. (3) Countable additivity: The measure of countably many disjoint sets $E_1, E_2, ...$ is equal to the sum of their measures. That is,

$$\mu(E_1 \cup E_2 \cup ...) = \mu(E_1) + \mu(E_2) + ...$$

(3) The measure of the n-box $[0, 1]^n$ is one. That is, $\mu([0, 1]^n) = 1$.

We saw that these axioms are contradictory when the domain of the measure is any subset of $\mathbf{R}^n$. Consequently, our approach was to limit the domain of the measure. The way we did this was to define the outer measure of any subset S of $\mathbf{R}^n$ by the formula

$$\mu^*(S) = \inf\{\sum_{i=1}^{\infty} \mathrm{Vol}(B_i) \mid B_1, B_2, ...\text{are open boxes whose union contains } S\}$$

We then defined a subset S of $\mathbf{R}^n$ to be measurable if for any subset E of $\mathbf{R}^n$ the outer measure of E is equal to the outer measure of the points common to E and S plus the outer measure of the points in E and not in S. That is,

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

Then we showed that the collection of measurable sets is a σ-algebra. That is, it contains the empty set, is closed under complements, and is closed under countable unions.

We have shown that the outer measure satisfies translation invariance for any set S. We have shown that the outer measure of the n-cube $[0,1]^n$ is one.

[1] What we will do now is show that when the outer measure's domain is restricted to measurable sets it is countably additive. We will then have completed our program of constructing a measure that satisfies our axioms.

We have shown previously that the outer measure is countably subadditive and monotone. Suppose we were able to show the outer measure of the union of two disjoint sets is the sum of their outer measures. The proof is then trivial. Let $E_1, E_2, ...$ be measurable and disjoint sets. Then by monotonicity

$$\mu^*(E_1 \cup E_2 \cup ...) \geq \mu^*(E_1, \cup E_2 \cup ... \cup E_n) = \mu^*(E_1) + \mu^*(E_2) + ... + \mu^*(E_n).$$

Since this holds for all n it holds in the limit. That is,

$$\mu^*(E_1 \cup E_2 \cup ...) \geq \mu^*(E_1) + \mu^*(E_2) + ...$$

The reverse inequality holds by subadditivity.

Let's show that if E_1 and E_2 are disjoint measurable sets then the outer measure of their union is the sum of their outer measures. Let S be a subset of $\mathbf{R}^n$. By measurability of $E_1 \cup E_2$ we have

$$\mu^*(S \cap (E_1 \cup E_2)) = \mu^*(S \cap (E_1 \cup E_2) \cap E_2) + \mu^*(S \cap (E_1 \cup E_2) - E_2)$$
$$= \mu^*(S \cap E_2) = \mu^*(S \cap E_1).$$

Taking $S = E_1 \cup E_2$ gives

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

So we're done. We have shown that the outer measure defined on the collection of measurable sets satisfies our axioms of measure. For this reason we give it a special name, we call it Lebesgue measure.

8. UNIQUENESS OF THE LEBESGUE MEASURE

We have shown that there exists a function μ^* from the set of measurable subsets of R^n to the extended real numbers $[0,\infty]$ such that

(1) μ is translation invariant: for any measurable set S the measure of μ is that same as the measure of μ translated by any vector v. That is, $\mu(S + \{v\}) = \mu(S)$.

(2) μ is countably additive: if $E_1, E_2, ...$ are measurable and disjoint then the measure of their union is the sum of their measures. That is, $\mu(E_1 \cup E_2 \cup ...) = \mu(E_1) + \mu(E_2) + ...$

(3) The measure of the n-cube is one. That is, $\mu([0,1]^n) = 1$.

The measure we found that satisfies these axioms was the outer measure μ^* defined by

$\mu^*(S) = \inf\{\text{Vol}(B_1)+\text{Vol}(B_2)+... \mid B_1, B_2, ...$ open boxes whose union contains $S\}$

for any subset S of $\mathbf{R}^n$. Here, an open box is the cartesian product of n open intervals.

A measurable subset of $\mathbf{R}^n$ is any subset E such that for each subset S of $\mathbf{R}^n$ the outer measure of S is equal to the outer measure of the points common to S and E plus the outer measure of the points in S but not in E. That is,

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

Because μ^* satisfies out axioms on the measurable sets we called it Lebesgue measure and denoted it by μ.

Here is a question: Is the Lebesgue measure the only measure that satisfies the axioms on the measurable sets?

Suppose m is a function from the measurable sets to the extended real numbers satisfying the same axioms. By countable additivity we can deduce that the measure m of any open box with sides of length $\frac{1}{m}$ is equal to the volume of this open box which is the Lebesgue measure of this open box.

Let E be a Lebesgue measurable subset of $\mathbf{R}^n$. let $\epsilon > 0$. There exist open boxes $B_1, B_2, ...$ whose union covers S such that

$$\mu(S) + \epsilon \geq \mu(B_1) + \mu(B_2) + ... = m(B_1) + m(B_2) + ... \geq m(B_1 \cup B_2 \cup ...) \geq m(S).$$

Since this holds for all ϵ we have $\mu(S) \geq m(S)$. I'll leave the reverse inequality for next time.

For the reverse inequality let's first consider a measurable subset S of the n-cube $[0,1]^n$. Let T be the set of points in $[0,1]^n$ and not in S. That is, T equals the set $[0,1]^n - S$. We have shown that $\mu(T)$ is at least $m(T)$. Since the n-cube is measurable, countable additivity of μ gives $1 = \mu(S) + \mu(T)$ and countable additivity of m gives $1 = m(S) + m(T)$. So $\mu(S) = 1 - \mu(T) \geq 1 - m(T) = m(S)$. By translation invariance of μ and m this holds whenever S is a translate of $[0,1]^n$.

Now suppose that S is any measurable subset of $\mathbf{R}^n$. Let v be a point of $\mathbf{R}^n$ with integer coordinates. Then

$$\mu(S) = \mu(S \cap \bigcup\{[0,1]^n + \{v\} \mid v \in \mathbf{Z}^n\}$$

$$= \sum_v \mu(S \cap ([0,1]^n + \{v\})) = \sum_v m(S \cap ([0,1]^n + \{v\}))$$

$$= m(S \cap \bigcup\{[0,1]^n + \{v\} \mid v \in \mathbf{Z}^n\} = m(S).$$

So we are done.

This is great because now we know that the Lebesgue measure satisfies our axioms on the measurable sets and also that it is the only function on the measurable sets that satisfies our axioms.

Would this still hold if we did not require the new measure to satisfy axiom 3: the measure of the n-cube is one? Well, suppose we had a function m from the measurable subsets to the extended real numbers satisfying translation invariance and countable additivity, and suppose that for this function $m([0,1]^n) = t$ where $0 < t < \infty$ (if $t = 0$ then the measure of any set is zero; if $t = \infty$ then the measure of most sets in infinite). But note that the function $\frac{1}{t}m$ satisfies all three axioms, so by our previous result this function is equal to the Lebesgue measure. So in this case for any measurable set S we have that $m(S) = t\mu(S)$.

One final thing for this email. Let's show that if I take a measurable set in $\mathbf{R}^n$ and rotate it then the Lebesgue measure of the original set is equal to the Lebesgue measure of the rotated set. Let g be our rotation function. Let m be the function on the measurable sets defined by the formula $m(S) = \mu(g(S))$. That is, m is the function that given a set computes the Lebesgue measure of the rotation of that set under the rotation function g. Note that this function is well defined because the rotation of a measurable set is measurable. We want to show that $m(S) = \mu(S)$ for every measurable subset S of $\mathbf{R}^n$. A smart way to do this is to show that m satisfies our three axioms. Then we can apply our last result that the Lebesgue measure is the only function satisfying these axioms.

Is m translation invariant?

Let S be a measurable subset of $\mathbf{R}^n$ and v a point in $\mathbf{R}^n$. Then because Lebesgue measure is translation invariant and a rotation is a linear transformation we have

$$m(S + \{v\}) = \mu(g(S + \{v\})) + \mu(g(S) + \{v\}) = \mu(g(S)) = m(S).$$

So m is translation invariant.

Is m countably additive? Let $S_1, S_2, ...$ be measurable and disjoint subsets of $\mathbf{R}^n$. Then because the rotation function g is a linear transformation and keeps disjoint sets disjoint and because the Lebesgue measure is countably additive we have

$$m(S_1 \cup S_2 \cup ...) = \mu(g(S_1) \cup g(S_2) \cup) = \mu(g(S_1)) + \mu(g(S_2)) + ... = m(S_1) + m(S_2) + ...$$

So m is countably additive.

We now know from our previous work that m is proportional to Lebesgue measure.

Does m assign a value of 1 to the n-cube?

Let t denote the number m assigns to the n-cube. We know that for any measurable set S of $\mathbf{R}^n$ we have $m(S) = t\mu(S)$.

Consider the unit ball B of $\mathbf{R}^n$. Because g is a rotation function it takes B to itself. Therefore $t\mu(B) = m(B) = \mu(g(B)) = \mu(S)$.

So $t = 1$ and we are done.

9. MEASURABLE FUNCTIONS

Our goal has been to develop a new theory of integration. Our motivation was to figure out when the integral of a function is equal to the sequence of integrals of functions that converge to the function. That is, if $f_1, f_2, ...$ is a sequence of functions converging (in some way) to a function f when does $\int f$ equal the limit of the sequence $\int f_1, \int f_2, ...$? This problem was motivated by our trying to find a solution to the heat equation.

So far we have defined Lebesgue measure. We know how to measure the area of a large class (the class of measurable subsets) of subsets of $\mathbf{R}^n$. The reason we constructed the Lebesgue measure is because we had in mind a way to use it to compute the Lebesgue integral of a function. The idea of the Lebesgue inegral was to slice the range of a function into intervals and to look at the area of the pre-image of each of these intervals. To get the area under the graph of the function we than just sum over the intervals of the partition taking the largest value the function takes in the interval and multiplying this by the area of the pre-image.

Before we do this we need to talks about what types of functions we can integrate. The trouble is that the pre-image of an interval may not be a measurable set. Since we don't know how to compute the area of a set which is not measurable we would not be able to integrate a function for which the pre-image of some interval in the range was not a measurable set. So let's make the following definition:

Definition: A function $f : \mathbf{R}^n \to \mathbf{R}$ is called measurable if the set $\{x \in \mathbf{R}^n \mid f(x) \leq t\}$ is a measurable set for all real numbers t.

Next time we will prove things about measurable functions.

Let's build some tools.

Claim 1: If $f : \mathbf{R}^n \to \mathbf{R}$ is a measurable function, then the set $\{x \in \mathbf{R^n} \mid f(x) < t\}$ is measurable.

Proof: We can write this set as the countable union of sets: $\bigcup_{n=1}^{\infty}\{x \in \mathbf{R}^n \mid f(x) \leq t - \frac{1}{n}\}$. Each set in this union is measurable by the definition of a measurable function. Because the collection of measurable sets is a σ-algebra, the countable union of measurable sets is measurable. QED

Now a definition. Perhaps we would like to integrate a function $f : \mathbf{R}^n \to \mathbf{R}^m$. In this case we can write f as the m-tuple of functions $(f_1, f_2, ..., f_m)$ where each coordinate of this m-tuple is a function from $\mathbf{R}^n to \mathbf{R}$. When we integrate f what we want is the m-tuple $(\int f_1, \int f_2, ..., \int f_m)$. So our definition for the measurability of f should be that each of the component functions is measurable.

Definition: A function $f : \mathbf{R}^n \to \mathbf{R}^m$ where $f = (f_1, f_2, .., f_m)$ is a measurable function if each f_i is a measurable function.

Since our definition of the Lebesgue integral will consider the pre-image of an interval in the range of a function we would like the pre-image of intervals to be measurable sets. Let's prove something like this for open boxes.

Claim 2: If $f : \mathbf{R}^n \to \mathbf{R}^m$ is a measurable function and B is an open box of $\mathbf{R}^m$, then the set $\{x \in \mathbf{R^n} \mid f(x) \in B\}$ is a measurable set.

Proof: The open box B can be written as $(a_1, b_2) \times (a_2, b_2) \times ... \times (a_m, b_m)$. We can write f as the m-tuple $(f_1, f_2, ..., f_m)$. Then the set we want to show is measurable can be written as the intersection of sets:

$$\bigcap_{i=1}^{m}(\{x \in \mathbf{R}^n \mid f_i(x) \leq a_i\}^c \cap \{x \in \mathbf{R}^n \mid f_i(x) \geq b_i\}^c)$$

which is measurable because the sets in this expression are measurable and the measurable sets form a σ-algebra. QED

Just for fun, we can prove the more general claim that the pre-image of any open set under a measurable function is a measurable set.

Claim 3: If $f : \mathbf{R}^n \to \mathbf{R}^m$ is a measurable function and U is an open subset of $\mathbf{R}^m$, then the set $\{x \in \mathbf{R}^m \mid f(x) \in U\}$ is measurable.

Proof: We can write U as the countable union of open boxes $B_1, B_2, ...$ with rational coordinates. The set we want to show is measurable can then be written as

$$\{x \in \mathbf{R}^n \mid f(x) \in \bigcup_{i=1}^{\infty} B_i\}$$

which equals the set

$$\bigcup_{i=1}^{\infty}\{x \in \mathbf{R}^n \mid f(x) \in B_i\}.$$

By Claim 2 this expression is the countable union of measurable sets. Since the set of measurable sets is a σ-algebra it is measurable. QED

And, just for fun, the converse to Claim 3 is true. That is,

Claim 4: If $f : \mathbf{R}^n \to \mathbf{R}^m$ is a function and for each open subset U of $\mathbf{R}^m$ the set $\{x \in \mathbf{R}^n \mid f(x) \in U\}$ is measurable, then f is a measurable function.

Proof: Let $x_1, x_2, ..., x_m$ be real numbers. We would like to show that the set $\{x \in \mathbf{R}^n \mid f(x) \leq (x_1, x_2, ..., x_m)\}$ is a measurable set. This is true because we can write this set as the intersection of sets, that by our hypothesis, are measurable sets:

$$\bigcap_{i=1}^{m}\{x \in \mathbf{R}^n \mid f_i(x) \in (x_i, \infty)\}^c.$$

QED

This last claim is interesting because it implies that the notion of a measurable function does not depend on the coordinate system we use on Euclidian space. It only depends on the open sets.

It will be useful to know whether the product and sum of measurable functions is measurable. And they are. The following claim helps us show this.

Claim 1: If $f : \mathbf{R}^n \to \mathbf{R}^m$ is a measurable function and $g : \mathbf{R}^m \to \mathbf{R}^l$ is a continuous function, then the composition $g \circ f : \mathbf{R}^n \to \mathbf{R}^l$ is a measurable function.

Proof: By Claim 3 and Claim 4 on the previous note we can show that $g \circ f$ is a measurable function by showing that the pre-image of any open set U under this function is a measurable set. Let's do this. But this is certainly true because one definition of a continuous function is that the inverse image each open set is an open set. So

$$(g \circ f)^{-1}(U) = f^{-1} \circ g^{-1}(U)$$

and this is the inverse image of an open set by a measurable function, which is a measurable set. QED

A corollary of this claim is that the sum of two measurable functions is a measurable function and the product of two measurable functions is a measurable function. This follows because addition is a continuous function from $\mathbf{R}^2$ to $\mathbf{R}$ and product is a continuous function from $\mathbf{R}^2$ to $\mathbf{R}$.

Here is an interesting example that shows the composition of two measurable functions need not be measurable.

Example: For each number t in the unit interval $[0, 1]$ consider the binary expansion of t:

$$\frac{i_1}{2} + \frac{i_2}{4} + \frac{i_3}{8}...$$

where $i_1, i_2, ...$ is a sequence of 0s and 1s. Note that this expansion is sometimes not unique. For example, 10101111111... where the 1s repeat is the same as 1011000000... where the zeros repeat. Let's rule out one of these cases and suppose that our sequence $i_1, i_2, ...$ never becomes an infinite sequence of 1s.

Consider the function $f : [0, 1] \to [0, 1]$ defined by the formula

$$f(t) = \frac{2i_1}{3} + \frac{2i_2}{9} + \frac{2i_3}{27}$$

. So this function maps the sequence 10010110... to 20020220... and reads it in base 3. The image of this function is contained in a famous set called the Cantor set. The Cantor set is the subset of the unit interval $[0, 1]$ whose base three expansions have no 1s. Another way to think about this set is that you begin with the unit interval $[0, 1]$, then you remove the middle third of this interval so your set is now $[0, 1/3] \cup [2/3, 1]$, and you then remove the middle thirds of both of these intervals, and so on forever. One of the reasons this set is interesting is that it is an uncountable and closed set with measure zero.

Getting back to our example, we know that the image of f is a subset of the Cantor set and so, by monotonicity of Lebesgue measure, its measure is zero. We also know that the function f is injective (one-to-one): each ternary expansion is unique because each binary expansion is unique. It is easy to show that any non-decreasing function is measurable. Since f is non-decreasing it is also measurable.

Now let's show that the composition of two measurable functions need not be a measurable function. Let h be any function from $[0,1]$ to the real numbers. Consider the function $g : [0,1] \to \mathbf{R}$ where $g(t) = h \circ f^{-1}(t)$ if $t \in f([0,1])$ and $g(t) = 0$ otherwise. Note that this function takes a value of 0 for all values in the unit interval, except those in the image of f. We have previously shown that any function that agrees with a measurable function except on a set of measure zero is also a measurable function. Therefore g is a measurable function because it agrees with the zero function (a measurable function) everywhere except on a set of measure zero.

What happens if you compose g with f. You get

$$g \circ f = h \circ f^{-1} \circ f = h.$$

But h was any real valued function on $[0,1]$. And since there are functions on $[0,1]$ that are not measurable we have constructed an example of composing two measurable functions and getting a function that is not measurable.

10. SEQUENCES OF MEASURABLE FUNCTIONS

Recall that the purpose of our new theory of integration is to allow us decide when the limit of a sequence of integrals $\int f_1, \int f_2, \int f_3, ...$ converges to the integral of the limit function f (the limit of the sequence of functions $f_1, f_2, ...$).

In our pursuit of this goal a natural question is whether the limit f of a sequence of measurable functions $f_1, f_2, ...$ is a measurable function. If this were not the case out theory would be hobbled by the requirement that we could only consider particular sequences of measurable functions. But fortunately this is not the case. In fact the point-wise limit of measurable functions is a measurable function.

Theorem 1: If $f_1, f_2, ...$ is a sequence of measurable functions converging point-wise to a function f, then f is a measurable function.

Proof: This is where the fact that the set of measurable sets being a sigma algebra is useful. Recall that a function $f : \mathbf{R}^n \to \mathbf{R}$ is a measurable function if the set $\{x \in \mathbf{R}^n \mid f(x) \leq t\}$ is a measurable set. We can rewrite this set as

$$\{x \in \mathbf{R}^n \mid \forall k > 0 \exists m \text{ such that } \forall n \geq m f_n(x) \leq t - 1/k\}.$$

This can be written as

$$\bigcap_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{n \geq m} \{x \in \mathbf{R}^n \mid f_n(x) \leq t - 1/k\}.$$

but, because each function f_n is a measurable function the sets in brackets are measurable. And because the set of measurable sets is a σ-algebra the expression, a countable intersection of a countable union of a countable intersection of measurable sets, is itself a measurable set. QED

That's good.

As we will see our theory of integration will be built on something called simple functions.

Definition: A simple function $f : \mathbf{R}^n \to \mathbf{R}$ is a measurable function whose image is a finite set.

Let f be a measurable function and denote its image by $\{y_1, y_2, \cdots, y_n\}$. Then there exist disjoint and measurable sets $E_1, E_2, ..., E_n$ on which f takes each of its values. We may then express the function f by the formula

$$f(x) = y_1 \chi_{E_1}(x) + y_2 \chi_{E_2}(x) + \cdots + y_n \chi_{E_n}(x)$$

where χ_S denotes the characteristic function on S, that is the function that takes a value of one on S and zero elsewhere. We will build up our theory of integration using these functions. In particular it will be very easy for us to take the integral of these functions. And we will approximate the integral of other functions by simple function that are close. Thus, another natural question is what sort of simple functions can we approximate using simple functions. The answer turns out to be all measurable functions.

Theorem 2: If f is a nonnegative and bounded measurable function, then there exists a sequence of simple functions $f_1, f_2, ...$ that converge pointwise to f.

Proof: For each $n = 1, 2, ...$ define the function $f_n(x) = \frac{k}{n}$ if $\frac{k}{n} \leq f(x) \leq \frac{k}{n+1}$ for $k = 1, 2, ...$ Each function f_n is measurable because f is a measurable function and to the sets on which each f_n takes a value of $\frac{k}{n}$ is a measurable set. And because f is bounded we have that the image of each f_n is a finite set. So $f_1, f_2, ...$ is a sequence of measurable functions which clearly converges to f. QED

11. LITTLEWOOD'S THREE PRINCIPLES

J.E.Littlewood was a British mathematician. He proposed three key principles of measure theory. The first is that every measurable set is almost a finite union of open boxes. The second is that every pointwise convergent sequence of measurable functions almost converges uniformly. The third is that every measurable function is almost a continuous function. Let's state and prove each of these.

Principle 1: Every measurable set is almost a finite union of open boxes.

The idea is the following.

Theorem 1: Let E be a measurable subset of $\mathbf{R}^n$ with finite measure. Littlewood's first principle says that if $\epsilon > 0$, then there exists a set $E' \subset \mathbf{R}^n$ such that $\mu(E - E'), \mu(E' - E) < \epsilon$.

The restriction that E be a set with finite measure is important. The result would otherwise not be true. For example, it would not be true if E were the union of open boxes of measure one in $\mathbf{R}^n$ centered at points whose coordinates are all prime numbers.

To prove the result we will approximate E with a compact set and then cut up a finite cover of this compact set.

Proof: Let δ_1 and δ_2 be positive real numbers. By a previous result there exists a compact set $K \subseteq E$ such that $\mu(E - K) < \delta_1$. Let $B_1, B_2, \cdots$ be open boxes whose union cover K such that $\mu(K) + \delta_2 \geq \mu(B_1) + \mu(B_2) + \cdots$. Since K is compact it is contained in the union of finitely many of the open boxes (say, $B_1, B_2, \cdots, B_n$). Then

$$\mu(K) + \delta_2 \geq \mu(B_1) + \mu(B_2) + \cdots \mu(B_n) \geq \mu(B_1 \cup B_2 \cup \cdots \cup B_n).$$

Denote the open box B_i by

$$(a_1^i, b_1^i) \times (a_2^i, b_2^i) \times \cdots (a_n^i, b_n^i)$$

. Let c_j^k be the k'th biggest element of $\{a_j^1, b_j^1, a_j^2, b_j^2, \cdots, a_j^n, b_j^n\}$. Consider the collection of open boxes

$$\{C \mid C \text{ can be written as } (c_1^m, c_1^{m+1}) \times (c_2^t, c_2^{t+1}) \times \cdots \times (c_n^s, c_n^{s+1})\}$$

which I'll denote by $\mathcal{C}$.

This collection of open boxes is finite. Its union is a subset of $B_1 \cup B_2 \cup \cdots \cup B_n$ and the difference $(B_1 \cup B_2 \cup \cdots \cup B_2) - \bigcup \mathcal{C}$ has Lebesgue measure zero since it it the union of subsets of $\mathbf{R}^n$ contained in subspaces of dimension less than n. Therefore the Lebesgue measure of $B_1 \cup B_2 \cup \cdots B_n$ is equal to the Lebesgue measure of $\bigcup \mathcal{C}$. Let E' denote $\bigcup \mathcal{C}$. By countable additivity and monotonicity of μ we have

$$\mu(E - E') = \mu(E - K) - \mu(E' - K) = \mu(E - K) - (\mu(E') - \mu(K \cap E'))$$
$$\leq \mu(E - K) + \mu(B_1 \cup B_2 \cup \cdots \cup B_n) - \mu(K) < \delta_1 + \delta_2.$$

Since we can choose $\delta_1 + \delta_2$ to be smaller than ϵ this proves the first inequality. Likewise

$$\mu(E' - E) = \mu(E' - K) - \mu(E - K) = \mu(E' - K) - \mu(E - K)$$
$$= \mu(E') - (\mu(E' \cap K) - \mu(E - K))$$

$$\leq \mu(B_1 \cup B_2 \cup \cdots \cup B_n) - \mu(K) + \mu(E - K) < \delta_1 + \delta_2.$$

QED

Principle 2: Every pointwise convergent sequence of measurable functions almost converges uniformly.

First of all, what does pointwise convergence and uniform convergence mean?

Let $f_1, f_2, ...$ be a sequence of functions. We say that this sequence converges pointwise to a function f if for each point x the sequence of numbers $f_1(x), f_2(x), ...$ converges to the number $f(x)$. Uniform convergence is stronger. Not only do we need the sequence of functions to converge pointwise but we each sequence of numbers $f_1(x), f_2(x), ...$ to converge at at least some speed. That is, for each number $\epsilon > 0$ there exists a number N such that for all points x we have $\mid f(x) - f_n(x) \mid < \epsilon$ whenever $n \geq N$.

The formal statement of Littlewood's second principle is the following.

Theorem 2: (Egorov) Let E be a measurable subset of $\mathbf{R}^n$ which has finite measure. If $f_1, f_2, ...$ is a sequence of measurable functions defined on E that converges pointwise to a function f, then for each $\epsilon > 0$ there exists a subset E' of E such that $\mu(E - E') < \epsilon$ such that $f_1, f_2, ...$ converges uniformly to f on E'.

We also need the hypothesis that E is a set with finite measure. To see why consider the following sequence of measurable functions. Let f_n denote the function from $\mathbf{R}$ to $\mathbf{R}$ such that $f_n(x) = 1$ when $x \geq n$ and $f_n(x) = 0$ otherwise. But the only sets on which this sequence converges uniformly are bounded above.

Proof: Consider the set of points in E such that some function in our sequence with an index greater than n is at least distance $\frac{1}{k}$ from f. That is, the set

$$S_{i,k} = \left\{ x \in E \mid \text{there exists } j \geq i \text{ such that } \mid f(x) - f_j(x) \mid \geq \frac{1}{k} \right\}.$$

A point x belongs to the intersection $S_{1,k} \cap S_{2,k} \cap ...$ if for every n the number $f_n(x)$ is at least distance $\frac{1}{k}$ from $f(x)$. This is not possible though because our sequence of functions $f_1, f_2, ...$ converges pointwise to f. Also, $E \supseteq S_{1,k} \supseteq S_{2,k} \supseteq ...$ and so by the continuity of measure (a result we proved earlier) $\lim_i \mu(S_{i,k}) = 0$. Therefore we can find a number n_k such that the measure of $S_{n_k,k}$ is less than $\frac{\epsilon}{2^k}$. That is, $\mu(S_{n_k,k}) < \frac{\epsilon}{2^k}$. We can do this process for all k to get the sets $S_{n_1,1}, S_{n_2,2}, ...$ Clearly we can choose the numbers $n_1, n_2, ...$ to be increasing. The set $S_{n_k,k}$ is a set of points where the function is not within $1/k$ of f. So let $E' = E - (S_{n_1,1} \cup S_{n_2,2} \cup ...)$. Does the sequence of functions $f_1, f_2, ...$ converge uniformly on this set? Well, let $\delta > 0$. Uniform convergence on this set means that there is a number N such that if x is a point in this set and $n \geq N$, then $f_n(x)$ is within distance δ of $f(x)$. For some k we have that $\frac{\epsilon}{2^k} < \delta$ and we know no point in the set E' belongs to set $S_{n_k,k}$. We can use n_k as N in the definition of uniform convergence. Hence the sequence

$f_1, f_2, ...$ converges uniformly to f on E'. What about the measure of $E - E'$. Well, this is just the measure of $S_{n_1,1} \cup S_{n_2,2} \cup ...$ and by our clever construction we have that

$$\mu(S_{n_1,1} \cup S_{n_2,2} \cup ...) \leq \mu(S_{n_1,1}) + \mu(S_{n_2,2}) + ... < \epsilon.$$

QED

Principle 3: Every measurable function is almost continuous.

Theorem 3: (Lusin) Let $f : \mathbf{R^n} \to \mathbf{R}$ be a measurable function. Then for each $\epsilon > 0$, there exists a continuous function $g : \mathbf{R}^n \to \mathbf{R}$ such that the set $\{x \in \mathbf{R}^n \mid f(x) \neq g(x)\}$ has measure $< \epsilon$.

Littlewood's description is accurate this time. We don't need any condition such as for the domain of the function to be finite. We will prove the theorem for when $f : E \to \mathbf{R}$ and E is a measurable subset of $\mathbf{R}^n$.

Proof: First assume that f is a simple function. Then we can write f as $y_1\chi_{E_1} + y_2\chi_{E_2} + ... + y_n\chi_{E_n}$ where $y_1, y_2, ..., y_n$ are real numbers and $E_1, E_2, ..., E_n$ are measurable and disjoint sets. Consider the set E_1. By a previous result there exists a compact subset K_1 of E_1 such that $\mu(E_1 - K_1) < \frac{\epsilon}{n}$. Do this for $E_2, ..., E_n$ too. There is a positive number c such that each of these sets is distance at least c apart. Now define the function g_i by saying that $g_i(x)$ is equal to 1 whenever x belongs to K_i, equal to $1 - \frac{2d(K_i,x)}{c}$ when $d(x, K_i) < c$, and equal to 0 otherwise. Then define $g = g_1 + g_2 + ... + g_n$. The function g is continuous because it is the sum of continuous functions. Also $f(x) = g(x)$ whenever x is in $K_1 \cup K_2 \cup ... \cup K_n$. Therefore the measure of the set $\{x \in E \mid f(x) \neq g(x)\}$ is equal to

$$\mu\left(E - \bigcup_{i=1}^{n} K_i\right) = \mu\left(\bigcup_{i=1}^{n}(E_i - K_i)\right) = \sum_{i=1}^{n} \mu(E_i - K_i) < \epsilon.$$

So we have proved the theorem for the case where f is a simple function.

Now consider the case where $f : E \to \mathbf{R}$ is any measurable function with $\mu(E) < \infty$. We can assume that f is negative because if the theorem is true for nonnegative functions it is also true for non nonnegative functions. This is because any nonnegative function can be written as the difference of two nonnegative functions, that is, we let f_- be the function whose value is $-f(x)$ whenever $f(x) \leq 0$ and 0 otherwise, and we let f_+ be the function whose value is $f(x)$ whenever $f(x) \geq 0$ and 0 otherwise. Then we can write the function f as $f_+ - f_-$. Using our assumption that the measure of E is finite we can also assume that the function f is bounded. If it isn't bounded we can choose a number n such that the measure of the set of points in E that take values above n has very small measure. This follows from the continuity of measure. Let's also assume that the function is bounded above by 1 - If we can prove the result for this case we can prove it for any bound by multiplying the function by a constant.

With these assumptions - that f is nonnegative and bounded above by 1, and our earlier assumption that the measure of E is finite, let's prove the theorem. Define for each $n = 0, 1, ...$ the simple function f_n by the formula

$$f_n(x) = \frac{i}{n} \text{ if } \frac{i}{n} \leq f(x) < \frac{i+1}{n} \text{ for } i = 0, 1, ...$$

This is a sequence of simple functions converging to f. Let $\epsilon > 0$. For each f_n there exists a continuous function g_n that agrees with f_n except perhaps on a set S_n with measure less than $\frac{\epsilon}{2^n}$. Each of the functions g_n is uniformly continuous. The sequence $g_1, g_2, ...$ converges pointwise to f on the set $E - (S_1 \cup S_2 \cup ...)$. By Egorov's theorem there exists a subset E' of $E - (S_1 \cup S_2 \cup ...)$ on which $g_1, g_2, ...$ converges uniformly to f and such that $\mu(E - (S_1 \cup S_2 \cup ...) - E') < \epsilon$. Denote the function that $g_1, g_2, ...$ converges to by g. This function is continuous because a sequence of uniformly continuous functions converge uniformly to it. Finally,

$$\mu\left(\{x \in E \mid f(x) \neq g(x)\}\right) \leq \mu\left(E - E'\right)$$

$$= \mu\left(S_1 \cup S_2 \cup ...\right) + \mu(E - (S_1 \cup S_2 \cup ...) - E') < 2\epsilon.$$

So we have proved the case for when f is defined on a set with finite measure.

The final case is as follows: Let $B(k)$ denote the open ball in $\mathbf{R}^n$ centered at the origin and of radius k. Let $S(k) = S(k) - S(k-1)$ for $k = 1, 2, \cdots$. Let f_k denote the function f restricted to the set $S(k)$. By our version of Lusin's theorem there exists a closed subset E_k of $S(k)$ and a continuous real valued function g_k on $\mathbf{R}^n$ such that $g_k = f_k$ on E_k and $\mu(S(k) - E_k) < \frac{\epsilon}{2^k}$.

Consider a pair of sets E_k and E_{k+1}. These sets are compact and disjoint and so the distance between their boundaries is positive. Define the real valued function $h_k : \mathbf{R}^n$ to take the value $t,\ g_k(x_k) + (1-t)g_{k+1}(x_{k+1})$ at any point that can be expressed as $tx_k + (1-t)x_{k+1}$ where x_k is a boundary point of E_k, x_{k+1} is a boundary point of E_{k+1}, and t is a number in the unit interval.

Define the function $g : \mathbf{R}^n \to \mathbf{R}$ by saying that $f(x) = g_k(x)$ whenever x belongs to $S(k)$ and $f(x) = h_k(x)$ whenever x can be expressed as $tx_k + (1-t)x_{k+1}$ where x_k is a boundary point of E_k, x_{k+1} is a boundary point of E_{k+1}, and t is a number in the unit interval.

This function is clearly continuous. The set of points at which f differs from g, that is , the set $\{x \in \mathbf{R}^n \mid f(x) \neq g(x)\}$, must belong to the disjoint union of sets $\bigcup_{k=1}^{\infty}(B(k) - E_k)$. So by the monotonicity and countable additivity of Lebesgue measure,

$$\mu(\{x \in \mathbf{R}^n \mid f(x) \neq g(x)\}) \leq \mu\left(\bigcup_{k=1}^{\infty}(B(k) - E_k)\right) = \sum_{k=1}^{\infty}\frac{\epsilon}{2^k} = \epsilon.$$

QED

12. THE LEBESGUE INTEGRAL

We have before given a definition of the Lebesgue integral. But that was only to motivate how to measure subsets of $\mathbf{R}^n$. Let's give a proper definition. In fact let's define the Lebesgue integral axiomatically. First, let E be a measurable subset of $\mathbf{R}^n$ that has finite measure. Here are four properties that we want the integral to satisfy.

(1) The integral over E of the function that takes a value of 1 on some subset S of E and zero elsewhere is the measure of S. That is $\int_E \chi_S = \mu(S)$.

(2) The integral is additive. If g is a measurable function and f is a measurable function defined on E, then the integral over E of $g+f$ is the integral over E of g plus the integral over E of f. That is, $\int_E (g+f) = \int_E g + \int_E f$.

(3) The integral is linear. If f is a measurable function defined on E and λ is a real number, then the integral over E of λ times f is equal to λ times the integral over E of f. That is, $\int_E \lambda f = \lambda \int_E f$.

(4) The integral satisfies bounded convergence. If $f_1, f_2, ...$ is a sequence of measurable functions defined on E that converges to a function f and there exists a number N such that $\mid f_1 \mid \leq N, \mid f_2 \mid \leq N, ...$, then the limit of the numbers $\int_E f_1, \int_E f_2, ...$ is equal to the integral over E of f. That is, $\lim_n \int_E f_n = \int_E \lim_n f_n$.

Our motivation for defining the Lebesgue integral was to prove theorems for when we could interchange the limit and integral operators. The bounded convergence property is such a theorem. Note though that we have assumed that the set E has finite measure.

A useful result that we will use to define the Lebesgue integral is that if f is a nonnegative, bounded, and measurable function defined on a subset E of $\mathbf{R}^n$ of finite measure, then the integral over E of f is equal to the $n+1$-dimensional Lebesgue measure of the area under the graph of f. That is,

$$\int_E f = \mu_{n+1}(\{(x,y) \in E \times \mathbf{R} \mid 0 \leq y \leq f(x)\}).$$

We can prove this using the bounded convergence property and the result that any nonnegative, bounded, and measurable function is the limit of simple functions.

We will define the integral of a nonnegative and measurable function by the $n+1$-dimensional Lebesgue measure of the area under its graph.

We will say that a nonnegative and measurable function f is integrable if its integral is finite.

We may as well define the integral of a measurable and nonnegative function f defined on a set E by the supremum of the integrals of all measurable functions defined on subsets E' of E with finite measure that are nonnegative and bounded.

We can also prove a general version of the dominated convergence property for when our function is defined on a set of possibly infinite measure. Get rid of the condition that E is a set of finite measure. Bound sequence by an integrable function g.

13. AXIOMS FOR THE LEBESGUE INTEGRAL

We want to construct a function $\int$ called the Lebesgue integral from the set of measurable functions to $[0, \infty]$. We would like this function to agree with the Riemann integral whenever its argument is a Riemann integrable function. We would also like to know where the Lebesgue integral is continuous. That is, what measurable functions f have the property that the sequence $\int f_1, \int f_2, ...$ converges to $\int f$ whenever $f_1, f_2, ...$ is a sequence of measurable functions converging to f. Here is a list of some properties we would like the Lebesgue integral to have:

(1) It is additive. If f and g are measurable functions then the integral of $f + g$ is equal to the integral of f plus the integral of g. That is,

$$\int (f + g) = \int f + \int g.$$

(2) It is linear. If f is a measurable function and λ is a positive real number then the integral of λf is equal to λ times the integral of f. That is,

$$\int \lambda f = \lambda \int f.$$

(3) If S is a measurable set then the integral of the characteristic function χ_S is the Lebesgue measure of S. That is,

$$\int \chi_S = \mu(S).$$

(4) Bounded convergence theorem. If $f_1, f_2, ...$ is a sequence of uniformly bounded measurable functions converging pointwise to some function f then the sequence $\int f_1, \int f_2, ...$ converges to $\int f$.

Let us first restrict the domain of the function $\int$ to simple functions that are positive on a set of finite measure.

It will turn out that demanding the above four properties for such functions defines the function $\int$.

It also defines the function $\int$ on the set of all measurable functions because our definition of the integral of an arbitrary measurable function will be determined by the limit of simple functions.

It will turn out that this function satisfies the four desired properties on the set of all measurable functions.

Claim 1: Let E be a measurable subset of $\mathbf{R}^n$ with finite measure. The unique real valued function $\int$ on the set of simple functions on E that satisfies conditions (1)-(3) is defined by

$$\int f = c_1\mu(E_1) + c_2\mu(E_2) + ... + c_m\mu(E_m)$$

for any simple function $f : E \to \mathbf{R}$.

Proof: The formula follows from condition (1), (2), and (3). That is condition (1) implies that $\int f$ is equal to the sum of the terms $\int c_i\chi_{E_i}$ from $i = 1$ to $i = m$. Condition (2) implies that each of the integrals $\int c_i\chi_{E_i}$ can be written as $c_i \int \chi_{E_i}$. Condition (3) implies that each of the terms $c_i \int \chi_{E_i}$ can be written as $c_i\mu(E_i)$. We also need to check that condition (4) holds. So let $f_1, f_2, ...$ be a sequence of simple functions converging to f.

Now let's show this formula satisfies the axioms. It is clear that conditions (2) and (3) hold using this formula. Let's show that condition (1) holds. Let f and g be simple functions on E. Then

$$f = c_1\chi_{E_1} + ... + c_m\chi_{E_m} \text{ and } g = b_1\chi_{S_1} + ... + b_k\chi_{S_k}.$$

where the sets $E_1, ..., E_m$ are pairwise disjoint and measurable with union equal to E, and the sets $S_1, ..., S_k$ are pairwise disjoint and measurable with union equal to E. The function $f + g$ is a simple function also. Let $E_{ij} = E_i \cap S_j$ for $i = 1, ..., m$ and $j = 1, ..., k$. These sets are pairwise disjoint and measurable and their union is equal to E. I claim that the simple function $\sum_{i=1}^m \sum_{j=1}^k (c_i + b_j)\chi_{E_{ij}}$ is equal to $f + g$. Consider a point x in E. This point is in exactly one E_i and exactly one S_j and so $f + g$ takes a value of $c_i + b_j$ on this point which is the same value as the function I proposed takes at the point x. Applying the formula for the integral of a simple function which we derived in this proof we get that

$$\int \sum_{i=1}^m \sum_{j=1}^k (c_i + b_j)\chi_{E_{ij}} = \sum_{i=1}^m \sum_{j=1}^k (c_i + b_j)\mu(E_{ij}).$$

We also have that

$$\int f + \int g = c_1\mu(E_1) + ... + c_m\mu(E_m) + b_1\mu(S_1) + ... + b_k\mu(S_k).$$

The right hand side of this expression can be written as

$$\sum_{i=1}^m \sum_{j=1}^k c_i\mu(E_{ij}) + \sum_{i=1}^m \sum_{j=1}^k b_j\mu(E_{ij})$$

which equal $\sum_{i=1}^m \sum_{j=1}^k (c_i + b_j)\mu(E_{ij})$. QED

Claim 2: Let E be a measurable subset of $\mathbf{R}^n$ with finite measure. The unique real valued function $\int$ on the set of nonnegative bounded measurable functions on

E that satisfies conditions (1)-(4) is defined by

$$\int f = \sup \left\{ \int g \mid g \leq f \text{ and } g \text{ a simple function} \right\}.$$

To prove this let's first prove a weaker condition

I'll prove this next time.

Recall the axioms we want the Lebesgue integral so satisfy:

(1) It is additive. If f and g are measurable functions then the integral of $f+g$ is equal to the integral of f plus the integral of g. That is,

$$\int (f+g) = \int f + \int g.$$

(2) It is linear. If f is a measurable function and λ is a positive real number then the integral of λf is equal to λ times the integral of f. That is,

$$\int \lambda f = \lambda \int f.$$

(3) If S is a measurable set then the integral of the characteristic function χ_S is the Lebesgue measure of S. That is,

$$\int \chi_S = \mu(S).$$

(4) Bounded convergence theorem. If $f_1, f_2, ...$ is a sequence of uniformly bounded measurable functions converging pointwise to some function f then the sequence $\int f_1, \int f_2, ...$ converges to $\int f$.

Claim 2: Let E be a measurable subset of $\mathbf{R}^n$ that has finite measure. The unique real valued function $\int$ on the set of bounded nonnegative measurable functions is given by

$$\int f = \sup \left\{ \int g \mid g \leq f \text{ where } g \text{ is a simple function} \right\}$$

Fist, a lemma.

Lemma 1: If g and h are simple functions on $\mathbf{R}^n$ such that $g \leq h$, then $\int g \leq \int h$.

Proof: Since g is a simple function it can be expressed as

$$h(x) = a_1 \chi_{E_1}(x) + ... + a_n \chi_{E_n}(x)$$

where the $E_1, ..., E_n$ are pairwise disjoint measurable subsets of $\mathbf{R}^n$. Since h is a simple function it can be expressed as $h(x) = b_1 \chi_{F_1}(x) + ... + b_n \chi_{F_n}(x)$ where the $F_1, ..., F_n$ are pairwise disjoint measurable subsets of $\mathbf{R}^n$. Let $G_1, ..., G_m$ be the collection of pairwise intersections of the sets $E_1, ..., E_n$, $F_1, ..., F_n$. There are numbers $c_1, ..., c_m$, $d_1, ..., d_n$ such that

$$h(x) = \sum_{i=1}^{m} c_i \chi_{G_i} \text{ and } g(x) = \sum_{i=1}^{n} d_i \chi_{G_i}.$$

SInce $g \leq h$ we have that $c_i \leq d_i$ for each i. Then

$$\int g = \sum_{i=1}^{m} c_i \mu(G_i) \leq \sum_{i=1}^{m} d_i \mu(G_i) \leq \int h.$$

QED

Definition: Let

$$t_- = \sup \left\{ \int g \mid g \leq f \text{ where } g \text{ is a simple function} \right\}$$

and

$$t_+ = \inf \left\{ \int h \mid f \leq h \text{ where } h \text{ is a simple function} \right\}.$$

Second, another lemma:

Lemma 2: $t_- = t_+$.

Proof: By Lemma 1, $t_- \leq t_+$. Define the function h on $\mathbf{R}^n$ by

$$h(x) = \frac{i+1}{n} \text{ whenever } \frac{i}{n} \leq f(x) \leq \frac{i+1}{n}.$$

Define the function g on $\mathbf{R}^n$ by

$$g(x) = \frac{i}{n} \text{ whenever } \frac{i}{n} \leq f(x) \leq \frac{i+1}{n}.$$

Then g and h are simple functions such that $g \leq f$ and $f \leq h$. Then

$$t_+ - t_- \leq \int h - \int g = \int h - g = \frac{\mu(E)}{n}.$$

Since this can be made as small as we like this shows the reverse inequality, that $t_+ \leq t_-$.

QED

Third, anther lemma

Lemma 3: If we let $\int f = t_-$ for each bounded and nonnegative function f on $\mathbf{R}^n$, then $\int$ satisfies condition 4. That is, if $f_1, f_2, ...$ is a uniformly bounded sequence of functions converging pointwise to f then the sequence $\int f_1, \int f_2, ...$ of numbers converges to $\int f$.

Proof: Let k be an integer. By Egorov's theorem there exists a subset E_k of E on which $f_1, f_2, ...$ converges uniformly to f and such that $\mu(E - E_k) < \frac{1}{k}$. We have that

$$\int f = \int \chi_{E_k} f + \int \chi_{E-E_k} f$$

so that

$$\mid \int f - \int \chi_{E_k} f \mid \leq \frac{\mu(E)}{k}.$$

Since the sequence of functions $f_1, f_2, ...$ converges uniformly to the function f on E_k there exists an integer m such that $\mid f(x) - f_n(x) \mid \leq \frac{1}{k}$ whenever $x \in E_k andn \geq m$.

This implies that for $n \geq m$ we have

$$| \int \chi_{E_k} f - \int \chi_{E_k} f_n | \leq \frac{\mu(E)}{k}.$$

Then

$$\int f - \int f_n = \left(\int f - \int \chi_{E_k} f \right) + \left(\int \chi_{E_k} f - \int \chi_{E_k} f_n \right) + \left(\int \chi_{E_k} f_n - \int f_n \right)$$

so that for $n \geq m$

$$| \int f - \int f_n | \leq \frac{\mu(E)}{k} + \frac{\mu(E)}{k} + \frac{M}{k}$$

where M is the bound for the sequence $f_1, f_2, ...$ The right hand side of this expression can be made arbitrarily small by choosing k big enough. This implies that the sequence $\int f_1, \int f_2, ...$ of numbers converges to $\int f$.

QED

Now let's prove the claim.

Proof of Claim 2: The formula

$$\int f = \sup\{ \int g \mid g \leq f \text{ where } g \text{ is a simple function} \}$$

(i.e. $\int f = t_-$)is the only possibility because f is equal to the limit of the functions g_n defined by

$$g_n(x) = \frac{i}{n} \text{ whenever } \frac{i}{n} \leq f(x) \leq \frac{i+1}{n}$$

and the limit of the sequence $\int g_1, \int g_2, ...$ is equal to t_-. So by condition 4 $\int f = t_-$. If the four conditions hold for this choice of $\int$ then we have proved claim 2. Let's check each of them.

Condition 4: We showed in Lemma 3 that condition 4 holds.

Condition 2:

$$\int \lambda f = \sup\{ \int \lambda g \mid g \leq f \text{ where } g \text{ is a simple function} \}$$

$$= \sup\{ \lambda \int g \mid g \leq f \text{ where } g \text{ is a simple function} \}$$

$$= \lambda \sup\{ \int g \mid g \leq f \text{ where } g \text{ is a simple function} \} = \lambda \int f.$$

Condition 3: Obviously$\int \chi_E = \mu(E)$.

Condition 1: Let f_n be the simple function defined by $f_n(x) = \frac{i}{n}$ whenever $\frac{i}{n} \leq f(x) \leq \frac{i+1}{n}$. Let g_n be the simple function defined by $g_n(x) = \frac{i}{n}$ whenever

$$\frac{i}{n} \leq g(x) \leq \frac{i+1}{n}$$

. Then the sequence $f_1, f_2, ...$ is uniformly bounded because g is bounded and converges pointwise to f and the sequence of functions $g_1, g_2, ...$ is uniformly bounded because f is bounded and converges pointwise to f. Then $f_n + g_n$ is a uniformly bounded sequence that converges pointwise to $f + g$. By condition 4 and the continuity of addition we have

$$\int f + g = \lim_n \int f_n + g_n = \lim_n \int f_n + \lim_n \int g_n = \int f + \int g.$$

QED.

We have shown that the unique function $\int$ from the set of functions (that are nonnegative, bounded, and measurable and take positive values on a set of finite measure) to the real numbers satisfying our four conditions is given by the equation

$$\int f = \sup \left\{ \int g \mid g \leq f \text{ where } g \text{ is a simple function} \right\}.$$

Now we would like to enlarge the domain of our function $\int$ to the set of functions that are nonnegative and measurable. That is, we are getting rid of the restriction of boundedness and being positive only on a set of finite measure.

One way to proceed is to continue to work with the definition $\int$ given above. But it will be more useful to define $\int f$ to be the Lebesgue integral of the area under the graph of f. That is,

$$\int f = \mu(\{(x, y) \in \mathbf{R}^{n+1} \mid 0 \leq y \leq f(x)\}).$$

It turns out that this definition of $\int$ agrees with the definition of $\int f$ as $\sup\{\int g \mid g \leq f, g \text{ simple}\}$ when f is bounded and positive only on a set of finite measure. We can show this by showing that this definition satisfies the conditions 1-4 for functions f that are bounded and positive only on a set of finite measure. We can also show it directly as follows.

Lemma 1: Let $f : \mathbf{R}^n \to \mathbf{R}$ be a bounded, nonnegative, measurable function that is positive only on a set of finite measure. Then

$$\sup \left\{ \int g \mid g \leq f \text{ where } g \text{ is a simple function} \right\}$$
$$= \mu(\{(x, y) \in \mathbf{R}^{n+1} \mid 0 \leq y \leq f(x)\})$$

Proof: Suppose we could prove the result when f is a simple function. Then if f is not a simple function we can define the function f_n by $f_n(x) = \frac{i}{n}$ whenever $\frac{i}{n} \leq f(x) \leq \frac{i+1}{n}$. This is a sequence of simple functions that is uniformly bounded and converges to f. So by condition 4 we know that $\int f = \lim_n \int f_n$. By our knowledge that the result holds for simple function we have that the right hand side of this is equal to $\lim_n \mu(\{(x, y) \in \mathbf{R}^{n+1} \mid 0 \leq y \leq f_n(x)\})$ and by continuity of measure we have that this is equal to $\mu(\{(x, y) \in \mathbf{R}^{n+1} \mid 0 \leq y \leq f(x)\})$, the thing we want it to be equal to. So let's suppose that f is a simple function. Because f is a simple function we can write $f = c_1\chi_{E_1} + c_2\chi_{E_2} + ... + c_n\chi_{E_n}$. We have that $\int f = c_1\mu(E_1) + c_2\mu(E_2) + c_n\mu(E_n)$. We can write $\{(x, y) \in \mathbf{R}^{n+1} \mid 0 \leq f(x) \leq y\}$ as the disjoint union of the sets $E_i \cup [0, c_i]$ and so its measure is also $c_1\mu(E_1) + c_2\mu(E_2) + ... + c_n\mu(E_n)$.

QED

Let's make this out new definition of the integral.

Definition 1: Let the Lebesgue integral be the function $\int$ from the set of nonnegative and measurable functions to $[0, \infty]$ defined by

$$\int f = \mu(\{(x, y) \in \mathbf{R}^{n+1} \mid 0 \leq y \leq f(x)\}).$$

This definition says that the Lebesgue integral of a nonnegative and measurable function is the area under its graph. For bounded functions that are positive only on a set of finite measure that area is finite. The area may be infinite if we relax either of these assumptions. For example consider he function $f : [0, 1] \to [0, \infty]$ defined by $f(x) = \frac{1}{x}$. The area under the graph of this function is $\lim_{x\to 0} -\ln(x) = \infty$. This function is positive only on a set of finite measure but it is not bounded. Now consider the function $f : [0, 1] \to [0, \infty]$ defined by $f(x) = \frac{1}{\sqrt{x}}$. The area under the graph of this function is 2. This function is not bounded but it still has a finite Lebesgue integral. Now consider the function $f : (-\infty, 0) \to [0, 1]$ defined by $f(x) = e^x$. The area under the graph of this function is 1 and this function is positive on a set of infinite measure. Clearly there are also nonnegative measurable functions that are positive on a set of infinite measure and the area under their graphs is infinite.

It turns out that for this more general definition of the Lebesgue integral our four conditions will continue to hold. Note that it is not much to say that the fourth condition will hold because if $f_1, f_2, ...$ is a uniformly bounded sequence of measurable functions converging pointwise to a function f then it must be that f is bounded. The thing we have to check for our new definition of the Lebesgue integral is that it might be that the sequence of functions converges to a function that is positive on a set with infinite measure.

Claim 1: The Lebesgue integral satisfies conditions 1-3.

Lemma 1: The Lebesgue integral satisfies condition 4: If $f_1 \leq f_2 \leq ...$ is a sequence of functions converging to a nonnegative measurable function f, then $\int f$ is the limit of the sequence $\int f_1, \int f_2, ...$

Proof: Let $S_i = \{(x, y) \in \mathbf{R}^{n+1} \mid 0 \leq y \leq f_i(x)\}$ and $S = \{(x, y) \in \mathbf{R}^{n+1} \mid 0 \leq y \leq f(x)\}$. The sequence $S_1, S_2, ...$ is an increasing sequence of sets. By the continuity of Lebesgue measure $\lim_n \mu(S_n) = \mu(S)$. That is, $\lim \int f_n = \int f$.

QED

This seems more general than condition 4 in that the sequence of functions does not need to be uniformly bounded. However it is less general in that the sequence of functions needs to be nondecreasing.

Lemma 2: The Lebesgue integral satisfies condition 2. That is, if λ is a positive real number then $\int \lambda f = \lambda \int f$.

Proof: We have proved previously that this holds when f is also bounded and only positive on a set of finite measure. Consider the function f_i defined by $f_i(x) = f(x)$ if $x \in [-i, i]$ and $f(x) \leq i$, $f_i(x) = 0$ if $x \notin [-i, i]$, and $f_i(x) = i$ if $x \in [-i, i]$ and $f(x) > i$. Each function f_i is a nonnegative, bounded, measurable function, that is positive only on a set of finite measure – so we know that the lemma holds for each f_i. We also have that $f_1 \leq f_2 \leq ...$ is a nondecreasing sequence of functions that converges to f. By Lemma 1 we have that $\int f_i$ converges to $\int f$. Therefore by continuity of multiplication

$$\int \lambda f = \lim \int \lambda f_i = \lim \lambda \int f_i = \lambda \int f.$$

QED

Lemma 3: The Lebesgue integral satisfies condition 3. That is, if S is a measurable subset of $\mathbf{R}^n$, then $\int \chi_S = \mu(S)$.

Proof: It turns out that if E is a measurable subset of $\mathbf{R}^n$ and E' is a measurable subset of $\mathbf{R}^m$ then $\mu_{n+m}(E \times E') = \mu_n(E) \times \mu(E')$. This fact is a problem shown on one of the problem sets for this course. Applying this result gives:

$$\int \chi_S = \mu_{n+1}(S \times [0,1]) = \mu_n(S)\mu_1([0,1]) = \mu_n(S).$$

QED

Lemma 4: The Lebesgue integral satisfies condition 1. That is, if f and g are nonnegative and measurable functions, then $\int f + g = \int f + \int g$.

Proof: We know that this is true when both f and g are nonnegative and positive only on a set of finite measure. Consider the function f_i defined by $f_i(x) = f(x)$ if $x \in [-i, i]$ and $f(x) \leq i$, $f_i(x) = 0$ if $x \notin [-i, i]$, and $f_i(x) = i$ if $x \in [-i, i]$ and $f(x) > i$. Consider the function g_i defined by $g_i(x) = g(x)$ if $x \in [-i, i]$ and $g(x) \leq i$, $g_i(x) = 0$ if $x \notin [-i, i]$, and $g_i(x) = i$ if $x \in [-i, i]$ and $g(x) > i$. Each f_i, g_i, and $f_i + g_i$ are nonnegative and positive only on a set of finite measure. The sequence $f_1, f_2, ...$ is nondecreasing and converges pointwise to f. The sequence $g_1, g_2, ...$ is nondecreasing and converges pointwise to g. The sequence $f_1 + g_1, f_2 + g_2, ...$ is nondecreasing and converges pointwise to $f + g$. Therefore, by Lemma 1 and the continuity of measure

$$\int f + g = \lim \int f_i + g_i = \lim \int f_i + \lim \int g_i = \int f + \int g.$$

QED.

That's nice. Now we know that our definition of the Lebesgue integral is the unique definition for measurable functions that are nonnegative, bounded, and positive only on a set of finite measure that satisfies conditions 1-4 and we also know that it satisfies conditions 1-3 and a variant of condition 4 on the set of measurable nonnegative functions.

Next time we'll define the integral for all measurable functions. You can probably guess the correct definition.

Let $f : \mathbf{R}^n \to \mathbf{R}$ be a measurable function. What is the integral of f? We know how to integrate nonnegative measurable functions and we know that the integral for nonnegative functions is additive and linear. So let's split up f in the following way. Define f_+ by $f_+(x) = f(x)$ if $f(x) > 0$ and $f(x) = 0$ otherwise. Define f_- by $f_-(x) = -f(x)$ if $f(x) < 0$ and $f(x) = 0$ otherwise. Then $f = f_+ + (-f_-)$. Then using the properties of the Lebesgue integral we have that

$$\int f = \int f_+ - \int f_-.$$

That is, we already knew how to integrate measurable functions. We just did not know it. There is one issue with this definition what if we get $\infty - \infty$ as $\int f$? That is, suppose the function we were trying to integrate was $sin(x)$. Our integral would then be the sum of infinitely many ones and infinitely many negative ones. This can be made to sum to any integer. So we would like to rule out this case. We will do so with the following definition:

Definition: Let f be a measurable function. The function f is called integrable if $\int \mid f \mid$ is finite.

Note that when a function is integrable then both $\int f_+ and \int f_-$ are finite so we won't get into the problem of having to subtract infinity from infinity. Also note that some nonnegative measurable functions are not integrable.

We will sometimes want to restrict ourselves to integrating integrable functions. Do the properties we want the Lebesgue integral to satisfy hold for these functions?

Lemma 1: The Lebesgue integral is linear (satisfies condition 2) for integrable functions. That is, if $\lambda > 0$ and f is a measurable function then $\int \lambda f = \lambda \int f$.

Proof: We have

$$\int \lambda f = \int (\lambda f)_+ - \int (\lambda f)_- = \lambda (\int f_+ - \int f_- \lambda \int f.$$

This can be generalised so that λ is any real number. It is clearly true if $\lambda = 0$. If $\lambda < 0$ then $(\lambda f)_+ = -\lambda f_- and (\lambda f)_- = -\lambda f_+$ and we get the same result.

QED

Lemma 2: The Lebesgue integral satisfies condition 3 for integrable functions. That is, if S is a measurable set then $\int \chi_S = \mu(S)$.

Proof: χ_S is a nonnegative measurable function and we showed this property before for nonnegative measurable functions.

QED

Lemma 3: The Lebesgue integral satisfies condition 4 for integrable functions. That is, if $f_1, f_2, ...$ is a sequence of measurable functions that is uniformly bounded converging to a function f then $\int f$ is the limit of $\int f_1, \int f_2, ...$

Proof: The sequence $(f_1)_+, (f_2)_+, ...$ is a sequence of nonnegative measurable functions that is uniformly bounded. The sequence $(f_1)_-, (f_2)_-, ...$ is also a sequence of nonnegative measurable functions that is uniformly bounded. We know that condition 4 holds for nonnegative measurable functions. Therefore, since addition is continuous,

$$\lim \int f_n = \lim \int (f_n)_+ - \lim \int (f_n)_- = \int f_+ - \int f_- = \int f.$$

QED

Lemma 4: Let f be a nonnegative measurable function. Let E and E' be disjoint measurable sets. Then

$$\int \chi_{E\cup E'} f = \int \chi_E f + \int \chi_{E'} f.$$

Proof: Because E and E' are disjoint we have that $\chi_{E\cup E'} = \chi_E + \chi_{E'}$. Since we have proved additivity for nonnegative functions we have that

$$\int \chi_{E\cup E'} f = \int (\chi_E + \chi_{E'}) f = \int \chi_E f + \int \chi_{E'} f.$$

QED

Lemma 5: The Lebesgue integral satisfies condition 1 (it is additive). If f and g are integrable functions then $\int f + g = \int f + \int g$.

Proof: We can split up the domain of $f + g$ into four different regions. Let E_1 denote the points where both f and g are positive. Let E_2 denote the points where both f and g are negative. Let E_3 denote the points not in E_1 or E_2 where $f + g$ is nonnegative. Let E_4 denote the points not in E_1 or E_2 where $f + g$ is negative.

By Lemma 4 $\int f + g = \sum_{i=1}^{4} \int \chi_{E_i}(f + g)$. If we can prove for each $i = 1, 2, 3, 4$ that $\int \chi_{E_i}(f + g) = \int \chi_{E_i} f + \int \chi_{E_i} g)$ then $\sum_{i=1}^{4} \int \chi_{E_i}(f + g)$ can be written as $\sum_{i=1}^{4} \int \chi_{E_i} f + \int \chi_{E_i} g) = \int f + \int g$ so we will be done.

We already proved this for i=1 because then both f and g are nonnegative. For $i = 1$ (i.e. both f and gare nonpositive) we can have that $\int f+g = -\int(-f)+(-g) = \int f + \int g$. Now consider $i = 3$. The set E_3 is the union of two regions: the set of points where $f(x)$ is nonnegative and larger than $-g(x)$ (because $g(x)$ is necessarily negative, otherwise his point x would belong to E_1) and the se of points where

$g(x)$ is nonnegative and larger than $-f(x)$. For the first of these regions we then have that $f+g = f - (-g)$ where both f and $-g$ are nonnegative functions. So $\int(-g) + \int f + g = \int f$ which implies that $\int f - \int(-g) = \int f + g$. By Lemma 1 the left hand side of this is $\int f + \int g$. We can do a similar trick on the other region.

Now consider $i = 4$. The set E_4 is the union of two regions also: the set of points where $f(x)$ is nonnegative and smaller than $-g(x)$ and the set of points where $g(x)$ is nonnegative and smaller than $-f(x)$. For the first of these regions we have that $(-f)$ is a nonpositive function and so on this region $\int(-f) + \int f + g = \int g$. This implies that $\int f + g = \int g - \int(-f)$. By Lemma 1 the right hand side of this expression is equal to $\int f + \int g$. A similar argument works for the other region.

QED

We now have the following.

Claim 1: The Lebesgue integral on the set of integrable functions satisfies conditions 1-4.

Proof: Lemma 1,2,3,5.

QED

We have achieved our goal of defining a new integral. Our definition is: if f is a nonnegative measurable function the $\int f = \mu_(\{(x, y \in \mathbf{R}^{n+1} \mid 0 \leq y \leq f(x))\})$. For an integrable function $\int f = \int f_+ - \int f_-$. This is the unique definition of the integral that satisfies conditions 1-4.

14. LEBESGUE'S DOMINATED CONVERGENCE THEOREM

We can think of the Lebesgue integral as a function from the set of integrable functions to the real numbers. The integral of any nonnegative measurable function is $\mu(\{(x,y) \in \mathbf{R}^{n+1} \mid 0 \leq y \leq f(x)\})$ and the integral of any integrable function f is defined as $\int f = \int f_+ - \int f_-$. This is the only function on the set of integrable functions that satisfies our four conditions. If you recall one of our motivations for defining the Lebesgue integral was to find conditions when the integral and limit operators could be interchanged. Condition 4 gave us one rule for this: if $f_1, f_2, ...$ is a sequence of uniformly bounded measurable functions (defined on a set of finite measure) converging pointwise to a function f, then $\lim \int f_n = \int f$. This is called the bounded convergence theorem. The hypothesis of this result can be relaxed to that the sequence of functions $f_1, f_2, ...$ need not be uniformly bounded by a number but can instead be uniformly bounded by an integrable function. This is called Lebesgue's Dominated Convergence Theorem.

Lebesgue's Dominated Convergence Theorem: Let $f_1, f_2, ...$ be a sequence of measurable functions converging pointwise to a function f. If g is an integrable function such that $\mid f_n \mid \leq g$ for all n, then f is integrable and $\lim \int f_n = \int f$.

Proof: We may assume that g is nonnegative (if not just use its absolute value). We can also assume that g is positive because each f_i is zero when g is zero. Let's also assume that the support of g is E. We have $\mid f \mid \leq g$. By the monotonicity of the integral this implies that f is integrable. Likewise, each f_i is integrable. Consider the set

$$S_n = \{x \in \mathbf{R}^n \mid \frac{1}{n} \leq g(x) \leq n\}.$$

The sequence of functions $f_1, f_2, \ldots$ is uniformly bounded on this set. What is the measure of S_n? We have that $\frac{1}{n}\chi_{S_n} \leq g$. This implies that $\mu(S_n) \leq n \int g$ and this is finite because g is integrable. The bounded convergence theorem can now be applied so that for each n $\lim_i \int \chi_{S_n} f_i = \int \chi_{S_n} f$.

Let T_n denote the set $E - S_n$. Notice that $\int f_i = \int \chi_{S_n} f_i + \int \chi_{T_n} f_i$. Therefore $\mid \int f - \int f_i \mid$ is no more than

$$\mid \int \chi_{S_n} f - \int \chi_{S_n} f_i \mid + \mid \int \chi_{T_n} f \mid + \mid \int \chi_{T_n} f_i \mid$$

Note that $T_1, T_2, \ldots$ is a decreasing sequence of sets such that $\bigcap_{n \geq 1} T_n = \emptyset$. Since we have assumed that g is positive there is some n such that x in T_n implies $g(x) \geq n$. For this n we have $n\mu(T_n) \leq \int g$ which is finite. Therefore $\mu(T_n)$ is finite. By the continuity of measure $\lim \mu(T_n) = 0$. Let $\epsilon > 0$. Then there exists an integer m such that $n \geq m$ implies for each i that $\mid \int \chi_{T_n} f - \int \chi_{T_n} f_i \mid \leq \frac{\epsilon}{2}$. For this m there exists i_m such that $i \geq i_m$ implies $\mid \int \chi_{S_n} f - \int \chi_{S_n} f_i \mid \leq \frac{\epsilon}{2}$.

QED

15. THE LEBESGUE SPACE

Let $E \subseteq \mathbf{R}^n$ be a measurable set. Define the Lebesgue space on E as the set

$$L^1(E) = \{f : E \to \mathbf{R} \mid f \text{ is integrable}\} \text{ mod functions that are the same almost everywhere.}$$

Recall the definition of an integrable function: the function $f : E \to \mathbf{R}$ is integrable if it is measurable and $\int \mid f \mid < \infty$. If f is an integrable function then we think of it as an element of $L^1(E)$ but what we really mean is that the class of functions that are equal to f almost everywhere is an element of $L^1(E)$.

Notice that $L^1(E)$ is a vector space: the sum of two integrable functions is integrable and a number multiplied by an integrable function is integrable. The Lebesgue integral is the unique real valued function on $L^1(E)$ satisfying our four axioms (additivity, linearity, the normalisation condition, and bounded convergence). It turns out that $L^1(E)$ does not have a finite basis. That is, it is an infinite dimensional vector space. We will want to talk about the basis of $L^1(E)$. But to do this we need to be able to define infinite sums of elements of $L^1(E)$and to do this we need a notion of convergence. That is, we need to give $L^1(E)$ a topology.

Definition: A sequence $f_1, f_2, ...inL^1(E)$ converges to a function f in $L^1(E)$if the sequence $\int \mid f_1 - f \mid, \int \mid f_2 - f \mid, ...$ converges to zero.

If f and g are integrable functions then let $d(f,g)$ denote $\int \mid f - g \mid$. There is a notion called a metric space which consists of a set X and a function d from $X \times X$ to the nonnegative real numbers. The idea is that $d(x,y)$ measures the distance between the points x and y in X. To be a metric space the function d must satisfy the following properties. (1) The distance between points x and y is the same as the distance between y and x, that is, for all x, y in X we have $d(x,y) = d(y,x)$. (2) The distance between two points if zero if and only if the two points are equal to one another, that is, $d(x,y) = 0$ if and only if $x = y$. (3) The distance from a point x to a point z is no more than the distance between x and a point y plus the distance between a y and z, that is, $d(x,z) \leq d(x,y) + d(y,z)$. This last condition is called 'the triangle inequality' because of its relation to the idea that any side of a triangle has length no more than the sum of the other two sides.

It turns out that the set $L^1(E)$ with the function $d(f,g) = \int \mid f - g \mid$ is a metric space. Let's prove this. Condition (1) is true because $\mid f - g \mid = \mid g - f \mid$ for all f and g in $L^1(E)$. For condition (2) we have that if $f = g$, then $d(f,g) = 0$ because the integral of the zero function is zero. It is possible to show that if the Lebesgue integral of a function is zero then that function is zero almost everywhere. Thus if $d(f,g) = 0$ then $f - g$ is equal to 0 almost everywhere and so f is equal to g almost everywhere. Note that her it is important that we think of elements of $L^1(E)$ as the class of functions that are the same almost everywhere. For condition (3) let f, g, and h belong to $L^1(E)$. Then $\mid f - g \mid \leq \mid f - h \mid + \mid h - g \mid$ and so my the linearity and monotonicity of the Lebesgue integral we have that $\int \mid f - g \mid \leq \int \mid f - h \mid + \int \mid h - g \mid$, that is, $d(f,g) \leq d(f,h) + d(h,g)$. So, we have shown that $L^1(E)$ with the function $d(f,g) = \int \mid f - g \mid$ is a metric space.

Why do we care about showing this? Well, there is a theory for metric spaces. Now that we know that $L^1(E)$ with the function $d(f,g) = \int \mid f - g \mid$ is a metric space we can apply to it this theory.

A notion that is related to a metric space is a normed vector space. A normed space is a vector space V and a function $\| \cdot \|$ from V to the nonnegative real numbers. The idea is given a point v in V that $\|v\|$ is the length of the vector v. To be a normed vector space the function $\| \cdot \|$ must satisfy the following properties. (1) The length of a vector v is zero if and only if v is the zero vector, that is, for all v in V we have $\|v\| = 0$ if and only if $v = 0$. (2) If α is a real number and v is a vector in V, then the length of the vector αv is the absolute value of α multiplied by the length of v, that is, for all $\alpha \in \mathbf{R}$ and for all $v \in V$ we have $\|\alpha v\| = \mid \alpha \mid \|v\|$. (3) The length of a vector $v + w$ is no more than the length of the vector v plus the length of the vector w, that is, for all v, w in V we have $\|v + w\| \leq \|v\| + \|w\|$. This last condition has the same intuition as the triangle inequality.

Notice that $L^1(E)$ with $\|f\| = \int |f|$ is a normed vector space. From our knowledge that $L^1(E)$ with the function $d(f,g) = \int \mid f - g \mid$ is a metric space it is clear that conditions (1) and (3) hold. For condition (2), let α be a real number and let f be an element of $L^1(E)$. Because the Lebesgue integral is linear we have $\|\alpha f\| = \int |\alpha f| = |\alpha| \int |f| = |\alpha| \|f\|$. So $L^1(E)$ with $\|f\| = \int |f|$ is a normed vector space.

Again, why do we care? I think we care for a similar reason that we care that $L^1(E)$ with the function $d(f,g) = \int \mid f - g \mid$ is a metric space. That is, there is a theory about normed vector spaces so by knowing that $L^1(E)$ with $\|f\| = \int |f|$ is a normed vector space we can apply this theory to it.

Also note that given a normed vector space V with the function $\| \cdot \|$ we get a metric space by defining $d(v, w)$ to be $\|v - w\|$. Indeed, we have already shown this to be the case for $L^1(E)$ with the norm $\|f\| = \int |f|$.

A desirable property of a normed vector space is that it is complete. That is, there are no sequences $v_1, v_2, ...$ that seem to be converging (in other words, are the points eventually get arbitrarily close to one another) but there is no point in the vector space that these points correspond to. For example the rational numbers are not complete because there is a sequence of rational numbers that converges to the square root of 2. We will call a normed vector space that is complete a Banach space.

It turns out that $L^1(E)$ with the norm $\|f\| = \int |f|$ is a Banach space. That is, if $f_1, f_2, ...$ is a sequence in $L^1(E)$ such that for each number $\epsilon > 0$ there exists an integer $k > 0$ such that $n, m \geq k$ implies that $\|f_n - f_m\| \leq \epsilon$ (such a sequence is called a Cauchy sequence), then there exists an element f of $L^1(E)$ such that the sequence $f_1, f_2, ...$ converges to f in the sense that $\lim \|f - f_n\| = 0$.

Theorem 1: The set $L^1(E)$ with the norm $\|f\| = \int |f|$ is a Banach space.

To prove this theorem let's first talk about the relationship between pointwise convergence and convergence using the norm $\int |f|$. Let $f_1, f_2, ...$ be a sequence in $L^1(E)$ and let f be an element of $L^1(E)$. Here are two natural questions.

Question 1: If $\lim \|f_n - f\| = 0$, then what can we say about $\lim |f_n(x) - f(x)|$?

Question 2: If for each x in E we have $\lim |f_n(x) - f(x)| = 0$, then what can we say about $\lim \|f_n - f\|$?

Consider the following example. Let $f_n = \chi_{S_n}$ where $S_1 = [0, 1]$, $S_2 = [0, \frac{1}{2}]$, $S_3 = [\frac{1}{2}, 1]$, $S_4 = [0, \frac{1}{3}]$, $S_5 = [\frac{1}{3}, \frac{2}{3}]$, $S_6 = [\frac{2}{3}, 1]$, and so on. This sequence does not converge pointwise at any point $x \in [0, 1]$. But we do have that $\lim \|f_n - f\| = 0$ because $\int |f_n| = \mu(S_n)$ and $\mu(S_n)$ converges to zero. This example shows that it is not true that convergence in the $L^1(E)$ norm implies pointwise convergence.

What about question 2. Consider this example. Let $f_n = \chi_{[n,n+1]}$. The sequence of functions $f_1, f_2, ...$ converges pointwise to the zero function because for any x we have that x is less than some integer m and so $|f_n(x)| = 0$ for all n at least equal to m. But we don't have that $f_1, f_2, ...$ converges to the zero function in the $L^1(E)$ norm because $\int |f_n| = 1$ for all n. This example shows that it is not true that pointwise convergence implies convergence in $L^1(E)$.

It turns out that the following is true: a sequence $x_1, x_2, ...$ converges to a point x if and only if when $y_1 = x_2 - x_1, y_2, x_3 - x_2, ...$ the sum $y_1 + y_1 + ...$ converges to x. A sum $z_1 + z_2 + ...$ is said to converge absolutely if the sum of the norm of its terms converges.

Definition: A sequence $x_1, x_2, ...$ converges quickly if the sum $y_1 + y_2 + ...$ where $y_1 = x_2 - x_1, y_2 = x_3 - x_2, ...$ converges absolutely.

Thus a sequence $f_1, f_2, ...$ in $L^1(E)$ converges quickly if the sum $\|f_2 - f_1\| + \|f_3 - f_2\| + ...$ converges.

It turns out a normed vector space is a Banach space if and only if each sequence that converges quickly converges in the space.

How can we use this to prove that $L^1(E)$ is a Banach space?

Well, let $f_1, f_2, ...$ be a Cauchy sequence in $L^1(E)$. Then we can find a subsequence of $f_1, f_2, ...$ that converges quickly. To do this, let k_i be the integer such that $n, m \geq k_1$ implies $\|f_n - f_m\| \leq \frac{1}{2^i}$, and do this for $i = 1, 2, ...$ Then $f_{k_1}, f_{k_2}, ...$ is a subsequence that converges quickly. It can be shown that this implies that there is a function f such that $f_{k_1}, f_{k_2}, ...$ converges almost everywhere to f. It is then possible to show that $f_1, f_2, ...$ converges almost everywhere to f. Doing this will show that $L^1(E)$ is a Banach space.

16. TONELLI AND FUBINI'S THEOREM

Given a measurable function $f : \mathbf{R}^m \times \mathbf{R}^n \to \mathbf{R}$ what is the relationship between $\int f(x,y)d(x,y)$, $\int \left(\int f(x,y)dx\right) dy$, and $\int \left(\int f(x,y)dy\right) dx$?

Let E be a measurable subset of $\mathbf{R}^m \times \mathbf{R}^n$. For each $x \in \mathbf{R}^m let E_x$ denote the set $\{y \in \mathbf{R}^n \mid (x,y) \in E\}$. Let's try to understand the relationship between the measure of E and the measure of the sets E_x.

Recall the definition of a measurable set: A set $E \subseteq \mathbf{R}^n$ is measurable if for all sets $F \subseteq \mathbf{R}^n$ the outer measure of F is equal to the outer measure of the points that are in both E and F plus the outer measure of the points that are in F but not in E. It would be nice if each set E_x (or at least almost all) thought of as a subset of $\mathbf{R}^n$ were measurable. Now define the function $f_E : \mathbf{R}^n \to [0, \infty]$ by $f_E(x) = \mu(E_x)$ if E_x is measurable and $f_E(x) = 0$ if E_x is not measurable. It would be nice if this were a measurable function. It would also be nice if the integral of this function was the measure of E, that is, $\int f_E = \mu(E)$.

Case 1) Let's start by assuming that $E \subseteq \mathbf{R}^m \times \mathbf{R}^n$ is an open box. Then E is equal to the cartesian product of a box $B_1 \subseteq \mathbf{R}^m$ and an open box $B_2 \subseteq \mathbf{R}^n$. That is, $E = B_1 \times B_2$. In this case $E_x = B_2$ if x is in B_1 and $E_x = \emptyset$ if x is not in B_1. We know that open sets are measurable so we know that E_x is a measurable set for all x. We also have that the f_E is a measurable function because $f_E = \mu(E_x)\chi_{B_1}$ and B_1 is a measurable set. Finally, we have that

$$\int f_E = \int \mu(B_2)\chi_{B_1} = \mu(B_2)\mu(B_1) = \mu(B_1 \times B_2).$$

Case 2) Now let's assume that E is the union of two open boxes E^1 and E^2. That is, $E = E^1 \cup E^2 = (B_1 \times B_2) \cup (B_3 \times B_4)$ where B_1, B_3 are open boxes of $\mathbf{R}^m$ and B_2, B_4 are open boxes of $\mathbf{R}^n$. Then we have that $E_x = E_x^1 \cup E_x^2$ which by Case 1 is the union of two measurable sets and is therefore measurable. Similarly,

$$f_E = f_{E^1} + f_{E^2} - f_{E^1 \cap E^2} = \mu(B_2)\chi_{B_1} + \mu(B_4)\chi_{B_3} - \mu(B_2 \cap B_4)\chi_{B_1 \cap B_3}$$

is a measurable function and

$$\int f_E = \mu(B_2)\mu(B_1) + \mu(B_4)\mu(B_3) - \mu(B_2 \cap B_4)\mu(B_1 \cap B_3) = \mu(E).$$

Continuing in this way we can show that the properties hold if E is the union of n open boxes. Can we also show this result for the limit case? Putting this slightly differently, given an open set E do the three properties

(1) The set E_x is measurable for almost every x, (2) The function $f_E : \mathbf{R}^m \to [0, \infty]$ is measurable, and (3) $\mu(E) = \int f_E$

hold. We can write E as the countable union of open boxes. Property (1) holds because E_x is either the empty set or a finite or countable union of open boxes. Now let E_k denote the union of the first k open boxes that E is made from. For each x we have that $E_{1,x}, E_{2,x}, ...$ is a nondecreasing sequence of sets whose measures converge to the measure of $\mu(E_x)$. Therefore $f_{E_1}, f_{E_2}, ...$ is a nondecreasing sequence of measurable functions whose limit is f. We have previously shown that the limit of measurable functions is also a measurable function. Therefore f is a measurable function. By the continuity of measure we know that the sequence $\mu(E_1), \mu(E_2), ...$ converges to $\mu(E)$. Therefore we know that $\int f_{E_1}, \int f_{E_2}, ...$ converges to $\mu(E)$. By the monotone convergence theorem we also know that $\int f_{E_1}, \int f_{E_2}, ...$ converges to $\int f$. Therefore $\int f = \mu(E)$. So we have shown that the three properties hold when E is any open set.

What can we say when E has measure zero? To say that E has measure zero means that for each number $\epsilon > 0$ there exists open boxes $B_1, B_2, ...$ whose union contains E and such that the sum of the volumes of these open boxes is no more than ϵ. Let $B_1, B_2, ...$ be such a sequence of open boxes for $\epsilon = \frac{1}{n}$. Let $F_n = B_1 \cup B_2 \cup ....$ We know that the properties (1),(2), and (3) hold for F.

We have that $f_E(x) = \mu(E_x) \le \mu(F_{n,x}) = f_{F_n}(x)$. We also have that $\inf f_{F_n} = \mu(F_n) \le \frac{1}{n}$. What is the measure of the set on which f_{F_n} takes values more than $\frac{1}{k}$. Call this set S_k. It must be that $\frac{1}{k}$ times the measure of this set is no more than $\frac{1}{n}$, that is, $\mu(S_k) \le \frac{k}{n}$. For each k we can choose n to make this as small as we like. Then we have that for each k the set on which f_{F_n} takes a value larger than $\frac{1}{k}$ can be made as small as we like. Therefore the set on which f_E takes a positive value must have measure zero. This shows property (1) because any set that has measure zero is measurable. It shows property (2) because the zero function in a measurable function. And it shows property (3) because the integral of the zero function is zero, the same as the measure of E. So we have shown that he three properties hold when E is a set of zero measure.

Now let's try to show the three properties are true when E is any measurable set. We have shown previously that any measurable set can be written as the countable intersection of open sets minus a set of measure zero. We have shown that the three properties hold when E is an open set or a set with zero measure. We will use this.

Our first goal is to show that the three properties hold when E is the countable intersection of open sets. Let's also suppose that E is bounded. Because the finite intersection of open sets is an open set we can think about our countable union of open sets as a decreasing sequence of open sets $U_0 \subseteq U_1 \subseteq ....$ We know the three properties hold for each U_i. We also know that $f_E \le f_{U_i}$ for each i. And we know that $E = \lim_i U_i$ and so $E_x = \lim_i U_{i,x}$. This implies that E_x is measurable because it is the countable intersection of measurable sets. Since E is bounded we know that for large enough i the set $U_{i,x}$ has finite measure. Therefore by the continuity of measure we have that $\mu(U_{i,x})$ converges to $\mu(E_x)$ for almost every x. That is, f_{U_i} converges pointwise almost everywhere to f_E. We have shown previously that the pointwise limit of measurable functions is a measurable function. This shows property (2), that f_E is a measurable function. Finally, we have for large enough i that U_i has finite measure and so f_{U_i} is an integrable function. Since for this i we have $f_E \le f_{U_i}$ we may apply the dominated convergence theorem to get that $\lim \int f_{U_i} = \int f_E$. The left hand side of this is $\lim \mu(U_i)$ and we have already shown that this is $\mu(E)$. So we have shown property (3).

Now let's suppose that E is the countable intersection of open sets minus a set of measure zero. As before we can arrange for these open sets to be decreasing so we have open sets $U_0 \subseteq U_1 \subseteq ...$ and a set of measure zero E_0 such that $E = (U_0 \cap U_1 \cap ...) - E_0$. Let's assume again that E is bounded. Let's also denote by F the union $U_0 \cap U_1 \cap ....$ We know that the three properties hold for F and we also know that the three properties hold for E_0 and also that f_{E_0} is zero almost everywhere. We have that $E_x = U_{i,x} - E_{0,x}$ which is measurable as the difference of two measurable sets. We have that $\mu(E_x) = \mu(U_{i,x})$ because we have previously shown that E_x is a set of measure zero. We then have by the dominated convergence theorem that $\mu(E) = \mu(F)$ and we know that $\mu(F) = \mu(F - E_0)$. This shows that the three properties hold when E is a bounded measurable set.

To show that the three properties hold when E is an arbitrary measurable set you could consider E intersected with an open ball of radius n and call this set E_n. The set E_n is a measurable set and the three properties hold for it. We then have that E_x is the union $E_{1,x} \cup E_{2,x} \cup ...$ which is measurable as the countable union of open sets. We also have that f_E is the pointwise limit of the functions $f_{E_1}, f_{E_2}, ...$ and since these are measurable functions the limit function f_E is a measurable function. Finally, we can apply the monotone convergence theorem. The sequence of functions $f_{E_1} \leq f_{E_2}, ...$ is nondecreasing and converges to f. This implies that $\lim \int f_{E_n} = \int f_E$. The left hand side of this expression is $\mu(E_n)$ and we know that $\mu(E_n)$ converges to $\mu(E)$. So we have shown that the three properties hold for an arbitrary measurable set.

How can we apply this result. At the beginning of this email I asked the following question: Given a measurable function $f : \mathbf{R}^m \times \mathbf{R}^n \to \mathbf{R}$ what is the relationship between $\int f(x,y)d(x,y)$, $\int \left(\int f(x,y)dx\right) dy$, and $\int \left(\int f(x,y)dy\right) dx$?

First, let's suppose that f is a nonnegative function. Then we have that

$$\int f = \mu\{(x,y,z) \in \mathbf{R}^m \times \mathbf{R}^n \times \mathbf{R} \mid 0 \leq z \leq f(x,y)\}$$

Let's apply our result. Denote by E the set

$$\{(x,y,z) \in \mathbf{R}^m \times \mathbf{R}^n \times \mathbf{R} \mid 0 \leq z \leq f(x,y)\}.$$

We have that E_x is the set

$$\{(y,z) \in \mathbf{R}^n \times \mathbf{R} \mid 0 \leq z \leq f(x,y)\}.$$

Our three properties say that E_x is measurable for almost every x, that f_E is a measurable function, and that $\mu(E) = \int f_E$. Let's expand this last expression. The left hand side is $\int f(x,y,d(x,y)$. Since $\mu(E_x)$ is the integral $\int f(x,y)dy$ we have that the right hand side is $\int \left(\int f(x,y)dy\right) dx$. That is,

$$\int f(x,y)d(x,y) = \int \left(\int f(x,y)dy\right) dx$$

Likewise, we can apply this to show that

$$\int f(x,y)d(x,y) = \int \left(\int f(x,y)dy\right) dx$$

too. The answer to our question is that they are all equal to one another. This is called Tonelli's theorem.

What if f is an arbitrary measurable function. Well, in this case we can split f up so that $f = f_+ - f_-$. In this case we need to avoid the case where both the integral of both f_+ and f_- is infinity. Requiring that f be an integrable function is more than enough for this. This is called Fubini's theorem.

It gives us a way to iteratively compute the integral of a function whose domain is a subset of $\mathbf{R}^n$ when n is more than 1.

17. CONVEX FUNCTIONS

Let E be a measurable subset of $\mathbf{R}^n$. Consider the Lebesgue space $L^1(E)$.

Taking the dot product of two vectors tells you something about the angle between the vectors. We would like to know if given functions f, g in $L^1(E)$ whether the product of these two functions is in $L^1(E)$. The reason we would like to know this is because then we could consider the integral of fg and this would give us some idea of the "angle" between f and g.

One question we want the answer to is this. If f, g are integrable functions, then is fg an integrable function. The following example shows that the answer is no.

Example: Let $E = [0,1]$ and let f be defined by $f(x) = \frac{1}{\sqrt{x}}$. Then f belongs to $L^1(E)$ because $\int |f| = 2\sqrt{x}|_0^1 = 2$. However the product of f with itself is not integrable because $\int |f^2| = \log(x)|_0^1 = \infty$.

Definition: A function $f : E \to \mathbf{R}$ is square integrable if it is measurable and $\int f^2 < \infty$.

Is it true that if a function is square integrable then it is integrable?

Consider this counterexample. The series $1 + \frac{1}{2} + \frac{1}{3} + \ldots$ diverges to infinity but the series $1 + \frac{1}{4} + \frac{1}{9} + \ldots$ converges to $\frac{\pi^2}{6}$. This is a counterexample because we can let E be the nonnegative real numbers and let f be the function given by $f(x) = \frac{1}{\lceil x \rceil}$. Then we get that $\int f = 1 + \frac{1}{2} + \frac{1}{3} + \ldots$ and $\int f^2 = 1 + \frac{1}{4} + \frac{1}{9} + \ldots$.

It is not possible to make such a counterexample when the measure of E is finite. That is, when $\mu(E) < \infty$, if $f : E \to \mathbf{R}$ is a square integrable function, then f is an integrable function.

Now let's define some more spaces. Let's denote by $L^2(E)$ the set of square integrable functions on the set E and let's define the norm for this space as $\left(\int |f|^2\right)^{\frac{1}{2}}$. In general we define the set $L^p(E)$ to be the set of measurable functions on E such that f^p is an integrable function. We put the norm $\left(\int |f|^p\right)^{\frac{1}{p}}$ on it. We need to show hat these are norms. Before we do this let's talk about convex functions.

Definition: Let $\phi : \mathbf{R}^n \to \mathbf{R}$. We say that ϕ is convex if for all $x, y \in \mathbf{R}^n$ and for all $\lambda \in [0,1]$ we have that

$$\phi(\lambda y + (1-\lambda)x) \leq \lambda\phi(y) + (1-\lambda)\phi(x).$$

Intuitively, this is saying that the line segment connecting any two points on the graph of the function of ϕ lies above the graph of ϕ.

It turns out that convex functions are continuous and that they are differentiable almost everywhere. Our goal is to show these things.

In our definition for a convex function let $y = x + h$ where h is some positive number. Note that $\lambda(x+h) + (1-\lambda)x = x + \lambda h$. Then we have that for a convex function $\phi\phi(x + \lambda h) \leq \lambda\phi(x+h) + (1-\lambda)\phi(x)$. Subtracting $\phi(x)$ from both sides and dividing by λh gives

$$\frac{\phi(x+\lambda h) - \phi(x)}{\lambda h} \leq \frac{\phi(x+h) + \phi(x)}{h}$$

. This inequality implies that for a convex function the slope of the line connecting x and $x+h$ for $h > 0$ is nonincreasing as h gets smaller.

Let's now show that for a convex function ϕ as h goes to zero from the right the function f defined by $f(h) = \frac{\phi(x+h)+\phi(x)}{h}$ converges to some number which we will call $d^+\phi(x)$. We have already shown that the function $f(h)$ is nonincreasing. This implies that the sequence of numbers $f(1), f(\frac{1}{2}), f(\frac{1}{3})...$ is nonincreasing. If we can show that this sequence is bounded below then we will have shown that $d^+\phi(x)$ exists (because any monotonic bounded sequence converges). To show this consider $x - h$. Applying our definition of a convex function with $\lambda = \frac{1}{2}$ gives $\phi(x) \leq \frac{1}{2}\phi(x-h) + \frac{1}{2}\phi(x+h)$. It follows that $\frac{\phi(x)-\phi(x-h)}{h} \leq \frac{\phi(x+h)-\phi(x)}{h}$.

That is, the slope of the line segment joining $x - h$ and x is no more than the slope of the line segment joining x and $x+h$. By the same logic as before we know that the function $g(h) = \frac{\phi(x)-\phi(x-h)}{h}$ is nondecreasing function of h. So by fixing any $h^* \leq 1$ gives us $\frac{\phi(x)-\phi(x-h)}{h}$ as a lower bound for $f(1), f(\frac{1}{2}), f(\frac{1}{3}), ...$

With the same sort of argument we can show that the sequence of numbers $g(1), g(\frac{1}{2}), g(\frac{1}{3}), ...$ converges. Let's call this limit $d^-\phi(x)$. Note also that since $g(h) \leq f(h)$ for each h we have that $d^-\phi(x) \leq d^+\phi(x)$.

We call $d^-\phi(x)$ the derivative at x from the left and we call $d^+\phi(x)$ the derivative at x from the right. We have just shown that for a convex function both of these derivatives exist and that the derivative from the left is no greater than the derivative from the right. Note that these derivatives being equal means that the function ϕ is differentiable at the point x. Note that for a convex function these derivatives need not be equal. For example for the absolute value function at the point $x = 0$ the derivative from the left if -1 and the derivative from the right is 1. It turns out that the set of points where a convex function is not differentiable has measure zero.

Now suppose we wanted to compare the derivative of the convex function ϕ at two points $x < y$. By letting $y = x + h$ we know that

$$d^-\phi(x) \leq d^+\phi(x) \leq \frac{\phi(y) - \phi(x)}{y - x} \leq d^-\phi(x) \leq d^+\phi(x).$$

This shows that the functions $d^-\phi$ and $d^+\phi$ are nondecreasing functions. We will prove later that a a differentiable function is convex if its derivative is nondecreasing.

Let's first consider continuity. It turns out that a convex function is continuous. Let's prove this.

Claim 1: If $\phi : \mathbf{R}^n \to \mathbf{R}$ is a convex function, then it is a continuous function.

Proof: Consider an interval $[a, b]$ and let x and y belong to this interval such that $x < y$. From before we have that $d^+\phi(a) \leq \frac{\phi(y)-\phi(x)}{y-x} \leq d^+\phi(b)$. For $C = \max(d^+\phi(a), d^+\phi(b))$ this implies that $|\phi(y) - \phi(x)| \leq C|x - y|$. This implies that ϕ is continuous on the interval $[a, b]$. (A function that satisfies this condition is called Lipschitz continuous). Repeating the argument for other intervals shows that the function ϕ is continuous everywhere.

QED

Another theorem.

Claim 2: Let ϕ be a convex function. Suppose that for a point x in the domain of ϕ there exists a number m such that $d^-(x) \leq m \leq d^+(x)$. Then for all y in the domain of ϕ we have that $my + (\phi(x) - mx) \leq \phi(y)$ This is saying that the line through the point $(x, \phi(x))$ with slope m lies below the graph of ϕ.

Proof: The inequality holds when $x = y$. Suppose that $x < y$. Rearranging the inequality we would like to show and dividing by $y - x$ gives

$$m \leq \frac{\phi(y) - \phi(x)}{y - x}.$$

We have shown previously that

$$d^-\phi(x) \leq d^+\phi(x) \leq \frac{\phi(y) - \phi(x)}{y - x}$$

so this inequality is true because by assumption $d^-(x) \leq m \leq d^+(x)$. A similar argument applies when $y < x$.

QED

The reason we introduced the notion of a convex function is to prove the following theorem.

Theorem 1: (Jensen's inequality) Let $\phi : \mathbf{R} \to \mathbf{R}$ be a convex function and let $f : E \to \mathbf{R}$ be an integrable function. Then

$$\phi\left(\int f\right) \leq \int \phi \circ f.$$

Proof: Choose a number m such that $d^- \left(\int f\right) \leq m \leq d^+ \left(\int f\right)$. By Claim 2 we have that there exist numbers m and b so that $mx+b \leq \phi(x)$ for all x with equality for $x = \int f$. By the linearity and monotonicity of the Lebesgue integral we have that

$$\phi\left(\int f\right) = m \int f + b \leq \int \phi \circ f.$$

QED

Now let's prove a criterion for deciding if a function is convex.

Claim 3: Suppose that $\phi : \mathbf{R} \to \mathbf{R}$ is differentiable. If the derivative of ϕ is nondecreasing, then ϕ is a convex function.

Proof: Let $h > 0$. We want to show that $\phi(x + \lambda h) \leq \lambda\phi(x + h) + (1 - \lambda)\phi(x)$. By the Mean Value Theorem there exists a point $x_1 \in [x, x + \lambda h]$ such that $\frac{\phi(x+\lambda h)-\phi(x)}{\lambda h} = \phi'(x_1)$. By the Mean Value Theorem there exists a point $x_2 \in [x + \lambda h, x + h]$ such that $\frac{\phi(x+h)-\phi(x+\lambda h)}{(1-\lambda)h} = \phi'(x_2)$. By hypothesis we have that $\phi'(x_1) \leq \phi'(x_2)$. This implies that $\phi(x + \lambda h) \leq \lambda\phi(x + h) + (1 - \lambda)\phi(x)$.

QED

Example: Consider the function f given by $f(x) = x^2$. This function has a nondecreasing derivative. By Claim 3 it is convex. More generally consider the functions $|x|^p$ for $p \geq 1$. Such a function has a nondecreasing derivative and so by Claim 3 is convex.

Let's now define Lebesgue spaces of order p.

Definition: Let $p \geq 1$. Let E be a measurable subset of $\mathbf{R}^n$. Define the Lebesgue space of order p to be the set of functions

$$L^p(E) = \left\{ f : E \to \mathbf{R} | f \text{ measurable and } \left(\int |f|^p\right)^{\frac{1}{p}} < \infty \right\}$$

modulo functions that are the same almost everywhere. Claim 4: If f, g belong to $L^p(E)$, then $f + g$ belongs to $L^p(E)$.

Proof: Since the absolute value raised to the power p defines a convex function we have that

$$\begin{aligned}\left(\int |f+g|^p\right)^{\frac{1}{p}} &= 2\left(\int |\frac{f+g}{2}|^p\right)^{\frac{1}{p}} \\ &\leq 2\left(\int |\frac{f}{2}|^p + \int |\frac{g}{2}|^p\right)^{\frac{1}{p}} \\ &= 2\left(\int |\frac{f}{2}|^p + |\frac{g}{2}|^p\right)^{\frac{1}{p}} < \infty.\end{aligned}$$

QED

Example: The function $x \mapsto e^x$ is a convex function. This follows from claim 3 because this function's first derivative is nondecreasing. So for example we have that for any real numbers a and b and any number $\lambda \in [0,1]$ that $e^{\lambda a+(1-\lambda)b} \leq \lambda e^a + (1-\lambda)e^b$.

Theorem 2: (Young's inequality) For any real numbers x and y and any integers $p, q > 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$ we have that $xy \leq \frac{1}{p}x^p + \frac{1}{q}y^q$. Proof: Let $\lambda = \frac{1}{p}$ where $p > 1$ and $\frac{1}{q} = (1-\lambda)$. Let $a = p \log x and let b = q \log(y)$. Then the convexity of the exponential function (using the notation above) tells us that $xy \leq \frac{1}{p}x^p + \frac{1}{q}y^q$. QED

Next time we will prove some more inequalities and relate them to Lebesgue spaces.

18. MORE INEQUALITIES

Recall the definition of a Lebesgue space of order p: Let E be a measurable subset of $\mathbf{R}^n$. The Lebesgue space of order p is the set of measurable functions $f : E \to \mathbf{R}$ (modulo functions that are the same almost everywhere) such that the expression $\left(\int |f|^p\right)^{\frac{1}{p}}$ is finite . We denote this set of functions by $L^p(E)$. The reason we say 'modulo functions that are the same almost everywhere' is that we want to show that the Lebesgue space of order p with the function $\|\cdot\|_{L^p} : L^p(E) \to \mathbf{R}$ defined by $\left(\int |f|^p\right)^{\frac{1}{p}}$ is a normed vector space (and later a Banach space). In order for the function $\|\cdot\|_{L^p}$ to be a norm we require that its value is zero if and only if the function it is considering is the zero function. However, any function that is the same almost everywhere to the zero function will also take a value of zero under this function. We get around this by thinking of the elements of $L^p(E)$ as the set of functions that are almost everywhere the same. (This raises an interesting idea: given two functions that are not the same almost everywhere it is impossible to change one of the functions on a set of measure zero, then to change it again on a set of measure zero, then again, and so on forever, to make it the same as the other function.) We have previously shown that the Lebesgue space of order 1 is a Banach space (and so also a normed vector space). What we would like to do now is to show that this is true for Lebesgue spaces of order more than 1. Another thing we would like to do is to figure out when the product of two functions in a Lebesgue space of order p is also in that space. To do this we introduced the idea of a convex function. The reason this will help us is that the function given by the expression $|x|^p$ is convex for all $p \geq 1$. Last time we developed the theory of convex function and finished by proving an inequality called Young's inequality (which follows from our showing that the exponential function is a convex function).

Young's inequality: For all real numbers x and y and all numbers $p, q > 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$ we have that $xy \leq \frac{1}{p}x^p + \frac{1}{q}x^q$.

The way we proved this was to note that the exponential function is convex so that for all real numbers a and b and all numbers λ in $[0,1]$ we have

$$e^{\lambda a+(1-\lambda)b} \leq \lambda e^{a} + (1-\lambda)e^{b}$$

and then to make the substitution

$$\lambda = \frac{1}{p}, 1-\lambda = \frac{1}{q}, a = p\log(x), b = q\log(y).$$

A corollary of Young's inequality is an inequality called Holder's inequality.

Corollary 1: (Holder's inequality) Let $f : E \to \mathbf{R}$ and $g : E \to \mathbf{R}$ be measurable functions and let $p, q \geq 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Then

$$\int |fg| \leq \left(\int |f|^{p}\right)^{\frac{1}{p}} \left(\int |g|^{q}\right)^{\frac{1}{q}}.$$

Proof: The inequality holds if either f or g takes the value of zero almost everywhere because then fg will take a value of zero almost everywhere and so the left hand side of the integral will be zero. It is clear that both sides of the inequality are nonnegative numbers. So let's assume neither f nor g takes the value zero almost everywhere. In this case the inequality also holds if one of the integrals on the right is infinite. So let's now also assume that neither of the integrals on the right hand side are infinite. That is, let's assume that f belongs to the Lebesgue space of order p on E and that g belongs to the Lebesgue space of order q on E. Now define the functions $\hat{f}$ by $\hat{f} = \frac{f}{(\int |f|^{p})^{\frac{1}{p}}}$ and $\hat{g} b y \hat{g} = \frac{g}{(\int |g|^{q})^{\frac{1}{q}}}$. If we can show that $\int \hat{f}\hat{g} \leq 1$ we will be done. Notice that Young's inequality tells us that $|\hat{f}\hat{g}| \leq \frac{1}{p}|\hat{f}|^{p} + \frac{1}{q}|\hat{g}|^{q}$. Integrating this expression gives

$$\int \hat{f}\hat{g} \leq \frac{1}{p}\int |\hat{f}|^{p} + \frac{1}{q}\int |\hat{g}|^{q} = \frac{1}{q} + \frac{1}{p} = 1.$$

QED

A special case of Holder's inequality when $p = 2$ is called The Cauchy-Schwartz Inequality. It get's its own name because it comes up quite often.

Holder's inequality tells us about when the product of two measurable functions is an element of $L^{1}(E)$.

Now let's prove an inequality called Minkowski's Inequality that will help us show that the function $\|\cdot\|_{L^p}$ on $L^{p}(E)$ defined for $f \in L^{p}(E)$ by the equation $\|f\|_{L^p} = \left(\int |f|^{p}\right)^{\frac{1}{p}}$ is a norm. Last time we showed that a Lebesgue space of order p is a vector space. Once we show this function is a norm we will know that the Lebesgue space of order p is a normed vector space.

Minkowski's inequality is a corollary of Holder's inequality.

Corollary 2: (Minkowski's inequality): Let $f, g \in L^p(E)$. Then

$$\|f+g\|_{L^p} \leq \|f\|_{L^p} + \|g\|_{L^p}.$$

Proof: We have previously shown this for the case where $p = 1$ so assume $p > 1$. Note that if the inequality holds when we replace f and g by $|f|$ and $|g|$ then it also holds for f and g. This is because when we replace $|f|$ and $|g|$ by f and g we have that the right hand side of the inequality is unchanged and the left hand side (by the monotonicity of the Lebesgue integral) cannot increase. So we may assume that both f and g are nonnegative functions. The reason this is useful is that we don't have to worry about the absolute values. Now let's try to prove the inequality. We have that $(f+g)^p = (f+g)^{p-1}f + \int (f+g)^{p-1}g$. By the monotonicity of the Lebesgue integral and then Holder's inequality we have that

$$\int (f+g)^p \leq \int (f+g)^{p-1}f + \int (f+g)^{p-1}g \leq (\|f\|_{L^p} + \|g\|_{L^p})\|(f+g)^{p-1}\|_{L_q}$$

What is $\|(f+g)^{p-1}\|_{L_q}$? Well we have that $\|(f+g)^{p-1}\|_{L_q} = \left(\int (f+g)^{(p-1)q}\right)^{\frac{1}{q}}$ and since $\frac{1}{p} + \frac{1}{q} = 1$ we have that $(p-1)q = p$ so that $\|(f+g)^{p-1}\|_{L_q} = \left(\int (f+g)^p\right)^{\frac{1}{q}}$. Going back to our our first inequality it follows that

$$\left(\int (f+g)^p\right)^{1-\frac{1}{q}} \leq (\|f\|_{L^p} + \|g\|_{L^p})\|(f+g)^{p-1}\|_{L_q}$$

and the left hand side of this is just $\|f+g\|_{L^p}$ since $1 - \frac{1}{q}$ is equal to $\frac{1}{p}$.

QED

Next time I will use these inequalities to show that each Lebesgue space of order p is a Banach space and to discuss when the product of two elements of $L^p(E)$ is also an element of $L^p(E)$.

19. L^P IS A BANACH SPACE

Recall that a Banach space is a vector space V together with a norm $\|\cdot\| : V \to \mathbf{R}$ such that if $v_1, v_2, \ldots$ is a sequence in V whose terms are getting arbitrarily close to one another then this sequence converges to a point in v in V and this convergence is measured using the norm.

Let E be a measurable subset of $\mathbf{R}^n$. We want to show that the set of functions $L^p(E)$ which is the set of functions $\{f : E \to \mathbf{R} \mid f \text{ is measurable and } \|f\|_{L^p} < \infty\}$ with $\|f\|_{L^p}$ given by $\left(\int |f|^p\right)^{\frac{1}{p}}$ is a Banach space.

It's easy to see that $L^p(E)$ is a vector space:

(1) If $\alpha \in \mathbf{R}$ and $f \in L^p(E)$, then $\alpha f \in L^p(E)$: We have that αf is a measurable function and $\|\alpha f\|_{L^p} < \infty$ because $\|f\|_{L^p} < \infty$.

(2) If f and g are in $L^p(E)$, then $f + g$ is in $L^p(E)$: the function $f + g$ is measurable and

$$\|f+g\|_{L^p} = \left(\int |f+g|^p\right)^{\frac{1}{p}} = 2\left(\int |\frac{f+g}{2}|^p\right)^{\frac{1}{p}}$$
$$\leq 2\left(\int \frac{1}{2}(|f|^p + |g|^p)\right)^{\frac{1}{p}} = (\|f\|_{L^p}^p + \|g\|_{L^p}^p)^{\frac{1}{p}} < \infty$$

so that $f+g$ belongs to $L^P(E)$.

(3) The zero element of $L^p(E)$ is the zero function. This is a measurable function since E is measurable and $\|0\| = 0 < \infty$. THerefore it is in $L^p(E)$.

It is also easy to show that $\|\cdot\|_{L^p}$ is a norm on $L^p(E)$:

(1) $\|f\|_{L^p(E)} = 0$ if and only if f is the zero element of $L^p(E)$: If f is the zero element of $L^p(E)$ then $\|f\|_{L^p} = (|f|^p)^{\frac{1}{p}} = 0$. If $\|f\|_{L^p} = 0$ then it must be that $\int |f|^p = 0$. We have previously shown that the integral of a nonnegative function is zero if and only if that function is zero almost everywhere. So $|f|^p = 0$ almost everywhere. This implies that $f = 0$ almost everywhere.

(2) If f is in $L^p(E)$, then $\|f\|_{L^p} \geq 0$: this is true because the integral of a nonnegative function is always nonnegative.

(3) If f and g are in L^p, then $\|f+g\|_{L^p} \leq \|f\|_{L^p} + \|g\|_{L^p}$: This is more difficult to show. Fortunately we have already proved it. We have previously proved Minkowski's inequality which states exactly this.

All we are left with is the task of showing that the normed vector space $L^p(E)$ with the norm $\|\cdot\|_{L^p}$ is complete: that is, if $f_1, f_2, \cdots$ is a sequence in $L^p(E)$ such that for each number $\epsilon > 0$ there exists an integer N such that $n, m \geq N$ implies that $\|f_n - f_m\| < \epsilon$, then there exists an element f of $L^p(E)$ such that this sequence converges to f, that is, $\|f - f_n\|_{L^p}$ converges to zero as n goes to infinity.

What would be a good guess for the function that this sequence converges to? Well, for each point x in E we have that $f_1(x), f_2(x), \cdots$ is a sequence of real numbers. Let's define f by letting $f(x)$ be the limit of this sequence of numbers when this limit exists and zero otherwise.

We want to show that the sequence of numbers $\|f - f_1\|_{L^p}, \|f - f_2\|_{L^p}, \cdots$ converges to zero. That is, we want to show that the sequence of numbers $(|f - f_1|^p)^{\frac{1}{p}}$, $\left(\int |f - f_2|^p\right)^{\frac{1}{p}}, \cdots$ converges to zero. This is equivalent to showing that the sequence of numbers $\int |f - f_1|^p, \int |f - f_2|^p, \cdots$ converges to zero.

What do we know about the sequence of functions $|f - f_1|^p, |f - f_2|^p, ...$? If the set on which the sequence of functions $f_1, f_2, ...$ does not converge has measure zero, then we know that this sequence of functions converges to the zero function almost everywhere. Note that the sequence of functions $f_1, f_2, \cdots$ is bounded. Let's show this. Let $\epsilon = 1$. Then we can find a number M such that $\|f_n - f_m\|_{L^p} < 1$ whenever $n, m \geq M$. This implies that the sequence is eventually within a radius of 1 from the function f_M and so if $n \geq M$, then

$$\|f_n\|_{L^p} = \|f_n - f_m + f_m\|_{L^p} \leq \|f_n - f_m\|_{L^p} + \|f_M\|_{L_p} = 1 + \|f_M\|_{L^p}.$$

The first $M-1$ terms of the sequence have finite norms so take the bound to be the maximum of these and $1+\|f_M\|_{L^p}$. This implies that the sequence of functions $|f-f_1|^p, |f-f_2|^p, ...$ is bounded. We then know this and that this sequence converges to the zero function almost everywhere. By the Bounded Convergence Theorem we have that $\lim_n \int |f-f_n|^p = 0$.

But we're still not done. In the last paragraph we assumed that the set of points where the sequence of functions $f_1, f_2, ...$ does not converge has Lebesgue measure zero. Notice that we can write the limit of this sequence as the limit of the sequence $f_1, f_1+(f_2-f_1), f_1+(f_2-f_1)+(f_3-f_2), ...$ We could even be more clever and let n_j be the number such that $\|f_m - f_n\|_{L^p} < \frac{1}{2^j}$ whenever $n, m \geq n_j$. We could then write the limit of the sequence $f_{n_1}, f_{n_2}, ...$ as the limit of the sequence $f_{n_1}, f_{n_1} + (f_{n_2} - f_{n_1}),\ f_{n_1} + (f_{n_2} - f_{n_1}) + (f_{n_3} - f_{n_2}) ....$ We then know that for this sequence we have

$$\sum_{j=1}^{\infty} \|f_{n_{j+1}} - f_{n_j}\|_{L^p} < \sum_{j=1}^{\infty} \frac{1}{2^j} = 1.$$

Let g_i denote $f_{n_i} - f_{n_{i-1}}$. We want to show that for almost all x we have that $g_1(x) + g_2(x) + ...$ converges. Consider the set of points x where $g_1(x)+g_2(x)+...$ does not converge. This set is contained in the set of points x such that $|g_1(x)| + |g_2(x)| + ...$ is equal to infinity. We can write this set in the following way. Let

$$S_{n,k} = \{x \in E \mid |g_1(x)| + |g_2(x)| + ... + |g_n(x)| \geq 2^k\}.$$

This is the set of points where $|g_1(x)| + |g_2(x)| + ...|g_n(x)|$ is at least 2^k. The set of points x where $|g_1(x)| + |g_2(x)| + ...$ is equal to infinity can be written as $\bigcap_{k=1}^{\infty} \bigcup n = 1^{\infty} S_{n,k}$. What can we say about the measure of $S_{n,k}$? We have the following inequality $2^k \chi_{S_{n,k}} \leq |g_1(x)| + |g_2(x)| + ... + |g_n(x)|$. Integrating this gives $2^k \mu(S_{n,k}) \leq \int |g_1(x)| + \int |g_2(x)| + ... + \int |g_n(x)|$ The right hand side is $\sum_{j=1}^{n} \|g_j\|_{L^p}$ and we know from above that this is no more than 1. Since the sets $S_{1,k}, S_{2,k}, ...$ are an increasing sequence of measurable sets we have by the continuity of measure that $\mu\left(\bigcup_{n=1}^{\infty} S_{n,k}\right) \leq \frac{1}{2^k}$.. By the continuity of measure again we have that $\mu\left(\bigcap_{k=1}^{\infty} \bigcup n = 1^{\infty} S_{n,k}\right) = 0$.

What we have shown is that if $f_1, f_2, ...$ is a sequence of functions in $L^p(E)$ and this sequence is a Cauchy sequence (i.e. gets closer and closer together in the L^p norm), then there exists a subsequence $f_{n_1}, f_{n_2}, ...$ that converges pointwise to a function f except maybe on a set of measure zero. This is enough because if we know that $f_1, f_2, ...$ is a Cauchy sequence and that some subsequence $f_{n_1}, f_{n_2}, ...$ of it converges pointwise to a function f then it must be that $f_1, f_2, ...$ converges to f in the L^p norm.

Let's prove this. Let $\epsilon > 0$. There exists an integer N such that $\|f - f_{n_i}\| < \frac{\epsilon}{2}$ whenever $i \geq N$. There also exists an integer M such that $\|f_n - f_m\| < \frac{\epsilon}{2}$ whenever $n, m \geq M$. This implies that

$$\|f - f_i\|_{L^p} = \|f - f_{n_i} + f_{n_i} - f_i\|_{L^p} \leq \|f - f_{n_i}\|_{L^p} + \|f_{n_i} - f_i\|_{L^p} < \epsilon$$

whenever i is at least the larger of M and N. Now we are done because we have shown that $L^p(E)$ is complete.

20. LINEAR FUNCTIONS AND THE DUAL SPACE

Let E be a measurable subset of $\mathbf{R}^n$. Consider the set of functions $L^p(E)$. We are now going to study functions $\lambda : L^p(E) \to \mathbf{R}$ which are linear. That is, if α is a real number and f is an element of $L^p(E)$, then $\lambda(\alpha f) = \alpha\lambda(f)$. And if f and g are elements of $L^p(E)$, then $\lambda(f + g) = \lambda(f) + \lambda(g)$.

We first just want to make some observations about linear functions on normed vector spaces. Let V and W be normed vector spaces. It is clear what it means for a function $\lambda : V \to W$ to be linear.

When is a linear function continuous? A sufficient condition is that there exists a constant C such that for all points x and y in V we have that the distance between the points $\lambda(x)$ and $\lambda(y)$ which are both points in W is no more than the constant C times the distance between the points x and y. That is,

$$\|\lambda(x) - \lambda(y)\|_W \leq C\|x - y\|_V.$$

The reason this implies that λ is a continuous function is that we can make $\lambda(y)$ as close as we want to $\lambda(x)$ by choosing y to be sufficiently close to x. Now it is clear that a function can be continuous without satisfying this condition. For example, consider the function $f : (0, 1) \to \mathbf{R}$ defined by the formula $f(x) = \frac{1}{x}$. This is a continuous function but does not satisfy the condition. For instance let $x = 1$. Then for some point y in $(0, 1)$ the distance between the image of x under f and the image of y under f converges to infinity as y gets closer and closer to zero. That is, $\lim_{y\to 0} |f(x) - f(y)| = \lim_{y\to 0} |1 - \frac{1}{y}| = \infty$. But the distance between x and y is at most one so there can never be a constant C which works.

But, as you might expect, for linear functions this condition is equivalent to continuity. To show this we only need to show that any linear function $\lambda : V \to W$ that is continuous satisfies the condition. Continuity of λ means that for all $\epsilon > 0$ there exists a $\delta > 0$ such that $\|\lambda(x) - \lambda(y)\|_W < \epsilon$ whenever $\|x - y\|_V < \delta$. First of all, suppose that λ is continuous at the point 0. This implies that λ is continuous at any point x in V.

To see this, let $\epsilon > 0$. We want to choose $\delta > 0$ such that $\|\lambda(x) - \lambda(y)\| < \epsilon$ whenever $\|x - y\| < \delta$. Since $\lambda(0) = 0$ and λ is a linear function this is the same as choosing $\delta > 0$ such that $\|\lambda(0) - \lambda(x-y)\| < \epsilon$ whenever $\|0 - (x-y)\| < \delta$ and this is what it means for λ to be continuous at the point 0. Note that the same argument applies to our condition. If the condition holds for $x = 0$, then it holds in general.

To see this, note that $\|\lambda(x) - \lambda(y)\|_W \leq C\|x - y\|$ is the same as $\|\lambda(0) - \lambda(y - x)\| \leq C\|0 - (y-x)\|$ and this true if the condition above holds when x (the one in the original statement of the condition) is equal to the zero element of V. So to prove that if a linear function is continuous, then it satisfies the condition we can prove that if a linear function is continuous at zero, then it satisfies the condition at $x = 0$. That is, we would like to show that if for each $\epsilon > 0$ there exists a $\delta > 0$ such that $\|\lambda(y)\| < \epsilon$ whenever $\|y\| < \delta$, then it is also true that there exists a constant C such that for all z in V we have that $\|\lambda(z)\| \leq C\|z\|$.

Let's prove this. For each z in V we have $\|\lambda(Kz)\| < \epsilon$ whenever $\|Kz\| < \delta$. Let $K = \frac{\delta}{\|z\|}$. Then we have that $\|\lambda(z)\| \leq \frac{\epsilon}{\delta}\|z\|$ whenever $\delta \leq \delta$. Let $C = \frac{\epsilon}{\delta}$. Then for all z in V we have $\|\lambda(z)\| \leq C\|z\|$.

The condition that for all $z \in V$ there exists a constant C such that $\|\lambda(z)\| \leq C\|z\|$ for a linear function is called boundedness and a linear function that satisfies it is called bounded. What we have shown is that a linear function is bounded if and only if it is continuous.

Here is an idea. Consider the set of all linear functions $\lambda : V \to W$. Let's denote this set by $\mathrm{Hom}(V, W)$. Let's define a function that tells us whether or not an element $\lambda of \mathrm{Hom}(V, W)$ is bounded. The function will take as its input the linear function λ and produce infinity if λ is not bounded and if λ is bounded the function will output the smallest bound C_0 such that for all z in V we have that $\|\lambda(z)\|_W \leq C_0\|z\|_V$. That is define the function from $\mathrm{Hom}(V, W)$ to $\mathbf{R}$ by the formula

$$C_0 = \inf\{C \in \mathbf{R} : \text{ for all } z \text{ in } V \text{ we have } \|\lambda(z)\| \leq C\|z\|\}$$

and denote it by $\|\lambda\|_{\mathrm{Hom}(V,W)}$. One thing we need to check is that this is a number that makes $\|\lambda(z)\| \leq C\|z\|$ for all z in V . That is, it might be that this infimum is not achieved. But it is achieved because we can find a sequence of numbers $C_1, C_2, ...$ converging to C_0 such that for each $n = 1, 2, ...$ and for each z in V we have $\|\lambda(z)\| \leq C_n\|z\|$. Taking the limit we get $\|\lambda(z)\| \leq C_0\|z\|$. Now it turns out that $\|\lambda\|_{\mathrm{Hom}(V,W)}$ is a norm and we will give it the name of the operator norm. Let's show that the operator norm is in fact a norm. We have to show three things:

(1) The operator norm only takes nonnegative values. (2) For each real number α and each $\lambda in \mathrm{Hom}(V,W)$ we have $\|\alpha\lambda\|_{\mathrm{Hom}(V,W)} = |\alpha|\|\lambda\|_{\mathrm{Hom}(V,W)}$. (3) The operator norm takes the value zero if and only if it is evaluated at the zero element of $\mathrm{Hom}(V,W)$. (4) The operator norm satisfies the triangle inequality: for all λ, ψ in $\mathrm{Hom}(V,W)$ we have

$$\|\lambda+\psi\|_{\mathrm{Hom}(V,W)} \leq \|\lambda\|_{\mathrm{Hom}(V,W)} + \|\psi\|_{\mathrm{Hom}(V,W)}.$$

Let's show each of these:

(1) You can see from the definition of C_0 above that the operator norm is nonnegative.

(2) This is also obvious from the definition of C_0 above.

(3) If λ is the zero function, then $\|\lambda(z)\| = 0 \leq 0\|z\|$ so that$\|\lambda\|_{\mathrm{Hom}(V,W)} = 0$. Now suppose that $\|\lambda\|_{\mathrm{Hom}(V,W)} = 0$. This means that for all z in the vector space V we have that $\|\lambda(z)\|_W \leq 0\|z\|_V = 0$ Since $\|\cdot\|_W$ is a norm this implies that $\lambda(z) = 0$. Therefore $\lambda(z) = 0$ for all z in V.

(4) Let λ and ψ be elements of $\mathrm{Hom}(V,W)$. Then for each z we have that

$$\|(\lambda+\phi)(z)\| = \|\lambda(z)\| + \|\phi(z)\| \leq \|\lambda\|_{\mathrm{Hom}(V,W)}\|z\| + \|\phi\|_{\mathrm{Hom}(V,W)}\|z\|.$$

This implies that

$$\|\lambda\|_{\mathrm{Hom}(V,W)} + \|\phi\|_{\mathrm{Hom}(V,W)}$$

is a bound for $\lambda+\phi$. Therefore

$$\|\lambda+\phi\|_{\mathrm{Hom}(V,W)} \leq \|\lambda\|_{\mathrm{Hom}(V,W)} + \|\phi\|_{\mathrm{Hom}(V,W)}.$$

So the operator norm is in fact a norm. We forgot to show that $\mathrm{Hom}(V,W)$ is a vector space. This is easy to show. The zero element if the zero function which is a linear function. The sum of two linear functions is a linear function. And a scalar multiple of a linear function is a linear function. Therefore $\mathrm{Hom}(V,W)$ is a vector space and since the operator norm is a norm it is a normed vector space when equipped with the operator norm.

Since we have a normed vector space a natural question to ask is whether it is Banach space. That is, it $\mathrm{Hom}(V,W)$ equipped with the operator norm complete. That is, if $\lambda_1, \lambda_2, ...$ is a Cauchy sequence in $\mathrm{Hom}(V,W)$, then is there an element $\lambda of \mathrm{Hom}(V,W)$ to which this sequence converges?

Let's try to prove this. Since $\lambda_1, \lambda_2, ...$ is a Cauchy sequence for each $\epsilon > 0$ there exists an integer N such that $m, n \geq N$ implies that $\|\lambda_n - \lambda_m\|_{\mathrm{Hom}(V,W)} < \epsilon$. This implies that for each $z \in V$ we have that $\lambda_1(z), \lambda_2(z), ...$ is a Cauchy sequence in W since $n, m \geq N$ implies that

$$\|\lambda_n(z) - \lambda_m(z)\|_W \leq \|\lambda_n - \lambda_m\|_{\mathrm{Hom}(V,W)}\|z\|_V < \epsilon\|z\|.$$

To get a guess for the function the sequence $\lambda_1, \lambda_2, ...$ converges to let's assume that W is itself complete. This means that each Cauchy sequence in W converges to some point. Since for each z in V the sequence $\lambda_1(z), \lambda_2(z), ...$ is a Cauchy sequence in W and W is complete we have that there is a point in W which we will denote by $\lambda(z)$ to which this sequence converges. This defines a function $\lambda : V \to W$. Let's show that this function λ is an element of $\mathrm{Hom}(V, W)$. That is, let's show that λ is a linear function.

Let x and y belong to V. Because each λ_n is a linear function and addition is continuous we have that

$$\begin{aligned}\lambda(x+y) &= \lim_n \lambda_n(x+y) = \lim_n \lambda_n(x) + \lambda_n(y)\\ &= \lim_n \lambda_n(x) + \lim_n \lambda_n(y) = \lambda(x) + \lambda(y).\end{aligned}$$

Let α be a real number and x an element of V. Because each λ_n is a linear function and multiplication is continuous we have that

$$\begin{aligned}\lambda(\alpha x) &= \lim_n \lambda_n(\alpha x) = \lim_n \alpha\lambda_n(x)\\ &= \alpha \lim_n \lambda_n(x) = \alpha\lambda(x).\end{aligned}$$

Therefore λ is a linear function from V to W.

To show that $\mathrm{Hom}(V, W)$ is complete we now only need to show that it is bounded and that the Cauchy sequence $\lambda_1, \lambda_2, ...$ in $\mathrm{Hom}(V, W)$ converges to λ where this convergence is in terms of the operator norm.

First note that you we can rewrite the definition of the operator norm. Our definition of the operator norm of a linear function $\lambda : V \to W$ is

$$\|\lambda\|_{\mathrm{Hom}(V,W)} = \inf\{C \in \mathbf{R} : \text{for all } x, \|\lambda(x)\|_W \leq C\|x\|_V\}.$$

This right hand side of this equation is the same as

$$\inf\{C \in \mathbf{R} : \text{for all } x \neq 0, \|\lambda(x)\|_W \leq C\|x\|_V\}$$

This is the same as

$$\sup\left\{\frac{\|\lambda(x)\|_W}{\|x\|_V} : x \in V - \{0\}\right\}.$$

And since λ is a linear function so that $\frac{\|\lambda(x)\|_W}{\|x\|_V} = \|\lambda\left(\frac{x}{\|x\|_V}\right)\|$ we can write y for $\frac{x}{\|x\|_V}$ and we get

$$\sup\{\|\lambda(y)\|_W : \|y\|_V = 1\}.$$

Note that we only need the condition $\|y\|_V = 1$ because $y = \frac{x}{\|x\|_V}$ where $x \in V - \{0\}$ implies that $\|y\|_V = 1$. And for each x in V there exists a $y = \frac{x}{\|x\|_V}$ in V such that $\|y\|_V = 1$. That is, the conditions $y = \frac{x}{\|x\|_V}$ where $x \in V - \{0\}$ and $\|y\|_V = 1$ are equivalent. So our new definition of the operator norm of a bounded linear function $\lambda : V \to W$, that is, and element of $\mathrm{Hom}(V, W)$ is $\|\lambda\|_{\mathrm{Hom}(V,W)} = \sup\{\|\lambda(y)\|_W :$

$\|y\|_V = 1\}$. Let's show that λ is bounded. Since $\lambda_1, \lambda_2, ...$ is a Cauchy sequence it is bounded. That is, there exists a number M such that $\|\lambda_n\|_{\mathrm{Hom}(V,W)} \leq M$ for all n. Let's assume that $\lim_n \|\lambda - \lambda_n\|_{\mathrm{Hom}(V,W)} = 0$. Therefore we have that

$$\|\lambda\|_{\mathrm{Hom}(V,W)} \leq \|\lambda_n\|_{\mathrm{Hom}(V,W)} + \|\lambda - \lambda_n\|_{\mathrm{Hom}(V,W)}.$$

Taking the limit at n goes to infinity gives

$$\|\lambda\|_{\mathrm{Hom}(V,W)} \leq M$$

which implies that M is a bound for λ. That is, for all $x \in V$ we have that $\|\lambda(x)\|_W \leq M\|x\|_V$. Finally, let's show that $\lim_n \|\lambda - \lambda_n\|_{\mathrm{Hom}(V,W)} = 0$. Let y belong to V such that $\|y\|_V = 1$. Then we have that $\lim_n \|\lambda(y) - \lambda_n(y)\|_W = 0$. From the definition of the operator norm we just derived this implies that λ_n converges to λ under the operator norm as n tends to infinity. That is, $\lim_n \|\lambda - \lambda_n\|_{\mathrm{Hom}(V,W)} = 0$.

We have shown that if W is complete then $\mathrm{Hom}(V, W)$ is a Banach space.

Given a normed vector space V we will study the set of real valued linear functions $\mathrm{Hom}(V, \mathbf{R})$. Since the real numbers are complete this set equipped with the operator norm is a Banach space. We will call $\mathrm{Hom}(V, \mathbf{R})$ the duel space of V and denote it by V^*.

21. THE HAHN BANACH THEOREM

Let V be a normed vector space. Recall that the dual of V denoted V^* is the set of bounded linear functions from V to the real numbers $\mathbf{R}$. Recall that a linear function $\lambda : V \to \mathbf{R}$ is called bounded if there exists a number C such that $\|\lambda(x)\|_W \leq C\|x\|_V$ for all x in V. Last time we proved that the dual space with the operator norm is a Banach space. Recall that the operator norm on the dual space of V takes as its input a bounded linear function $\lambda : V \to \mathbf{R}$ and outputs the smallest bound for λ. That is

$$\|\lambda\|_{V^*} = \inf\{C \in \mathbf{R} : \text{ for all } x \text{ in } V, |\lambda(x)| \leq C\|x\|_V\}.$$

Last time we showed that another way to define the operator norm $\|\cdot\|_{V^*}$ is by the formula

$$\|\lambda\|_{V^*} = \sup\{|\lambda(x)|_W : \|x\|_V = 1\|\}.$$

Let's now consider the dual space of the dual space. That is, V^* denotes the dual space of V and V^{**} denotes the dual space of the dual space of V. The dual space of the dual space of V is the set of bounded and linear functions $f : V^* \to \mathbf{R}$. That is, an element of V^{**} is a function f that associates each bounded and linear function $\lambda : V \to \mathbf{R}$ with a real number and that are themselves bounded and linear. The operator norm on the dual space of the dual space of V is defined as before except now instead of using the norm on V we use the operator norm on V^{**}. That is, $\|\cdot\|_{V^{**}}$ is defined for any bounded and linear function $f : V^* \to \mathbf{R}$ by the formula

$$\|f\|_{V^{**}} = \inf\{C \in \mathbf{R} : \text{for all } \lambda \text{ in } V^*, |f(\lambda)| \leq C\|\lambda\|_{V^*}\}$$

Another way to define this is by the formula

$$\|f\|_{V^{**}} = \sup\{|f(\lambda)| : \|\lambda\|_{V^*} = 1\}.$$

This second definition says that the operator norm on the dual space of the dual space of V is found by looking at the value of f on elements λ of the dual space of V that have at their smallest bound the number 1. Such functions λ are such that for all elements x of V we have $|\lambda(x)| \leq \|x\|_V$ and for some y in V we have $|\lambda(y)| = \|y\|_V$.

Consider the function that takes each element v of V to the function $e(v)$ which is an element of V^{**} defined by the formula $e(v)(\lambda) = \lambda(v)$. So this function takes as its input an element of V and gives as its output the map that will evaluate any bounded and linear function $\lambda : V \to \mathbf{R}$ at that element.

First of all, let's check that $e(v)$ is an element of the dual space of the dual space of V.

Is $e(v)$ a linear function? Let λ_1 and λ_2 be elements of V^*. We have

$$e(v)(\lambda_1 + \lambda_2) = (\lambda_1 + \lambda_2)(v) = \lambda_1(v) + \lambda_2(v).$$

Let α be a real number and let λ be an element of V^*. We have

$$e(v)(\alpha\lambda) = (\alpha\lambda)(v) = \alpha\lambda(v).$$

So e(v) is a linear function.

Is $e(v)$ bounded? First of all, suppose that $e(v)$ is bounded whenever $\|v\|_V = 1$. This means that there exists a constant C such that $|e(v)(\lambda)| \leq C\|\lambda\|_{V^*}$ for all $\lambda in V^*$. This can be rewritten as $|\lambda(v)| \leq C\|\lambda\|_{V^*} for all \lambda in V^*$. For an arbitrary v let $y = \frac{v}{\|v\|_V}$. Then we have that $|\lambda(y)| \leq C\|\lambda\|_{V^*}$ which implies that $|\lambda(v)| \leq (C\|v\|_V)\|\lambda\|_{V^*}$ so $e(v)$ is bounded.

Let's show that $e(v)$ is bounded whenever $\|v\|_V = 1$. So suppose v is an element of V such that $\|v\|_V = 1$. Consider $\|e(v)\|_{V^{**}}$. This is equal to $\sup\{|e(v)(\lambda)| : \|\lambda\|_{V^*} = 1\}$ which is equal to $\sup\{|\lambda(v)| : \|\lambda\|_{V^*} = 1\}$.

That $\|\lambda\|_{V^*} = 1$ means that $|\lambda(x)| \leq \|x\|_V$ for all x in V. In particular we have $|\lambda(v)| \leq 1$ and since this is true for all λ in V^* such that $\|\lambda\|_{V^*} = 1$ we have that $\|e(v)\|_{V^{**}} \leq 1$.

Therefore for each v in V we have that $e(v)$ is a bounded and linear function and so an element of the dual space of the dual space of V. That is, for all v in V the function $e(v)$ is an element of V^{**}.

What can we say about the function that maps each point in V to the element $e(v)$ of V^{**}? Let's make some claims.

Claim 1: The function $e : V \to V^{**}$ is injective. That is, if $e(v) = e(u)$, then $v = u$. Claim 2: The function $e : V \to V^{**}$ is an isometry onto its image. That is, for each v in V we have that $\|v\|_V = \|e(v)\|_{V^{**}}$. Claim 3: For any nonzero v in V there exists a function $\lambda : V \to \mathbf{R} in V^*$ such that $\|\lambda\|_{V^*} = 1$ and $\lambda(v) = \|v\|_V$. Claim 4: For any vector v in V such that $\|v\|_V = 1$ there exists λ in V^* such that $\|\lambda\|_{V^*} = 1$ and $\lambda(v) = 1$.

Today we will prove the Hahn-Banach Theorem which will show that all of these claims are true.

The Hahn-Banach Theorem: Let V be a normed vector space and V_0 a subspace of V. Any bounded and linear map $\lambda_0 : V \to \mathbf{R}$ can be extended to a bounded linear map $\lambda : V \to \mathbf{R}$ where the norm of the extension does not need to be any bigger than the norm of the linear map, that is, $\|\lambda\|_{V^*} = \|\lambda_0\|_{V_0^*}$.

Consider the function $e : V \to V^{**}$. Is this a linear function?

Let v and u belong to V. We have that $e(v + u)$ is the bounded and linear function in V^{**} that maps each element $\lambda of V^* to \lambda(v + u) = \lambda(v) + \lambda(u)$. This is the same as the function $e(v) + e(u)$.

Let α be a real number and let v be an element of V. We have that $e(\alpha v)$ maps each element $\lambda of V^* to \lambda(\alpha v) = \alpha\lambda(v)$. This is what the function $\alpha e(v)$ does.

So $e : V \to V^{**}$ is a linear function. Is it bounded? To show it is bounded we need to show that for each v in V there exists a number C such that $\|e(v)\|_{V^{**}} \leq C\|v\|_V$. This is the same as saying that the set of numbers $\|e(v)\|_{V^{**}}$ such that v is in V and $\|v\|_V = 1$ is bounded above. Suppose $\|v\|_V = 1$. We have that $\|e(v)\|_{V^{**}}$ is equal to $\sup\{|e(v)(\lambda)| : \|\lambda\|_{V^*} = 1\}$ and this is equal to $\sup\{|\lambda(v)| : \|\lambda\|_{V^*} = 1\}. That \|\lambda\|_{V^*} = 1$ means that $|\lambda(x)| \leq \|x\|_V$ for all x in V. Therefore $\|e(v)\|_{V^{**}} \leq 1$ which implies that e is bounded.

What is the relationship between the four claims and the Hahn-Banach Theorem?

First of all, I claim that Claim 3 and Claim 4 are equivalent. Since Claim 4 is a special case of Claim 3 we have that Claim 3 implies Claim 4. Now suppose that Claim 4 is true. Let's try to prove claim 4. Let v be a nonzero vector in V. Let y denote the vector $\frac{1}{\|v\|_V}v$. Then $\|y\|_V = 1$. So by Claim 4 there exists a bounded linear function $\lambda : V \to \mathbf{R}$ such that $\|\lambda\|_{V^*} = 1$ and $\lambda(y) = 1$. We can rewrite the last equation as $\lambda(v) = \|v\|_V$.

Now suppose that claim 2 is true. That is, suppose that the map $e : V \to V^{**}$ is an isometry onto its image which means that for each v in V we have $\|v\|_V = \|e(v)\|_{V^{**}}$. The right hand side of this is equal to $\sup\{|\lambda(v)| : \|\lambda\|_{V^*} = 1\}$ and it can be shown that this supremum is achieved. That is, there exists a function λ in

V^* with $\|\lambda\|_{V^*} = 1$ such that $\lambda(v) = \|v\|_V$. But this is Claim 3. Therefore Claim 2 implies Claim 3.

Does Claim 3 imply Claim 2? Suppose that λ is an element of V^* such that $\|\lambda\|_{V^*} = 1$. This means that $|\lambda(v)| \leq \|v\|_V$ for all v in V. Since $\|e(v)\|_{V^{**}}$ is equal to $\sup\{|\lambda(v)| : \|\lambda\|_{V^*} = 1\}$ we have that $\|e(v)\|_{V^{**}} \leq \|v\|_V$. By Claim 3 we have that there exists a $\lambda in V^*$ such that $\|\lambda\|_{V^*} = 1$ and $\lambda(v) = \|v\|_V$. Therefore $\|e(v)\|_{V^{**}} = \|v\|_V$. So Claim 3 does imply Claim 2.

So far we have shown that Claim 2 is equivalent to Claim 3 and that Claim 3 is equivalent to Claim 4.

Note that Claim 2 implies Claim 1 because any isometry is automatically injective. To see this suppose $v \neq u$. Then there is some positive distance between them. This implies there is a positive distance between $e(v)$ and $e(u)$ so that $e(v) \neq e(u)$.

Given that e is injective, if Claim 2, were true then certainly we could prove that Claim 1 implies Claim 2. I'm not sure whether this means that Claim 1 implies Claim 2. When are two true theorems equivalent?

What is the relationship between Claims 2,3, and 4 and the Hahn-Banach Theorem? Let's suppose that the Hahn-Banach Theorem is true. Let's try to prove that Claim 4 is true.

Let v be a vector in V such that $\|v\|_V = 1$. Consider the subspace V_0 of V consisting of all scalar multiples of v. Let $\lambda_0 : V \to \mathbf{R}$ be defined by the formula $\lambda(x) = \alpha$ where α is the unique number such that $x = \alpha v$. This is clearly a linear function. Is it bounded? It is bounded if there exists a constant C such that $|\lambda(x)| \leq C\|x\|$ for each $x \in V_0$. We get equality by choosing $C = 1$ so $\|\lambda\|_{V^*} = 1$. By the Hahn-Banach Theorem there exists a bounded linear map $\lambda : V \to \mathbf{R}$ that is an extension of λ_0 and is such that $\|\lambda_0\|_{V_0^*} = \|\lambda\|_{V^*}$. That is $\|\lambda\|_{V^*} = 1$. We have that $\lambda(v) = \lambda_0(v) = 1$ because λ is an extension of λ_0.

This shows that if the Hahn-Banach Theorem is true then Claim 4 is true. What about the reverse. I'm not sure about the reverse.

Let's start by proving Claim 4 for finite dimensional vector spaces.

Claim 4: If $v \in V$ such that $\|v\|_V = 1$, then there exists $\lambda \in V^*$ such that $\lambda(v) = 1$ and $\|\lambda\|_{V^*} = 1$.

Proof: Let's prove this by induction.

Case 1: Suppose that the dimension of V is equal to the number 1. Let v be in V be such that $\|v\|_V = 1$. Define $\lambda : V \to \mathbf{R}$ by $\lambda(v) = 1$ and for x in V there exists a unique number α such that $x = \alpha v$, define $\lambda(x) = \alpha$. We have shown previously that such a λ is a bounded and linear function from V to $\mathbf{R}$ such that $\|\lambda\|_{V^*} = 1$.

Case 2: Suppose that the dimension of V is equal to the number 2. Let v be a point in V such that $\|v\|_V = 1$. Let V_0 be the one dimensional subspace of V consisting of all scalar multiples of v. By Case 1 there exists a bounded and linear function $\lambda_0 : V_0 \to \mathbf{R}$ such that $\|\lambda_0\|_{V^*} = 1$ and $\lambda_0(v) = 1$. Note that λ_0 is the only choice with these properties. Since the dimension of V is equal to the number 2 there exists a vector y that is in V but not in V_0. We can assume that $\|y\|_V = 1$ To define an extension $\lambda : V \to \mathbf{R}$ of λ_0 to the whole space we only need to choose the number t such that $\lambda(y) = t$. We want to have $\|\lambda\|_{V^*} = 1$. This implies that we need $|\lambda(y)| \leq 1$. We need more than this though. We need that for every x in V such that $\|x\|_V = 1$ we have $|\lambda(x)| \leq 1$. What we want is to be able choose a number t such that this is true.

Each point x such that $\|x\|_V = 1$ can be written as a $v + by$ where both a and b belong to $[-1, 1]$. We know that $1 = \|x\|_V \leq |a| + |b| \leq 2$. What we need to do is to choose t such that whenever $\|x\|_V = 1$ we have $|\lambda(x)| \leq 1$. We have that $\lambda(x) = a + bt$. We want the following two inequalities to hold: $-1 \leq a + btanda + bt \leq 1$. these can be rewritten as $\frac{-1-a}{b} \leq t$ and $t \leq \frac{1-a}{b}$. It can be shown that there is a t that satisfies these inequalities.

QED

Not that in this proof we have proved the Hahn-Banach Theorem when the vector space V is finite dimensional. Next time we will prove the Hahn-Banach Theorem when V is possible infinite dimensional.

Recall the Hahn-Banach Theorem: Let V_0 be a subspace of a normed vector space V. Any bounded and linear function $\lambda_0 : V_0 \to \mathbf{R}$ can be extended to a bounded and linear function $\lambda : V \to \mathbf{R}$ such that $\|\lambda\|_{V^*} = \|\lambda_0\|_{V_0^*}$.

We saw last time that this theorem has various implications. For example it implies that the function $e : V \to V^{**}$ defined by $e(v)(\lambda) = \lambda(v)$ is an isometry. That is $\|v\|_V = \|e(v)\|_{V^{**}}$. It also implies the equivalent statement that for each v in V such that $\|v\|_V = 1$ there exists a λ in V^* such that $\|\lambda\|_{V^*} = 1$ and $\lambda(v) = \|v\|_V$. Last time we proved this and in the process of proving it we proved the Hahn-Banach theorem when V has finite dimension.

Today we will prove the Hahn-Banach theorem when V possibly has infinite dimension. Our approach will be the following. Let P be the set of ordered pairs (W, g) where W is a subset of V that contains V_0 and $g : W \to \mathbf{R}$ is a bounded an linear function that is an extension of λ_0 such that $\|g\|_{W^*} = \|\lambda_0\|_{V_0^*}$.

Let's define the following ordering $<_P$ on the set P. For any two elements (W, g) and (W', g') of P let's say that $(W, g) <_P (W', g')$ to mean that $W \subsetneq W'$ and g' is an extension of g.

Suppose we could prove the existence of an element (W', g') in P such that for all (W, g) in P not equal to (W', g') it is not true that $(W', g') <_P (W, g)$.

This must mean that $W' = V$. If it did not mean this then we could find a vector v in V that is not in W'. We could then consider the set smallest subspace W of V containing $W' \cup \{v\}$ and construct using the procedure of last time a bounded and linear function $g : W \to \mathbf{R}$ such that $\|g\|_{W^*} = \|\lambda_0\|_{V_0}$. But this means that (W, g) is in P and $(W', g') <_P (W, g)$.

So our goal for the proof of the Hahn-Banach Theorem is to prove the existence of an element (W', g') in P such that for all (W, g) in P it is not true that $(W', g') <_P (W, g)$. We will call such an element (W', g') a maximal element of P.

Thus our goal is to prove the existence of a maximal element in the set P. Let's do this in a lemma.

Definition: A partially ordered set is a set P with a binary relation $<_P$ that is reflexive (not $x <_P x$ for all $x \in P$), transitive ($x <_P y$ and $y <_P z$ implies $x <_P z$), and asymmetric (if $x <_P y$, then not $y <_P x$.)

You can see that P with the binary relation $<_P$ is a partially ordered set.

Definition: A linearly ordered set is a set P with a binary relation $<_P$ such that P with the binary relation $<_P$ is a partially ordered set and such that $<_P$ is complete (for all x, y are distinct elements of P, then $x <_P y$ or $y <_P x$).

You can see that P with the binary relation $<_P$ is not a linearly ordered set.

Definition: A partially ordered set P is well ordered if each nonempty subset Q of P has a least element: that is, an element x in Q such that $x <_P q$ for each q in Q.

Zorn's Lemma: Let P be a partially ordered set such that each linearly ordered subset of P has an upper bound. Then P has a maximal element.

Proof: Suppose that P is a partially ordered set such that each linearly ordered subset of P has an upper bound and that P has no maximal element.

Let Q be a linearly ordered subset of P. We have assumed that Q has an upped bound: that is, there exists an element $\lambda(Q)$ in P such that $q <_P \lambda(Q)$ for each q in Q.

Let's say that a subset $Q \subseteq P$ is a good chain if Q is well-ordered and each element x in Q satisfies the formula $x = \lambda(\{q \in Q : q <_P x\})$.

Note that there is no largest good chain in P. To see this let Q be a good chain in P. Consider the set $Q \cup \lambda(Q)$. This set is well ordered. Let x be in $Q \cup \lambda(Q)$. If $x \in Q$ we have that $x = \lambda(\{q \in Q : q <_P x\})$ because Q is a good chain. Otherwise, $x = \lambda(Q)$. Therefore $Q \cup \lambda(Q)$ is also a good chain. Since any singleton set is a good chain this implies that there is no largest good chain in P.

Now I claim that if Q and Q' are both good chains in P then exactly one of the following conditions holds:

(1) $Q = Q'$. (2) There exists an element $q_0 \in Q$ such that $Q' = \{q \in Q : q <_P q_0\}$. (3) There exists an element $q_0' \in Q'$ such that $Q = \{q' \in Q' : q' <_P q_0'\}$.

Note that the three possibilities are mutually exclusive because (2) implies that $Q \subsetneq Q'$ and (3) that $Q' \subsetneq Q$.

What are the alternatives? It may be that there is an element q in Q that is not in Q' and that there is an element q' in Q' that is not in Q. Let's choose q in $Q$and not in Q' such that if x is in Q and $x <_P q$, then x is in Q'.

Consider the set $\{x \in Q : x <_P q\}$. Call this set W. If W is equal to Q' we are done. But we know this is not the case because there exists a q' in Q' that is not in Q. So it must be that $W \subsetneq Q'$. Let c denote the least element of $Q' - W$. Then we have that W is equal to $\{x \in Q' : x <_P c\}$. Since both Q and Q' are good chains we have that $c = \lambda(W)$ and that $d = \lambda(W)$ so that q actually does belong to Q'. A contradiction.

I now claim that if $\{Q_\alpha\}$ is a collection of good chains, then the union $\bigcup Q_\alpha$ is also a good chain. We have to show that $\bigcup Q_\alpha$ is a well-ordered set and that each element x in $\bigcup Q_\alpha$ satisfies the formula $x = \lambda(\{q \in \bigcup Q_\alpha : q <_P x\})$.

To show that $\bigcup Q_\alpha$ is well ordered we need to show that it is a linearly ordered set and that each nonempty subset of $\bigcup Q_\alpha$ has a least element.

The set $\bigcup Q_\alpha$ is a partially ordered set because it is a subset of P and P with the binary relation $<_P$ is a partially ordered set. To show that $\bigcup Q_\alpha$ is a totally ordered set we just need to show that if q and q' are distinct elements of $\bigcup Q_\alpha$, then $q <_P q'$ or $q' <_P q$. There exists a good chain Q such that q is an element of Q. There exists a good chain Q' such that q' is an element of Q'. We have shown that one of these must be a subset of the other so for one of these both q and q' are elements. Since both Q and Q' are good chains this implies that $q <_P q'$ or $q' <_P q$. Therefore $\bigcup Q_\alpha$ is a linearly ordered set.

Let's show that each nonempty subset W of $\bigcup Q_\alpha$ has a least element. A least element of W is an element x in W such that for all y in $W - \{x\}$ we have $x <_P y$. It is not clear to me that the $\bigcup Q_\alpha$ is a well-ordered set. Apparently from here it can be shown that it is.

If $\bigcup Q_\alpha$ is well-ordered for each x in $\bigcup Q_\alpha$ we have that $x = \lambda(\{q \in \bigcup Q_\alpha : q <_P x\})$. The reason this is true is because $\bigcup Q_\alpha - \{q \in \bigcup Q_\alpha : q <_P x\}$ is a nonempty subset of $\bigcup Q_\alpha$ and so has a least element x.

If this is true then the union of all good chains is the largest good chain. But we have shown that there is no largest good chain. This contradiction completes the proof of Zorn's lemma.

QED

Now let's prove the Hahn-Banach Theorem.

Proof of the Hahn-Banach Theorem: As we defined it before the proof of Zorn's Lemma the set P with the binary relation $<_P$ is a partially ordered set. Suppose that Q is a linearly ordered subset of P. We would like to show that there exists an upper bound for Q. We have that $Q = \{(W_\alpha, g_\alpha)\}_\alpha$ where each W_α is a subspace of V that contains V_0 and each $g_\alpha : W_\alpha \to \mathbf{R}$ is a bounded and linear function that is an extension of λ_0 and is such that $\|g_\alpha\|_{W_\alpha^*} = \|\lambda_0\|_{V_0^*}$. Let's define W to be $\bigcup_\alpha W_\alpha$ and define g by saying that $g(x) = g_\alpha(x)$ whenever $x \in W_\alpha$. Note that the definition for g makes sense because Q is linearly ordered. If x happens to belong to W_{α_1} and W_{α_2}, then since Q is linearly ordered it must be that one of these is a strict subset of the other and also that $g_{\alpha_1}(x) = g_{\alpha_2}(x)$.

First note that W is a subspace of V that contains V_0. It is clear that it is contained in V and that it contains V_0 because each term of the union of which it is made does. But is it also a vector space: it has a zero element; if x and y belong to W then $x + y$ is in W because Q is linearly ordered there exists a set W_α that contains both and so contains $x+y$; if $c \in \mathbf{R}$ and x is in W then cx is in W because x belongs to some vector space W_α and so this vector space also contains cx.

Let's now show that $g : W \to \mathbf{R}$ is a bounded and linear function. Let x and y belong to W then since Q is totally ordered there exists a W_α that contains both of the points x and y and so the point $x + y$. Then

$$g(x + y) = g_\alpha(x + y) = g_\alpha(x) = g_\alpha(y) = g(x) + g(y)$$

. Not let c be a real number and x an element of W. Then there exists a vector space W_α that contains x and so also contains cx. Then $g(cx) = g_\alpha(cx) = cg_\alpha(x) = cg(x)$. So g is a linear function.

Let's show that g is bounded. We have for each α that $\|g_\alpha\|_{W_\alpha^*} = \|\lambda_0\|_{V_0^*}$. Suppose there were to exist an x in W such that $\|x\|_V = 1$ and $|g(x)| > \|\lambda_0\|_{V_0^*}$. There exists an α such that x belongs to W_α and so $g(x) = g_\alpha(x)$. We then have that $|g_\alpha(x)| > \|\lambda_0\|_{V_0^*}$ which means that $\|g_\alpha\|_{W_\alpha^*} > b\|\lambda_0\|_{V_0^*}$ which is not so.

Therefore $g : W \to \mathbf{R}$ is a linear and bounded function that is an extension of $\lambda_0 : V_0 \to \mathbf{R}$ and is such that $\|g\|_{W^*} = \|\lambda_0\|_{V_0^*}$.

This shows that (W, g) is an upper bound for Q.

By Zorn's Lemma the set P has a maximal element (M, λ) and it must be that $M = V$ otherwise we could find a larger element using the procedure in the proof of yesterday.

QED

Here is an application of the Hahn-Banach Theorem: Let E be a measurable subset of $\mathbf{R}^n$ with finite Lebesgue measure. Let V_0 denote the set of all bounded and measurable real valued functions on E. Let V denote the set of measurable functions on E. Then V is a vector space and V_0 is a subspace of V. Let's turn V into a normed vector space by giving it the supremum norm: that is, if f is in V, then the supremum norm of f is the number $\|f\|$ defined by the formula $\|f\| = \sup\{|f(x)| : x \in E\}$. Now think about the Lebesgue integral as a real valued function on V_0. This function is linear. It is also bounded because if f belongs to V_0 and $\|f\| = 1$, then $\int f \leq \mu(E)$. In other words the operator norm of the Lebesgue integral on V_0 is equal to $\mu(E)$: $\|\int\|_{V_0^*} = \mu(E)$. Recall The Hahn-Banach Theorem: Let V be a normed vector space and $V_0 \subseteq V$ a subspace. Any bounded and linear function $\lambda_0 : V_0 \to \mathbf{R}$ can be extended to a bounded and linear function $f : V \to \mathbf{R}$ such that $\|f\|_{V^*} = \|\lambda_0\|_{V_0^*}$.

So by The Hahn-Banach Theorem the Lebesgue integral $\int$ on the set of bounded and measurable real valued functions on E can be extended to a bounded and linear function $\hat{\int}$ on the set of all measurable functions in such a way $\|\hat{\int}\|_{V^*} = \|\int\|_{V_0^*}$.

22. THE DUAL OF L^P AND THE HAHN DECOMPOSITION THEOREM

Let E be a measurable subset of $\mathbf{R}^n$ with finite measure. Recall that $L^1(E)$ is the set of measurable functions $f : E \to \mathbf{R}$ such that $\|f\|_{L^1} = \int |f| < \infty$. Recall that the $L^1(E)^*$ is the set of bounded and linear functions $\lambda : L^1(E) \to \mathbf{R}$ where bounded is defined using the operator norm $\|\lambda\|_{L^{1*}}$ which is the smallest number C such that —$\lambda(f)| \leq C\|f\|_{L^1}$. Also, recall that the operator norm of a function $\lambda in L^1(E)^*$ can be written as

$$\sup\{|\lambda(f)| : f \in L^1(E) \text{ and } \|f\|_{L^1} = 1\}.$$

Let $\lambda \in L^1(E)^*$. Today we will study this function. It turns out that there exists a measurable function $f : E \to \mathbf{R}$ such that for each g in $L^1(E)$ we have $\lambda(g) = \int gf$. A similar result will apply for $L^p(E)$ where $p > 1$.

The function λ is continuous (recall that a linear function is bounded if and only if it is continuous). Note that if S is a measurable subset of E then χ_S belongs to $L^1(E)$. Suppose we knew the value of $\lambda(\chi_S)$ for each measurable subset S of E. Because λ is linear, if g is a simple function (and so is in $L^1(E)$), then we know the value of $\lambda(g)$. For an arbitrary function g in $L^1(E)$ we can construct a sequence of simple functions $g_1, g_2, ...$ that converge to g in the L^1 norm. By the continuity of λ we have that $\lim_i \lambda(g_i) = \lambda(g)$.

So we only need to know the value of λ on χ_S when S is a measurable subset of E to know what value λ takes at an arbitrary element of $L^1(E)$. Notice that this is kind of like a measure. We want to assign to each measurable subset S of E a number $\lambda(\chi_S)$. The major difference here is that $\lambda(\chi_S)$ may be negative.

Does there exist a number M such that $-M \leq \lambda(\chi_S) \leq M$ for all measurable subsets S of E? Since λ is bounded there exists a number C such that $|\lambda(\chi_S)| \leq C\int \chi_S \leq C\mu(E)$. So the answer to this question is "yes". For instance $M = C\mu(E)$ works.

Let's introduce the following definition. Let E be a measurable subset of $\mathbf{R}^n$. A finite signed measure is a function ν from the collection of measurable subsets of E to the real numbers that satisfies the following two properties:

(1) There exists a number M such that $-M \leq \nu(S) \leq M$ for each measurable subset S of E. (2) If $S_1, S_2, ...$ are disjoint and measurable subsets of E, then $\nu(S_1 \cup S_2 \cup ...) = \nu(S_1) + \nu(S_2) + ...$

Note that $|\nu(S_1)| + |\nu(S_2)| + ... \leq 2M$ so that the series $\nu(S_1) + \nu(S_2) + ...$ converges absolutely and so converges.

I claim that the function ν defined by $\nu(S) = \lambda(\chi_S)$ is a finite signed measure. We have already proved property (1) holds for $M = C\mu(E)$. Let's show that property (2) holds too. Let $S_1, S_2, ...$ be a sequence of measurable and disjoint subsets of E. For each positive integer n we have that

$$\nu(\bigcup_{i=1}^{n} S_i) = \lambda(\chi_{\bigcup_{i=1}^{n} S_i}) = \lambda(\sum_{i=1}^{n} \chi_{S_i}) = \sum_{i=1}^{n} \nu(S_i).$$

The continuity of λ gives the result.

Let's call ν positive if it is nonnegative and let's call it negative if it is nonpositive.

Let's try to prove the following theorem:

The Hahn Banach Theorem: Let ν be a finite signed measure on E. Then E can be written as the disjoint union of the sets E_+ and E_- such that the restriction of ν to E_+ is the positive finite signed measure ν_+ and the restriction of ν to E_- is the negative finite signed measure ν_-.

Proof: I claim that if $S_1, S_2, ...$ are measurable subsets of E such that $\nu|_{S_i}$ is positive for each $i = 1, 2, ...$, then $\nu|_{\bigcup S_i}$ is positive.

To see this let S be a subset of $\bigcup S_i$. Then by the countable additivity of ν we have that

$$\nu(S) = \nu(S \cap S_1) + \nu(S \cap (S_2 - S_1)) + \nu(S \cap (S_3 - (S_2 - S_1))) + ...$$

where each term in this sum is a positive number because each $\nu|_{S_i}$ is positive. This implies that $\nu(S)$ is a positive number so that $\nu|_{\bigcup S_i}$ is positive.

Consider all real numbers of the form $\nu(S)$ where $\nu|_S$ is positive. This set is nonempty (e.g. you can take S to be the empty set). It is bounded below because each number in the set is nonnegative and it is bounded above because ν is a finite signed measure. So we can let $x = \sup\{\nu(S) \mid S \subseteq E \text{ and } \nu| \text{ is positive}\}$. $S_1, S_2, ...$ are measurable subsets of E such that $\nu|_{S_i}$ is positive for each $i = 1, 2, ...$ and such that the sequence of numbers $\nu(S_1), \nu(S_2), ...$ converges to x. Let $E_+ = \bigcup S_i$. By our claim we know that $\nu|_S$ is positive. We also have that for each i, $x \geq \nu(E_+) \geq \nu(S_i)$. This implies that $\nu(E_+) = x$ and we have that $v|_{E_+}$ is positive. Note that E_+ is also the union of all subsets S of E such that $\nu|_S$ is positive.

Let's E_- denote the set $E - E_+$. Let's show that $\nu|_{E_-}$ is negative. Suppose not. Then there exists a subset S of E_- such that $\nu(S) > 0$. I will show that we can find a subset S' of this set such that $\nu|_{S'}$ is positive. This would be a contradiction because E_+ is the union of all sets on which ν is positive. Define

$$\text{Badness}(S) = \sup t \geq 0 : B \subseteq S \text{ and } t = -\nu(B).$$

Choose a set B_1 such that $\nu(B_1) \leq \frac{-\text{Badness}(S)}{2}$. Choose a set B_2 such that $\nu(B_2) \leq \frac{-\text{Badness}(S-B_1)}{2}$. And so on. Because ν is countably additive we have that the badness of S is no more than half the badness of $S - B_1$. This is because we removed at least half the badness of S when we removed B_1. Likewise we have that $\text{Badness}(S - B_1 - B_2 - ... - B_n) \leq \frac{\text{Badness}(S)}{2^n}$ so that the badness of $S - B_1 - B_2 - ... - B_n$ converges to 0. Let $S' = S - B_1 - B_2 - ....$ We have that $\text{Badness}S' = 0$ which implies that $\nu|_{S'}$ is positive. Note also that $\nu(S') > \nu(S) > 0$. Therefore where is no subset S of E_- such that $\nu(S) > 0$. This implies that $\nu|_{E_-}$ is negative.

QED

23. THE RADON-NIKODYM THEOREM AND THE HAHN-BANACH THEOREM

We have proved The Hahn Decomposition Theorem. Let's use it to understand the relationship between Lebesgue measure μ and some finite signed measure ν. Let E be a measurable subset of $\mathbf{R}^n$ that has finite Lebesgue measure and suppose that ν is a finite signed measure on E.

Let t be a real number and consider the the function given by the formula $\nu - t\mu$. Note that because the Lebesgue measure on E is a finite signed measure this is a finite signed measure. In face, we can make the following observation.

Observation: the set of finite signed measures on E is a vector space.

Let's apply The Hahn Decomposition Theorem to $\nu - t\mu$. This says that there exists a set E_t such that $(\nu - t\mu)|_{E_t}$ is positive and $(\nu - t\mu)|_{E-E_t}$ is negative.

Let's write out what this means. If S is a measurable subset of E_t, then $\nu(S) \geq t\mu(S)$. And if S is disjoint from E_t, then $\nu(S) \leq t\mu(S)$.

You can now see how The Hahn Decomposition Theorem can tell us something about the relationship between ν and μ.

Question: What we would like to do now is to find the relationship between E_t and $E_{t'}$ when $t' > t$.

One would expect that as t gets bigger the set on which $\nu - t\mu$ is positive becomes smaller (this is because μ is positive on all of E). So we might expect that $E_{t'} \subseteq E_t$.

This is almost true. The reason it is not just plain true is that the Hahn-Decomposition is not unique. We can always move countably many sets of measure zero between the sets E_+ and $E - E_+$ without affecting the positivity or negativity of the measure there. It is almost true because we cannot move any set of nonzero measure between E_+ and $E - E_+$ without breaking the positivity or negativity of the measure.

What to do? It will be convenient for us to only consider the case where t is a rational number. Define the set F_t by the formula $F_t = \bigcup_{t' \geq t, t' \in \mathbf{Q}} E_{t'}$.

Note that $(\nu - t\mu)|_{F_t}$ is positive and that $(\nu - t\mu)_{E-F_t}$ is negative.

Now the answer to the question is "yes": if $t' > t$, then $F_{t'} \subseteq F_t$.

So let's suppose that E_t actually denotes F_t and continue our discussion using the notation E_t.

Let's now define the function $f : E \to \mathbf{R}$ by the formula $f(x) = \sup\{t \in \mathbf{Q} : x \in E_t\}$. Does this formula make any sense? Could it be that this supremum does not exist? There are two cases in which this supremum might not exist.

Case 1: $x \in E_t$ for all t. That is, $x \in \bigcap E_t$. I claim that $\bigcap E_t$ is a set of measure zero.

For each rational number t we have that $\nu(\bigcap E_t) \geq t\mu(\bigcap E_t)$. Now note that if $\mu(\bigcap E_t) > 0$ then $\nu(\bigcap E_t) = \infty$ and this if impossible because ν is a finite signed measure. So $\bigcap E_t$ is a set of Lebesgue measure zero.

Case 2: $x \notin E_t$ for all t. That is, $x \in E - \bigcup E_t$.

This implies that for each rational number t we have $\nu(E - \bigcup E_t) \leq t\mu(E - \bigcup E_t)$.

Let's apply the same argument as before. Suppose that $\mu(E - \bigcup E_t) > 0$. Then it must be that $\nu(E - \bigcup E_t) = -\infty$ and this contradicts that ν is a finite signed measure. Therefore $E - \bigcup E_t$ is a set of Lebesgue measure zero.

In all other cases the function is defined because the set $\{t \in \mathbf{Q} : x \in E_t\}$ is nonempty and it does not contain all rational numbers t. Since it does not contain all rational numbers t there exists a rational number t' such that $x \notin E_{t'}$. As we have shown above this implies that $x \notin E_t$ for all $t > t'$. Therefore t' is an upper bound for the set $\{t \in \mathbf{Q} : x \in E_t\}$. These two things implies that the supremum of this set exists.

What does this mean? It means that the function f is defined except perhaps on a subset Z of E that has Lebesgue measure zero. So let's redefine the domain of f to be the set $E - Z$.

I now claim that f is a measurable function.To see this note that we can define f using the following formula:

$$f(x) = \sup\{t\chi_{E_t}(x) : t \in \mathbf{Q}\}.$$

That is, we can define f the be the pointwise supremum of measurable functions. We have previously shown that the pointwise supremum of measurable functions is a measurable function. Therefore f is a measurable function.

Our goal is to see what we can say about the relationship between ν and μ.

Consider the set $Y = \{x \in E : a < f(x) < b\}$. This is a subset of $E - Z$. We know that if x belongs to Y, then $a < f(x) < b$. This means that $x \notin E_a$ and that $x \in E_b$. So Y is a subset of $E_b - E_a$.

This implies that if S is a measurable subset of Y, then S is a subset of E_b so that $\nu(S) \geq b\mu(S)$, and S is a subset of $E - E_a$ so that $\nu(S) \leq a\mu(S)$. That is, if S is a measurable subset of Y, then $b\mu(S) \leq \nu(S) \leq a\mu(S)$. We now have an estimate for ν in terms of μ.

How can we make this estimate more precise? Easy. Just take a and b to be closer to one another.

Now suppose that $g : E - Z \to \mathbf{R}$ is a simple function. This means that $g(E - Z)$ is a finite set $\{v_1, v_2, ..., v_k\} \subseteq \mathbf{R}$. Then we can write $E - Z = Y_1 \cup Y_2 \cup ... \cup Y_k$ where Y_i is the subset of $E - Z$ on which g takes the value v_i.

Suppose that $f \geq g$. Then we have that $\int f d\mu \geq \int g d\mu = \sum_i v_i \mu(Y_i)$. But since $f \geq g$ we have that $f(x) \geq v_i foreachxinY_i$. This implies that $Y_i \cap E_{v_i} = \emptyset$ which implies that $\nu(Y_i) \leq v_i \mu(Y_i)$. Therefore

$$\int f d\mu \geq \sum_i \nu(Y_i) = \nu(E - Z).$$

Now suppose that $f \leq g$. Then we have that $\int f d\mu \leq \int g d\mu = \sum_i v_i \mu(Y_i)$. Since $f \leq g$ we have that $f(x) \leq v_i$ for each $x \in Y_i$. This implies that

$$Y_i \subseteq E_{v_i} = \emptyset sothat \nu(Y_i) \geq v_i \mu(Y_i).$$

Therefore

$$\int f d\mu \leq \sum_i \nu(Y_i) = \nu(E - Z).$$

Therefore

$$\nu(E - Z) = \int f d\mu.$$

Exactly the same logic applies if S is a measurable subset of E. Then we could define a simple function $g : S - Z \to \mathbf{R}$ and the previous argument would show that

$$\nu(S - Z) = \int f \chi_S d\mu.$$

In general we have proved that if ν is a measurable subset of E, then there exists a set Z of Lebesgue measure zero such that

$$\nu(S) = \nu(S \cap Z) + \int f \chi_S d\mu$$

This is called The Radon-Nikodym Theorem. Let's state it as a theorem.

The Radon-Nikodym Theorem: Let E be a measurable subset of $\mathbf{R}^n$ and ν a finite signed measure on E. Then there exists a subset Z of E with Lebesgue measure zero and a measurable function $f : E - Z \to \mathbf{R}$ such that

$$\nu(S) = \nu(S \cap Z) + \int f \chi_S d\mu$$

for all measurable subsets S of E.

In our proof we assumed that E had finite Lebesgue measure. To prove the case where Ehas infinite Lebesgue measure just cut E up into countably many sets of finite Lebesgue measure, apply our argument to each piece.

24. THE CHANGE OF VARIABLES FORMULA

Recall The Radon-Nikodym Theorem: Let E be a measurable subset of $\mathbf{R}^n$ and let ν be a finite signed measure on E. Then there exists a subset Z of E which is a set of Lebesgue measure zero and a measurable function $f : E - Z \to \mathbf{R}$ such that if S is a measurable subset of E, then

$$\nu(S) = \nu(S \cap E) + \int f d\mu.$$

The reason this theorem is nice is that it tells us the relationship between any finite signed measure and Lebesgue measure.

Definition: A finite signed measure is absolutely continuous with respect to Lebesgue measure if whenever a set has Lebesgue measure zero it also has ν measure zero.

So for a finite signed measure that is absolutely continuous with respect to μ the conclusion of the Radon-Nikodym theorem is that if S is a measurable subset of E, then $\nu(S) = \int f d\mu$.

Now let $g : E \to \mathbf{R}$ be a simple function with image $v_1, v_2, ..., v_k$ on the sets $Y_1, Y_2, ..., Y_k$. Then

$$\int g d\nu = \sum_{i=1}^{k} v_i \nu(Y_i) = \sum_{i=1}^{k} v_i \int \chi_{Y_i} f d\mu = \int g f d\mu.$$

This can be generalised so that if g is a measurable function (I suppose with respect to both ν and Lebesgue measure), then

$$\int g d\nu = \int g f d\mu.$$

This is quite powerful. For instance, in probability we may have a probability measure with which we want to integrate random variables to compute their expected value. The Radon-Nikodym Theorem tells us that we can turn this integral into a Lebesgue integral.

Here's another application of The Radon-Nikodym Theorem. Suppose that U is a bounded and open subset of $\mathbf{R}^n$ and that $h : U \to V$ is a bijection. In this situation we can define a measure ν on U by the formula $\nu(S) = \mu(h(S))$ for each subset S of U. Note that the measure ν might not be absolutely continuous with respect to μ. This will happen if the function h maps a subset of E that has positive Lebesgue measure to a set that has zero Lebesgue measure. Let $g : V \to \mathbf{R}$ be a simple function. Then the image of g is finite and we have that $g(x) = \sum_{i=1}^{n} v_i \chi_{Y_i}(x)$

where Y_i is the set of points on which the function g takes the value v_i. We have that

$$\int g d\mu = \sum_{i=1}^{n} v_i \mu(Y_i).$$

Since the function h is a bijection we have that $g \circ h : U \to \mathbf{R}$ is a simple function that can be written as $\sum_{i=1}^{n} v_i \chi_{h^{-1}(Y_i)}$. This implies that

$$\int g \circ h d\nu = \sum_{i=1}^{n} v_i \nu(h^{-1}(Y_i)) = \sum_{i=1}^{n} v_i \mu(Y_i).$$

Therefore

$$\int_V g d\mu = \int_U g \circ h d\nu.$$

Since ν is a finite signed measure and assuming it is absolutely continuous The Radon-Nikodym Theorem implies there exists a set of measure zero Z and a function $f : U - Z \to \mathbf{R}$ such that

$$\int_U g \circ h d\nu = \int_{U-Z} (g \circ h) f d\mu.$$

Therefore

$$\int_V g d\mu = \int_{U-Z} (g \circ h) f d\mu.$$

What we have done is successfully changed the domain of integration. This is called The Change of Variables Formula. It is possible to write down an explicit formula for f in terms of the function h.

Recall that $f(x)$ is defined as the largest number t such that the largest set E_t where $(\nu - t\mu)|_{E_t}$ is positive contains the point x. In this case we have that $\nu = \mu \circ h$. It turns out that $f(x)$ is given by the absolute value of the determinate of the derivative of h evaluated at the point x. So The Change of Variables Formula is

$$\int_V g d\mu = \int_{U-Z} (g \circ h) |\det(Dh(x))| d\mu.$$

25. THE RADON-NIKODYM THEOREM AND THE DUAL OF L^P

Let's return to the task of trying to understand the dual space of $L^p(E)$. Recall that the set $L^p(E)$ is the set of functions $f : E \to \mathbf{R}$ (where E is a measurable subset of $\mathbf{R}^n$) that have finite L^p norm. The L^p norm is defined by the formula $\left(\int |f|^p\right)^{\frac{1}{p}}$. We have previously shown that $L^p(E)$ with this norm is a Banach space which means that it is a normed vector space that is complete. Recall that the dual of $L^p(E)$ denoted $L^p(E)^*$ is the set of linear and bounded functions $\lambda : L^p(E) \to \mathbf{R}$. Recall that a linear function $\lambda : L^P(E) \to \mathbf{R}$ is bounded if there exists a number C such that $|\lambda(f)| \leq C\|f\|_{L^p}$ for all f in $L^p(E)$ (recall that this definition was motivated by Lipschitz continuity and in fact a linear function is bounded if and only if it is continuous). The smallest number C that works in this inequality defines the operator norm which is the norm we use on the dual of $L^p(E)$. The operator norm of λ can also be expressed as the magnitude of the largest number that λ takes on elements of $L^p(E)$ with L^p norm 1. That is, $\|\lambda\|_{L^{p*}} = \sup\{|\lambda(f)| : \|f\|_{L^p} = 1\}$. Before we went and proved The Hahn-Decomposition Theorem and The Radon-Nikodym Theorem we defined the idea of a finite signed measure and this was motivated by the function that sends a measurable subset S of E to the number $\lambda(\chi_S)$. Let's get back to this. We will assume that the Lebesgue measure of E is finite.

Call this function ν. I claim that ν is a finite signed measure and that ν is absolutely continuous with respect to μ. To see this note that λ is bounded so there exists a number C such that for all measurable subsets S of E we have that

$$|\nu(S)| = |\lambda(\chi_S)| \leq C\|\chi_S\|_{L^p} = C\left(\int \chi_S^p\right)^{\frac{1}{p}} = \mu(S)^{\frac{1}{p}}.$$

So ν satisfies the 'finite' part of being a finite signed measure. Note that this inequality also implies ν is absolutely continuous with respect to μ (that is, $\nu(S) = 0$ whenever $\mu(S) = 0$) To show that ν is a finite signed measure we just need to show that it is countably additive. Let $S_1, S_2, ...$ be disjoint measurable subsets of E. We want to show that $\nu(S_1 \cup S_2 \cup ...) = \nu(S_1) + \nu(S_2) + ...$ That is we want to show that $\lambda(\chi_{\bigcup_{i=1}^{\infty} S_i}) = \lim_n \lambda(\chi_{\bigcup_{i=1}^{n} S_i})$. Since λ is bounded it is continuous so we have that the right hand side of this is equal to $\lambda(\lim_n \chi_{\bigcup_{i=1}^{n} S_i})$. If we can show that $\lim_n \chi_{\bigcup_{i=1}^{n} S_i}$ converges to $\chi_{\bigcup_{i=1}^{\infty} S_i}$ we will be done. But we have to be a bit careful here. Remember our purpose. We will eventually want to approximate general functions in $L^p(E)$ by simple functions. The notion of convergence in which we will approximate these functions is given by the L^p norm. We would use pointwise convergence if we wanted to approximate the functions pointwise. It is not completely clear to me why we are not doing this yet. In any case we have that

$$\|\chi_{\bigcup_{i=1}^{\infty} S_i} - \chi_{\bigcup_{i=1}^{n} S_i}\|_{L^p} = \|$$

$$\chi_{\bigcup_{i>n} S_i}\|_{L^p} = \mu(\cup_{i>n} S_i)^{\frac{1}{p}}.$$

By the continuity of Lebesgue measure and the fact that E has finite measure we know that as n tends to infinity the Lebesgue measure of $\bigcup_{i>n} S_i$ tends to zero. Therefore $\chi_{\bigcup_{i=1}^{n} S_i}$ converges to $\chi_{\bigcup_{i=1}^{\infty} S_i}$ in terms of the L^p norm as n tends to infinity. So we have shown that ν is a finite signed measure and that it is absolutely continuous with respect to Lebesgue measure.

What does The Radon-Nikodym Theorem tell us about the relationship between ν and Lebesgue measure. It tells us that there exists a set Z of Lebesgue measure zero and a measurable function $f : E - Z \to \mathbf{R}$ such that for each measurable subset S of E we have $\nu(S) = \int \chi_S f d\mu$. That is, $\lambda(\chi_S) = \int \chi_S f d\mu$.

Let $g : E \to \mathbf{R}$ be a simple function with image $\{v_1, v_2, ..., v_k\}$ on the sets $Y_1, Y_2, ..., Y_k$. Then by the linearity of λ we have that

$$\lambda(g) = \sum_{i=1}^{k} v_i \lambda(\chi_{Y_i}) = \sum_{i=1}^{k} v_i \int \chi_{S_i} f d\mu = \int g f d\mu.$$

So knowing the function f tells us what λ does on simple functions. Can we now approximate any element of $L^p(E)$ by a sequence of simple functions and get a similar formula? That will be true if the function on $L^p(E)$ that maps g to $\int g f d\mu$ is continuous. That is, if $g_1, g_2, ...$ is a sequence of functions in $L^p(E)$ converging to g then does the sequence of numbers $\int g_1 f d\mu, \int g_2 f d\mu, ...$ converge to $\int g f d\mu$? A sufficient condition for this when $p > 1$ is that f belong to $L^q(E)$ where $\frac{1}{p} + \frac{1}{q} = 1$. The reason is that in this case Holder's inequality applies (recall that Holder's inequality says that whenever p and q are numbers larger than one such that $\frac{1}{p} + \frac{1}{q} = 1$ and g belong to $L^p(E)$ and f belongs to $L^q(E)$, then $\int g f d\mu \leq \|g\|_{L^p} \|f\|_{L^q}$). We then have that

$$\int (g - g_i) f d\mu \leq \|g - g_i\|_{L^p} \|f\|_{L^q}$$

and the right hand side converges to zero as n tends to infinity.

Let's try to show that f belongs to $L^q(E)$. We know that f is a measurable function. To show that it is an element of $L^q(E)$ we just need to show that its L^q norm is finite. That is, we need to show that $\left(\int |f|^q\right)^{\frac{1}{q}} < \infty$.

Recall that a definition of $\int |f|^q$ is the supremum of the collection of numbers $\int h^q$ where h is a simple function such that $0 \leq h \leq |f|$. We then have that

$$\int h^q = \int h^{q-1} g \leq \int h^{q-1} |f| = \int h^{q-1} \pm f$$

where the sign $\pm$ is meant to mean that the sign of f may vary from point to point. But recall that since h is a simple function it is certainly in $L^p(E)$ and so the right hand side is just the value our linear function λ takes at $\pm h^{q-1}$ where the sign is a plus when f is positive and a minus when f is negative (this will also belong to $L^p(E)$). And remember that λ is bounded. Therefore we have that

$$\int h^q \leq \lambda(\pm h^{q-1}) \leq \|\lambda\|_{L^{p*}} \| \pm h^{q-1}\|_{L^p} = \|\lambda\|_{L^{p*}} \left(\int h^{(q-1)p}\right)^{\frac{1}{p}}.$$

Notice that $(q-1)p = q$ so the inequality can be rewritten as $\left(\int h^q\right)^{\frac{1}{q}} \leq \|\lambda\|_{L^{p*}}$. Now let $h_1, h_2, ...$ be a sequence of simple functions converging to $|f|$ and such that $0 \leq |h| \leq f$. By The Dominated Convergence Theorem and the continuity of raising an expression to a power of $\frac{1}{q}$ we have that $\left(\int h_i^q\right)^{\frac{1}{q}}$ converges to $\left(\int |f|^q\right)^{\frac{1}{q}}$. Therefore $\left(\int |f|^q\right)^{\frac{1}{q}} \leq \|\lambda\|_{L^{p*}}$ which shows that f is an element of $L^q(E)$. Therefore the map from $L^p(E)$ to the real numbers defined by the formula $g \mapsto \int gfd\mu$ is continuous. Therefore, if $g_1, g_2, ...$ is a sequence of simple functions converging to the function g in the L^p norm then the sequence of numbers $\int g_1 fd\mu, \int g_2 fd\mu, ...$ converges to $\int gfd\mu$. This shows that if g belongs to $L^p(E)$, then $\lambda(g) = \int gfd\mu$. So the function f which is an element of $L^q(E)$ completely defines the function λ which is an element of $L^p(E)^*$. This gives us an interesting relationship between the dual space $L^p(E)^*$ and $L^q(E)$.

Our conclusion is that every bounded linear function $\lambda : L^p(E) \to \mathbf{R}$ is given by $\lambda(g) = \int gf$ for some $f \in L^q(E)$ where $\frac{1}{p} + \frac{1}{q} = 1$.

Consider the map from $L^q(E)$ to $L^p(E)^*$ defined by the formula $f \mapsto \int \cdot f$. The above shows that this map is a surjection. We also have that the inverse map is an isometry because $\|f\|_{L^q} = \|\lambda\|_{L^{p*}}$. We have shown above that $\|f\|_{L^q} \leq \|g\|_{L^{p*}}$. Let's show that equality holds. Recall that $\|\lambda\|_{L^{p*}} = \sup\{|\lambda(g)| : \|g\|_{L^p} = 1\}$. We want to find a function g in $L^q(E)$ such that $\int gf = \|f\|_{L^q}$. It can be shown that this will be the case then $|g| = \lambda |f|^{\frac{p}{q}}$ where λ is a constant. Therefore the inverse map is an isometry from $L^p(E)^*$ to $L^q(E)$.

What can we say when $p = 1$. We can't use Holder's inequality in this case. Hopefully the same thing will hold though. Let's claim that $\|f\|_{L^q} \leq \|\lambda\|_{L^{p*}}$. Consider the set $S = \{x \in E : f(x) > \|\lambda\|_{L^{p*}}\}$. Then since λ is bounded we have that

$$\|\lambda\|_{L^{p*}} \mu(E) \leq \int \chi_S f = \lambda(\chi_S) \leq \|\lambda\| \mu(S).$$

If $\mu(S) \neq 0$, then the first inequality is strict. So it must be that $\mu(S) = 0$. Therefore $\|f\|_{L^p} \leq \|\lambda\|_{L^{p*}}$ almost everywhere. And this is enough to show $\lambda(g) = \int gfd\mu$ for any g belonging to $L^p(E)$.

26. THE BAIRE CATEGORY THEOREM

Let X be a metric space. Let Y be a subset of X. The set Y is closed if y belongs to Y whenever $y_1, y_2, ...$ is a sequence of points in Y that converges to y. The closure $\bar{Y}$ of Y is the smallest closed set containing Y. The set Y is called dense if its closure is equal to X. That is, Y is dense if $\bar{Y} = X$.

For example, let E be a measurable subset of $\mathbf{R}^n$. Let X be the set of measurable real valued functions on E. Let Y be the set of simple functions on E. We have previously shown that any measurable function $f : E \to \mathbf{R}$ can be written as the limit of simple functions. Therefore Y is not a closed set. But Y is dense because $\bar{Y} = X$. Now let $X = L^p(E)$. Then Y is a subset of X. By the same logic Y is not closed but it is dense since any function $f \in L^p(E)$ can be written as the limit of a sequence of simple functions. Since this sequence of simple functions must belong to any closed set containing Y its limit f belongs to every closed set containing Y and hence f belongs to the closure of Y.

Let X be a metric space. If a subset Y of X is closed, then the set $X - Y$ is called open. An open subset Z of X may also be defined as follows: for each point z in Z there exists an $\epsilon > 0$ such that the set $B_\epsilon(z) = \{x \in Z : d(x, z) < \epsilon\}$ (called an open ball) is a subset of Z.

We would like to determine conditions under which the intersection of dense sets is a dense set. It seems plausible that this could be true since dense sets are in a sense large sets. Let's try to think of some counterexamples.

Counterexample 1: Let X be the real numbers. Then the rational numbers $\mathbf{Q}$ and the rational numbers translated by the square root of two $\mathbf{Q} + \{\sqrt{2}\}$ are both dense sets in X. But since they are disjoint their intersection is the? which is not dense in X (the empty set is a closed set and so is its own closure).

Let's rule this out by assuming each of our dense sets is an open set.

Counterexample 2: Let X be the rational numbers $\mathbf{Q} = \{q_1, q_2, ...\}$. Let $U_1 = \mathbf{Q} - \{q_1\}, U_2 = \mathbf{Q} - \{q_2\}, \cdots$ and so on. Each of these sets is dense. But their intersection is empty and the empty set is not dense here.

Let's rule this out by assuming that the metric space X is complete (that is, a metric space in which all Cauchy sequences converge).

It turns out that with these two conditions the intersection of countably many dense sets will also be dense. This result is called The Baire Category Theorem.

Theorem: (The Baire Category Theorem) Let X be a complete metric space. If $U_1, U_2, ...$ are dense open sets in X, then $\bigcap_i U_i$ is dense in X.

It is possible to show that a subset Y of a metric space X is dense if and only if Y has a nonempty intersection with every nonempty open subset U of X and this is true if and only if Y has a nonempty intersection with every open ball $B_\epsilon(x)$ for each $x \in X$ and each $\epsilon > 0$.

Let's use this to prove The Baire Category Theorem.

Proof of The Baire Category Theorem: Let x_0 be a point in X and $\epsilon_1 > 0$. Let's show there is a point y in $B_{\epsilon_1}(x)$ such that $y \in U_1 \cap U_2 \cap ....$

Since U_1 is dense there exists a point x_1 that is in both U_1 and $B_{\frac{\epsilon_1}{2}}(x_0)$. Since $B_{\frac{\epsilon_1}{2}}(x_0) \cap U_1$ is open there exists a number $\epsilon_2 > 0$ such that $B_{\epsilon_2}(x_1)$ is a subset of $B_{\frac{\epsilon_1}{2}}(x_0) \cap U_1$.

Now repeat the process with U_2. That is, since U_2 is dense there exists a point x_2 that is in both U_2 and $B_{\frac{\epsilon_2}{2}}(x_1)$. Since $B_{\frac{\epsilon_2}{2}}(x_1) \cap U_2$ is open there exists a number $\epsilon_3 > 0$ such that $B_{\epsilon_3}(x_2)$ is a subset of $B_{\frac{\epsilon_2}{2}}(x_1) \cap U_2$.

Doing this forever creates a sequence $x_1, x_2, ...$ of points such that

$$x_1 \in B_{\frac{\epsilon_1}{2}}(x_0) \cap U_1, x_2 \in B_{\frac{\epsilon_2}{2}}(x_1) \cap U_2, ...$$

where

$$B_{\frac{\epsilon_1}{2}}(x_0) \cap U_1 \supseteq B_{\epsilon_2}(x_1) \supseteq B_{\frac{\epsilon_2}{2}}(x_1) \cap U_2 \supseteq B_{\epsilon_3}(x_2), ...$$

Note that for each i we have that

$$B_{\frac{\epsilon_i}{2}}(x_{i-1}) \cap U_i \supseteq \text{closure}(B_{\frac{\epsilon_{i+1}}{2}}(x_i) \cap U_{i+1}) \supseteq B_{\frac{\epsilon_{i+1}}{2}}(x_i) \cap U_{i+1}.$$

The sequence $x_1, x_2, ...$ is clearly a Cauchy sequence because $\epsilon_{i+1} \le \frac{\epsilon_i}{2}$. For some $\delta > 0$ if you want $d(x_n, x_m) < \delta$ then just choose N such that $\epsilon_N < \delta$. Then $n, m \ge N$ implies that x_n, x_m belong to $B_{\frac{\epsilon_N}{2}}(x_{N-1})$ so that $d(x_n, x_m) < \epsilon_N < \delta$. That is, the sequence $x1, x2, ...$ is a Cauchy sequence.

Since X is complete there is a point y in X such that $\lim_i x_i = y$. For each i the sequence $x1, x2, ...$ is eventually in closure$(B_{\frac{\epsilon_{i+1}}{2}}(x_i) \cap U_{i+1})$ and so y is eventually in this set. But this implies that for each i the point y is in $B_{\frac{\epsilon_i}{2}}(x_{i-1}) \cap U_i$. Therefore y belongs to $B_{\epsilon_0}(x_0)$ and y belongs to $U1?U2?...$. So $\bigcap_i U_i$ is dense.

QED

27. THE OPEN MAPPING THEOREM

The word isomorphism means equal shape. A set X with some structure S_X and a set Y with some structure S_Y have the same shape if there is a bijection between X and Y that also gives a bijection between the interactions elements of X under structure S_X and the interactions of the elements of Y under structure S_Y. Everything you can do in language Y can be translated into language X and everything you do in language X can be translated into language Y and these translations agree going back and forward.

For example if we just had the sets X and Y then saying they are isomorphic means that they have the same number of elements. If in addition we had an binary relation R_X on X and a binary relation R_Y on Y then we could say X and Y are isomorphic if there is a bijection F between X and Y such that $(x_1, x_2) \in R_X$ if and only if $(f(x_1), f(x_2)) \in R_Y$. If X and Y are vector spaces then we could say that they are isomorphic if there is a bijection F between X and Y such that $x_1 + x_2 = x_3$ if and only if $F(x_1) + F(x_2) = F(x_3)$ and $\alpha x_1 = x_2$ if and only if $\alpha F(x_1) = F(x_2)$ (that is, F is a linear function). If X and Y are topological spaces then we would say they are isomorphic if there exists a bijection F between X and Y such that F is also a bijection between the open sets of X and the open sets of Y such that $X_1 \cap X_2 = X_3$ if and only if $F(X_1) \cap F(X_2) = F(X_3)$ and $\bigcup X_\alpha = X_1$ if and only if $\bigcup F(X_\alpha) = F(X_1)$ (these conditions follow automatically from the properties of a bijective function).

What if X and Y are Banach spaces? A Banach space has the structure of both a topological space and a vector space. It also has the structure of a norm. We might say that two Banach spaces are isomorphic if there is a bijection F from

X to Y such that F is an isomorphism of the vector spaces X and Y, F is an isomorphism of the topological spaces X and Y, and F is an isomorphism of the norms of X and Y. What might it mean for F to be an isomorphism of the norms of X and Y? It means that $\|x\|_X = c$ if and only if $\|F(x)\|_Y = c$. It turns out this condition implies that F is an isomorphism of the topological spaces X and Y.

Let's try to understand isomorphisms vector spaces that are also isomorphisms of topological spaces. We will prove a theorem about these isomorphisms related to Banach spaces. We will show that if X and Y are Banach spaces and if $F : X \to Y$ is an isomorphism of vector spaces that is continuous and surjective, then F is also an isomorphism of topological spaces. This result is known as The Open Mapping Theorem.

The Open Mapping Theorem: If $F : X \to Y$ is an isomorphism of vector spaces (that is, a linear function) that is continuous and surjective, then F is an isomorphism of topological spaces.

The function F is an isomorphism of topological spaces V and W if F gives a bijection between the open sets of V and W. One definition of F being continuous is that $F^{-1}(A)$ is an open subset of V whenever A is an open subset of W. This shows that any open set in W is associated with an open set in V. We want to show that this map is a bijection. Since F is injective and surjective the map F^{-1} that associates open subsets of W with open subsets of V is map is injective. Is the map surjective? That is, if B is an open subset of V does there exist an open subset A of W such that $F^{-1}(A) = B$? This is equivalent to saying that if B is an open subset of V then does there exist an open subset A of W such that $F(B) = A$? That is, if B is an open subset of V is $F(B)$ an open subset of W? So what the Open mapping theorem is saying is if $F : X \to Y$ is an isomorphism of vector spaces that is continuous then $F(B)$ is an open subset of Y whenever B is an open subset of X.

Let's prove The Open Mapping Theorem. First of all let's make a simplification. Suppose we could prove the theorem for a function F whose operator norm was equal to 1. That is, an F such that the smallest number C such that $\|F(x)\|_Y \leq C\|x\|_X$ for all x in X is 1. Suppose then we wanted to prove the theorem for an F with an arbitrary operator norm? Since F is a linear and continuous there exits a number C which is the smallest number such that $\|F(x)\|_Y \leq C\|x\|_X$ for all x in X. If this number C is zero then F is the zero function. Since F is assumed to be surjective it must be that Y is a singleton. Since Y is an open set and any nonempty subset B of X gets mapped to Y which is an open set (and the empty set gets mapped to the empty set, another open subset of Y) we have that F is open.

If C is not equal to zero we can consider the function $\frac{1}{C}F$. The operator norm for this function is equal to 1 so we can apply the theorem and we then have that if B is an open subset of X then $\frac{1}{C}F(B) = A$ is an open subset of Y. Is the set $F(B) = CA$ an open subset of Y. It is easy to show that it is. The set $CA = \{y \in Y : y = Ca \text{ for some } a \in A\}$. To show that CA is open we need to show that if y is a point of CA then there exists a positive number ϵ such that the open ball of radius ϵ around y, that is the set $B_\epsilon(y) = \{z \in Y : \|y - z\|_Y < \epsilon\}$, is a subset of CA. Since y belongs to CA there exists a point a in A such that $y = Ca$. Since A is open there exists a positive number ϵ such such that $B_\epsilon(a)$ is a subset of A. I claim that $B_{C\epsilon}(y)$ is a subset of CA.

So let z belong to $B_{C\epsilon}(y)$. Then $\|y-z\|_X < C\epsilon$. Dividing by C gives $\|a-\frac{z}{C}\| < \epsilon$. Therefore $\frac{z}{C}$ belongs to $B_\epsilon(a)$ and hence belongs to A. This implies that $z = C\frac{z}{C}$ belongs to CA. Therefore $B_{C\epsilon}(y)$ is a subset of CA. This shows that CA is an open subset of Y. The point is that if we can prove The Open Mapping Theorem when the operator norm of F is equal to 1 then we have proved The Open Mapping Theorem. Since it will probably make it easier to prove the theorem let's assume that the operator norm of F is equal to 1.

Let B be an open subset of X. We want to show that $F(B)$ is an open subset of X. One way to do this is to show that for each point y in $F(B)$ there exists a positive number ϵ such that the open ball $B_\epsilon(y)$ is a subset of $F(B)$.

Let x be a point in B. Since B is open there exists a positive number ϵ' such that $B_{\epsilon'}(x)$ is a subset of B. Suppose we could find a positive number ϵ such that $B_\epsilon(0_Y)$ is a subset of $F(B_{\epsilon'}(0_X))$. Since F is continuous we then have that $F^{-1}(B_\epsilon(0_Y)) \cap B_{\epsilon'}(0_X)$ (Call this set A) is an open subset of $B_{\epsilon'}(0_X)$ that maps under F to $B_\epsilon(0_Y)$. Then we have that $F(A + \{x\}) = F(A) + F(x)$ is an open subset of $F(B)$ containing $F(x)$. This will show that $F(B)$ is open in Y.

Let's do one more simplification. Suppose we could prove this when ϵ' is equal to 1. That is, suppose we could find an $\epsilon > 0$ such that $B_\epsilon(0_Y)$ is a subset of $F(B_1(0_X))$. This is the same as saying that $\epsilon' B_{\epsilon'}(0_Y)$ is a subset of $\epsilon' F(B_1(0_X))$. By the linearity of F this is the same as saying that $B_{\epsilon'\epsilon}(0_Y)$ is a subset of $F(B_{\epsilon'}(0_X))$, which is what we want. So we only need to prove the result for when $\epsilon' = 1$.

Let me now restate The Open Mapping Theorem and what we have to show. The open mapping theorem says that if $F : X \to Y$ is a map between Banach spaces and is an isomorphism of vector spaces that is surjective and continuous then it is an isomorphism of topological spaces.

To finish the proof of this theorem we need to show that there exists a positive number ϵ such that $F(B_1(0_X))$ is a subset of $B_\epsilon(0_Y)$ where the F here has an operator norm equal to 1.

Now consider the open balls $B_1(0_X), B_2(0_X), ...$ The union of these open balls is X. Since F is surjective $\bigcup_{i=1}^{\infty} F(B_i(0_X)) = Y$. This implies that $\bigcap(Y -$ closure$[F(B_i(0_X)]) = \emptyset$.

Recall The Baire Category Theorem. The Baire Category Theorem says that in a complete metric space the countable intersection of dense and open sets is also dense. Since X is a Banach space it is a complete metric space. Since the empty set is not dense in Y and each of the sets $Y -$ closure$[F(B_i)(0_X)]$ is open (this is why we took the closure) The Baire Category Theorem implies that for some n the set $Y -$ closure$[F(B_n(0_X))]$ is not dense in Y. This means that there exists an $\epsilon > 0$ and a point x in X such that the open ball $B_\epsilon(x)$ does not intersect the set $Y -$ closure$[F(B_n(0_X))]$. This means that $B_\epsilon(x)$ is a subset of closure$[F(B_n(0_X))]$. The argument above for why we could let $\epsilon' = 1$ shows that we can also assume that $n = 1$.

There are two obstacles to complete the proof. Obstacle number one is that the ball $B_\epsilon(x)$ is centered at x and x may not be equal to 0_Y as we want. Obstacle number two is that this open ball is a subset of the closure of $F(B_1(0_X))$ and we want our the final open ball to e a subset of $B_1(0_X)$.

Suppose that y is a point in Y such that $\|y\|_Y < \epsilon$. Then $x + y$ and $x - y$ belong to $B_\epsilon(x)$ so they are also in closure$[F(B_1(0_X))]$. This implies that for any positive number δ such that $\epsilon > \delta$ there exists points v and w in $B_1(0_X)$ such that $\|(x+y) - F(v)\|_Y < \delta and \|(x-y) - F(w)\| < \delta$. By the triangle inequality we also have that $\|F(w - v) - 2y\|_Y < 2\delta$ which implies that $\|F(\frac{w-v}{2}) - y\| < \delta$. Since $B_1(0_X)$ is convex we have that $\frac{w-v}{2}$ belongs to $B_1(0_X)$.

So we have just shown that if y belongs to $B_\epsilon(0_Y)$ then for any $\delta > 0$ we can choose z in $B_1(0_X)$ such that $\|F(z) - y\|_Y < \delta$.

Let's take δ equal to ϵ^2 (suppose $\epsilon < 1$). Then we can find a point z in $B_1(0_X)$ such that $\|y - F(z)\|_Y < \epsilon^2$. This implies that $\|\frac{y-F(z)}{\epsilon}\|_Y < \epsilon$ so we can find a point z' such that $\|\frac{y-F(z)}{\epsilon} - F(z')\|_Y < \epsilon^3$. This implies that $\|y - F(z) - \epsilon F(z')\|_Y < \epsilon^2$. It also implies that $\|\frac{y-F(z)-\epsilon F(z')}{\epsilon}\|_Y < \epsilon$. Therefore there exists a point z'' in $B_1(0_X)$ such that $\|\frac{y-F(z)-\epsilon F(z')}{\epsilon} - F(z'')\|_Y < \epsilon^4$. This implies that $\|y - F(z) - \epsilon F(z') - \epsilon^2 F(z'')\|_Y < \epsilon^3$. We can continue this process so that we get a sequence of points $F(z), F(z) - \epsilon F(z'), F(z) - \epsilon F(z') - \epsilon^2 F(z''), ...$ This sequence converges to y.

Because F is linear this implies that $F(z + \epsilon z' + \epsilon^2 z'' + ...) = y$. Let $\hat{z}$ denote the point $z + \epsilon z' + \epsilon^2 z'' + ...$ We have that $\|\hat{z}\|_X \leq \frac{1}{1-\epsilon}$. We should have chosen the epsilons in the previous paragraph more carefully so that this is less than one. Then we would have shown that $B_\epsilon(0_Y)$ is a subset of $F(B_1(0_x))$. This is unclear and correct.

28. SETS WITH THE SAME SHAPE

Let X be a set with language S_X and let Y be a set with language S_Y. Why is is useful to know whether (X, S_X) is isomorphic to (Y, S_Y)? I think it is because problems in (X, S_X) can be posed in (Y, S_Y), solved there and the solutions translated back to (X, S_X).

Here is an example. Let $f : (0, \infty) \to \mathbf{R}$ be the the natural logarithm. This is a bijection from $\mathbf{R} to (0, \infty)$. It translates any statement $x_1 \cdot x_2 = x_3$ to the statement $\log(x_1) + \log(x_2) = \log(x_3)$ and it translates this statement back to $x_1 \cdot x_2 = x_3$.

If we know how to solve the statement $x_1 \cdot x_2 = x_3$ for some positive real numbers x_1, x_2 and x_3 then we can also solve the statement $y_1 + y_2 = y_3$ whenever $y_1 = \log(x_1), y_2 = \log(x_2), and y_3 = \log(x_3)$. This is because f is injective.

If we know how to solve the statement $y_1 + y_2 = y_3$ for some real numbers y_1, y_2, and y_3 then we can solve the statement $x_1 \cdot x_2 = x_3$ whenever $x_1 = e^{y_1}, x_2 = e^{y_2}$, and $x_3 = e^{y_3}$. This is because the inverse is injective.

Note that not any bijection will work. For instance suppose that $g : (0, \infty) \to \mathbf{R}$ is a bijection. Let's try to translate the statement $x_1 \cdot x_2 = x_3$ where x_1, x_2, and x_3 are positive real numbers. Let $y_1 = g(x_1), y_2 = g(x_2)$, and $y_3 = g(x_3)$. We have that $g(x_1) \cdot g(x_2) = g(x_3)$.

Suppose we only know how to solve statements of the form $z_1 + z_2 = z_3$ when z_1, z_2, and z_3 are real numbers. The function g helps us solve the equation $x_1 \cdot x_2 = x_3$ if and only if $g(x_1) + g(x_2) = g(x_3)$.

What is an example of this? Suppose g is defined by $g(x) = \sqrt{x}$. Let's try to solve $e \cdot \pi = x$ for x. Suppose we know how to solve all equations of the form $z_1 + z_2 = z_3$ whenever z_1, z_2, and z_3 are positive real numbers. This knowledge and the knowledge of g does not help us solve the equation $e \cdot \pi = x$ because g does not satisfy the property that $g(x_1) + g(x_2) = g(x_1 \cdot x_2)$ whenever x_1, x_2 and x_3 are positive real numbers. But if we use the function f defined by $f(x) = \log(x)$ then we have $\log(e) + \log(\pi) = \log(x)$ and we know how to solve this equation so that we know $\log(x)$ and from this we know x.

29. AXIOMS FOR TH LOGARITHMIC FUNCTION

Let $f : (0, \infty) \to \mathbf{R}$ be a bijection such that for all x_1, x_2, x_3 in $(0, \infty)$ it is true that $x_1 \cdot x_2 = x_3$ if and only if $f(x_1) + f(x_2) = f(x_3)$. We have seen that any logarithmic function f (i.e. a function f defined by the rule $f(x) = \log_b(x)$ where $b > 0$) is such a function.

A natural question to ask is whether f must be a logarithmic function. Are the above properties for $f : (0, \infty) \to \mathbf{R}$ just another way of saying that f is a logarithmic function? That is, are the properties

(1) $f : (0, \infty) \to \mathbf{R}$, (2) $fisabijection$, (3) for all x_1, x_2, x_3 in $(0, \infty)$ it is true that $x_1 \cdot x_2 = x_3$ if and only if $f(x_1) + f(x_2) = f(x_3)$,

axioms for the logarithmic function?

Another way to state property number (3) is: for all x_1, x_2, x_3 in $(0, \infty)$ it is true that $f(x_1 \cdot x_2) = f(x_1) + f(x_2)$.

Let's rearrange the problem slightly.

Do the properties

(1) $g : \mathbf{R} \to (0, \infty)$, (2) g is a bijection, (3) for all y_1, y_2, y_3 in $\mathbf{R}$ it is true that $y_1 + y_2 = y_3$ if and only if $g(y_1) \cdot g(y_2) = g(y_3)$

imply that g is defined by the rule $g(y) = b^y$ for some real number $b > 0$?

Another way to state property number 3 is:

(3') for all y_1, y_2, y_3 in $\mathbf{R}$ it is true that $g(y_1 + y_2) = g(y_1) \cdot g(y_2)$.

What is the value of g at 0? By property (3') we have $g(0 + 0) = g(0) \cdot g(0)$ which implies that $g(0) = 1$.

Let's denote $g(1)$ by b. Then we have $g(n) = b^n$ whenever n is a nonnegative integer. We have that $g(n - 2n) = g(n)g(-2n)$ so $\frac{g(-n)}{g(-n) \cdot g(-n)} = b^n$ so $g(-n) = \frac{1}{b^n}$.

Let n be a nonzero integer. We have

$$\underbrace{g(\frac{1}{n})...g(\frac{1}{n})}_{\text{n times}} = g(1) = b.$$

Therefore $g(\frac{1}{n}) = b^{\frac{1}{n}}$.

Let m also be a non-zero integer. We have

$$g(\frac{m}{n}) = \underbrace{g(\frac{1}{n})...g(\frac{1}{n})}_{\text{m times}} = (b^{\frac{1}{n}})^m = b^{\frac{m}{n}}.$$

Since the map that takes the real number x to the number b^x is continuous and any real number can be written as the limit of rational numbers we have that $g(x) = b^x$. That is,

$$g(x) = \lim_i g(\frac{m_i}{n_i}) = \lim_i b^{\frac{m_i}{n_i}} = b^x.$$

whenever $\frac{m_1}{n_1}, \frac{m_2}{n_2}, ...$ is a sequence of rational numbers whose limit is x.

$g(1) = b$ is non-negative because by property number 1 g takes values in the positive real numbers. $g(1) = b$ is not equal to 0 because by property number 2 g is a surjection.

30. RELATIVES OF THE OPEN MAPPING THEOREM

Theorem 1: $\sqrt{2}$ is an irrational number.

Theorem 2: There are infinitely many prime numbers.

Both these theorems are true but they are not close relatives. You can't prove one easily from the other.

Theorem 3: $\sqrt{2}+1$ is an irrational number.

Theorem 1 and 3 are relatives. Theorem 3 can be easily proved using Theorem 1 and Theorem 1 can be easily proved using Theorem 3.

Theorem 4: Every number can be uniquely written as the product of prime numbers.

Theorem 2 and 4 are also relatives. But they are relatives in a different way that Theorem 1 and 3 are relatives. Theorem 2 can be proved by applying Theorem 4. I don't think Theorem 4 can be proved easily from Theorem 2.

Recall The Open Mapping Theorem: If $F : V \to W$ is a bounded linear surjection of Banach spaces, then F is open (i.e. if U is an open subset of V, then $F(U)$ is an open subset of W).

A close relative of this theorem is the following:

Theorem 1: If $F : V \to W$ is a bounded linear bijection of Banach spaces, then F^{-1} is continuous.

This theorem follows directly from The Open Mapping Theorem when F is a bijection. The Open Mapping Theorem also follows from this theorem.

Let $F : V \to W$ be a bounded linear surjection of Banach spaces. Let's partition V by where F sends its elements. That is, define a relation $\sim$ by saying that for two elements $v_1, v_2 \in V$ we have $v_1 \sim v_2$ if and only if $F(v_1) = F(v_2)$. This relation partitions V (the relation is an equivalence relation). If v is an element of V denote the equivalence class that corresponds to v by $\bar{v}$. It can be shown that $v_1 \sim v_2$ if and only if there is an element v in the kernel of F such that $v_1 = v_2 + v$. ($v = v_1 - v_2$ works and if there exists an element v in the kernel of F such that $v_1 = v_2 + v$ then $F(v_1) = F(v_2 + v) = F(v_2)$ so that $v_1 \sim v_2$.) The partition of V is denoted by $V/\ker(F)$.

Define the map $\bar{F} : V/\ker(F) \to W$ by saying that $\bar{F}(\bar{v}) = F(v)$. It is possible to show that the set $V/\ker(F)$ with the norm defined by

$$\|\bar{v}\|_{V/\ker(F)} = \inf\{\|x\|_V : x \in \bar{v}\}$$

is a Banach space. Therefore $\bar{F} : V/\ker(F) \to W$ is a bounded linear bijection of the Banach spaces $V/\ker(F) and W$. Theorem 1 then implies that $\bar{F}^{-1}$ is continuous.

Consider the map $\pi : V \to V/\ker(F)$ that takes an element v to its equivalent class $\bar{v}$. That is, π is defined by the equation $\pi(v) = \bar{v}$. It is possible to show that π^{-1} is continuous. We then have that $F = \pi \circ \bar{F}$. It is possible to show that π is open. So if U is an open subset of V, the $F(U)$ is an open subset of W because $F(U) = \pi(\bar{F}(U))$. This proves The Open Mapping Theorem.

The Open Mapping Theorem implies Theorem 1 without much thinking. Showing that Theorem 1 implies The Open Mapping Theorem requires much more thinking.

Let's apply Theorem 1. Let $f : X \to Y$ be a function of metric spaces. One way to represent f is as a set $\Gamma(f)$ of ordered pairs (x, y) such that $f(x) = y$ and $y_1 = y_2$ whenever (x, y_1) and (x, y_2) belong to the set. Sometimes the set $\Gamma(f)$ is called the graph of f.

One way to say that f is continuous is that if $x_1, x_2, ...$ is a sequence in X which converges to x then the sequence $f(x_1), f(x_2), ...$ converges to $f(x)$.

What does continuity imply about the graph of f. One thing it implies is that the graph of f is a closed set. That is, if $(x_1, y_2), (x_2, y_2), ...$ is a sequence in $\Gamma(f)$ which converges to (x, y), then the point (x, y) belongs to $\Gamma(f)$. You can see that this is true: since $x_1, x_2, ...$ is a sequence in X which converges to x, then the sequence $f(x_1), f(x_2), ...$ converges to $f(x)$. Since f is a function we have that $y_1 = f(x_1), y_2 = f(x_2)$, and so on. This implies that the sequence $y_1, y_2, ...$ converges to $f(x)$ so that $f(x) = y$. That is, (x, y) belongs to $\Gamma(f)$.

The converse is not always true. It is true if Y is compact (compact means that if $y_1, y_2, ...$ is a sequence in Y then it has a subsequence $y_{n_1}, y_{n_2}, ...$ that converges to a point y in Y.)

It is also true if both X and Y are Banach spaces and f is a linear function.

This second statement is called The Closed Graph Theorem. As well as stating The Closed Graph Theorem we can also prove it.

The Closed Graph Theorem: If $F : X \to Y$ is a linear map of Banach spaces, then F is continuous if and only if $\Gamma(F)$ is closed.

Proof: We have already shown that if F is continuous then $\Gamma(F)$ is closed.

Conversely, assume that $\Gamma(F)$ is closed. Consider the set $X \times Y$ with (what we would like to be a) norm $\|(x, y)\| = \|x\|_X + \|y\|_Y$. It can be shown that this is a Banach space whenever X and Y are Banach spaces (which we are assuming them to be). Because $\Gamma(F)$ is closed it is also a Banach space with this norm.

Consider the projection map $F_1 : \Gamma(F) \to X$ defined by $F_1(x, y) = x$ and $F_2 : \Gamma(F) \to Y$ defined by $F_2(x, y) = y$. These are linear maps of Banach spaces. It is easy to see these maps are continuous (or what is the same thing bounded).

The function F_1 is a bijection. Therefore by Theorem 1 its inverse F_1^{-1} is continuous. But we can write $F = F_2 \circ F_1^{-1}$. Therefore F is a continuous function because it is the composition of two continuous functions. QED

So from Theorem 1 we can without much thinking prove The Closed Graph Theorem. Is the reverse true? Let's try to prove Theorem 1 again but this time using The Closed Graph Theorem.

Theorem 1: If $F : X \to Y$ is a bounded linear bijection of Banach spaces, then F^{-1} is continuous.

Proof: Since F is linear being bounded is the same as being continuous. Therefore F is continuous. By The Closed Graph Theorem this is the same as having a closed graph. Therefore F has a closed graph. Since F is a bijection $\Gamma(F) = \Gamma(F^{-1})$. So F^{-1} has a closed graph which by The Closed Graph Theorem is the same as being continuous. Therefore F^{-1} is continuous. QED

This shows that Theorem 1 follows without much thinking from The Closed Graph Theorem.

What we have shown is that The Open Mapping Theorem, Theorem 1, and The Closed Graph Theorem are close relatives. Each follows easily from the other.

31. AN APPLICATION OF THE CLOSED GRAPH THEOREM

Let V and W be Banach spaces. Recall that the dual space of V which is denoted V^* is the set of bounded linear functions from V to $\mathbf{R}$. Likewise, the dual space of W which is denoted W^* is the set of bounded linear functions from W to $\mathbf{R}$.

Suppose we have a function $F : V \to W$ and a function $G : W^* \to V^*$. Let's say that these two functions are adjoint if for all vectors v in V and for all functions λ in W^* we have that $(G(\lambda))(v)$ is equal to $\lambda(F(v))$.

An obvious guess for G is the function that maps each λ in W^* to the function $\lambda \circ F$. The problem is that the function $F \circ \lambda$ may not be an element of V^* for all λ in W^*.

One requirement is that F is a linear function. Suppose that F is not a linear function. Then either there exists points v_1 and v_2 in V such that $F(v_1 + v_2) \neq F(v_1) + F(v_2)$ or there exists a point v in V and a real number α such that $F(\alpha v) \neq \alpha F(v)$. For the first case let

$$\lambda_0 : \text{span}\{F(v_1 + v_2), F(v_1) + F(v_2)\} \to \mathbf{R}$$

be the linear function that takes the value 1 on $F(v_1 + v_2)$ and the value 0 on $F(v_1) + F(v_2)$. By The Hahn-Banach Theorem there exists a linear function $\lambda : W \to \mathbf{R}$ that is an extension of λ_0 (and such that the operator norm of λ_0 is equal to the operator norm of λ). But then we have that

$$\lambda \circ F(v_1 + v_2) = 1 \neq 0 = \lambda \circ F(v_1) + \lambda \circ F(v_2)$$

so that $\lambda \circ F$ is not a linear function. For the second case let

$$\lambda_0 : \text{span}\{F(\alpha v), \alpha F(v)\} \to \mathbf{R}$$

be the linear function that takes the value 1 on $F(\alpha v)$ and the value 0 on $\alpha F(v)$. By the Hahn-Banach Theorem there exists a linear function $\lambda : W \to \mathbf{R}$ that is an extension of λ_0 (and such that the operator norm of λ_0 is equal to the operator norm of λ). But then we have that

$$\lambda \circ F(\alpha v) = 1 \neq 0 = \lambda \circ \alpha F(v) = \alpha \lambda \circ F(v)$$

so that $\lambda \circ F$ is not a linear function.

Suppose that F is a continuous function. Since F is linear this is the same as saying that Fv is bounded. Then there exists a number C such that for all v in V we have $\|F(v)\|_W \leq C\|v\|_V$. Since F is a linear function we have that $\lambda \circ F$ is a linear function. To show that $\lambda \circ F$ belongs to V^* (and hence that the map $\lambda \mapsto \lambda \circ F$ is a map from $W^* \to V^*$) we only need to show it is bounded. Since λ belongs to W^* it is a bounded linear map and so there exists a constant K such that for all w in W we have $|\lambda(w)| \leq K\|w\|_W$. Let v belong to V. We have that

$$|\lambda \circ F(v)| = |\lambda(F(v))| \leq K\|F(v)\|_W \leq KC\|v\|_V.$$

Therefore $\lambda \circ F$ is bounded and so an element of V^*.

What we have just shown is that if F is a bounded linear map then it has an adjoint and its adjoint is the map $G : W^* \to V^*$ that maps λ to $\lambda \circ F$.

It should not be too surprising that a version of the converse is true: if F and Gare adjoints then F is a bounded linear map.

We have already shown that F is linear. We will use The Closed Graph Theorem to show that F is bounded (which is the same as continuous for a linear map). Consider a point (v, w) in the graph of F. Then we have that $F(v) = w$. This implies that $\lambda(F(v)) = \lambda(w)$ and since F and G are adjoint it implies that $\lambda(w) = G(\lambda)(v)$. Now suppose that for all functions λ in W^* we had $\lambda(w) = G(\lambda)(v)$. Since F and G are adjoint we have that $\lambda(w) = \lambda(F(v))$ for all $\lambda \in W^*$. By The Hahn-Banach Theorem this implies that $w = F(v)$. To see this suppose that $w \neq F(v)$. Define the linear function $\lambda_0 : \text{span}\{w, F(v)\} \to \mathbf{R}$ that takes the value 1 on w and the value 0 on $F(v)$. By The Hahn-Banach Theorem there exists a linear function $\lambda : W \to \mathbf{R}$ that is an extension of λ_0 and such that has the same operator norm as λ_0 (so that λ belongs to W^*). But then we have that $\lambda(w) \neq 1 \neq 0 = \lambda(F(v))$.

So we have that (v, w) belongs to $\Gamma(F)$ if and only if $\lambda(w) = (G(\lambda))(v)$ for all $\lambda in W^*$. That is,

$$\Gamma(F) = \bigcap_{\lambda \in W^*} \{(v, w) : \lambda(w) = (G(\lambda))(v)\}.$$

Because each $\lambda in W^*$ is continuous this is an intersection of closed sets. This intersection of closed sets is always a closed set. Therefore the graph of F is closed.

Now recall The Closed Graph Theorem: If $F : V \to W$ is a linear map of Banach spaces then F is continuous if and only if the graph of F is closed.

By The Closed Graph Theorem F is continuous. Therefore F is bounded.

So $F : V \to W and G : W^* \to V^*$ are adjoint if and only if F is a bounded linear map.

32. THE UNIFORM BOUNDEDNESS PRINCIPLE

Suppose that $F_1, F_2, ...$ is a sequence of bounded linear functions from the Banach space V to the Banach space W. What does it mean for this sequence to converge to a linear function $F : V \to W$?

One possibility is to use the norm of W. That is, to consider for each point v in V the sequence $F_1(v), F_2(v), ...$ We would then say that the sequence $F_1, F_2, ...$ converges to the linear function F if for each point v in V the sequence $F_1(v), F_2(v), ...$ of points in W converges to $F(v)$. That is, $\lim_n \|F_n(v) - F(v)\| = 0$ for each v in V. This is called pointwise convergence.

Another possibility is the use the operator norm (the norm on $\mathrm{Hom}(V, W)$). That is to say that the sequence $F_1, F_2, ...$ converges to F if $\lim_n \|F_n - F\|$ (where for G in $\mathrm{Hom}(V, W)$ (that is, a bounded linear function $G : V \to W)\|G\|$ is defined to be the smallest number C such that $\|G(v)\|_W \leq C\|v\|_V$ for all v in V; this is the same number as the largest value of $\|G(v)\|_W$ for all v such that $\|v\|_V = 1$). This is called operator norm convergence.

What is the relationship between pointwise convergence and operator norm convergence?

We have that

$$\|F_n - F\| = \sup\{\|F_n(v) - F(v)\|_W : \|v\|_V = 1\}.$$

Pointwise convergence means that $\lim_n \|F_n(v) - F(v)\|_W = 0$ for all v in V. You can see from the above equation that convergence in the operator norm implies that $\lim_n \|F_n(v) - F(v)\|_W = 0$ for each v such that $\|v\|_V = 1$. If u is an arbitrary element of V let $v = \frac{u}{\|u\|_V}$. Then $\|v\|_V = 1$ so that $\lim_n \|F_n(v) - F(v)\| = 0$. And this implies that $\lim_n \frac{1}{\|u\|_V}\|F_n(u) - F(u)\| = 0$ which implies that $\lim_n \|F_n(u) - F(u)\| = 0$.

The converse is not true: pointwise convergence does not imply convergence in the operator norm. Here is an example that proves this to be so. Let $V = L^1(\mathbf{R})$ and let $W = \mathbf{R}$. Define the function F_i by the equation $F_i(f) = \int f\chi_{[i,i+1]}$. The sequence of functions $F_1, F_2, ...$ converges pointwise to zero. If it were not the case then for each integer N we could choose an integer n such that $\int |f|\chi_{[0,n]} \geq N$. The sequence of function is $f\chi_{[0,1]}, f\chi_{[0,2]}, ...$ is a nondecreasing sequence of nonnegative measurable functions. By the monotone convergence theorem $\int |f| = \lim_n \int |f|\chi_{[0,n]} = \infty$. Therefore f does not belong to $L^1(\mathbf{R})$. The sequence $F_1, F_2, ...$ does not converge to zero in the operator norm since $\left|\int f\chi_{[i,i+1]}\right| = 1$ whenever

$f = \chi_{[i,i+1]}$ and in this case $\|f\|_{L^1} = \int |f| = 1$. Therefore the operator norm of any F_i is at least 1.

Note that convergence in the operator norm is only for bounded linear functions. Not only must each term of the sequence be a bounded linear function but the limit must be a bounded linear function. Pointwise convergence is well defined for arbitrary functions.

Suppose that $F_1, F_2, ...$ is a sequence of bounded linear functions from the Banach space V to the Banach space W which converge pointwise to a function F.

Is F bounded?

We will show that the answer is "yes".

This follows from a theorem called The Uniform Boundedness Principle.

The Uniform Boundedness Principle: Suppose V and W are Banach spaces, and we are given a set $\{T_\alpha\}$ of bounded linear functions such that for all v in V we have that $\{T_\alpha(v)\}$ is bounded (that is, there exists a number N such that for each α we have $\|T_\alpha(v)\|_W \leq N$). Then $\{T_\alpha\}$ is bounded (that is, there exists a number C such that the operator norm $\|T_\alpha\| \leq C$ for each α).

How does this imply that if $F_1, F_2, ...$ is a sequence of bounded linear functions converging pointwise to F then F is bounded. We have that for each v in V the set $\{F_1(v), F_2(v), ...\}$ is bounded (because this sequence converges). Therefore there exists a number C such that the operator norm of each function F_i is at most C. That is if v is such that $\|v\|_V = 1$, then $\|F_i(v)\|_W \leq C$. Since $F(v) = \lim_i F_i(v)$ and the norm on W is a continuous function we have that

$$\|F(v)\|_W = \|\lim_i F_i(v)\|_W = \lim_i \|F_i(v)\|_W \leq \lim_i C = C.$$

Therefore F is bounded.

How can we prove The Uniform Boundedness Principle? Let B_N denote the set $\{v \in V : \|T_\alpha(v)\|_W \leq N \text{ for all } \alpha\}$. Then $V = \bigcup_{N=1}^{\infty} B_N$. Each set B_N is closed because the norm on W is a continuous function and each T_α is a continuous function. The sequence of sets $B_1, B_2, ...$ is increasing. Recall The Baire Category Theorem.

The Baire Category Theorem: Let X be a complete metric space. The union of countably many nowhere dense subsets of X is not equal to X.

The Baire Category Theorem implies that for some N we have that B_N is not nowhere dense. That is, B_N contains a nonempty open set. This implies there is a point v in B_N and a number $\epsilon > 0$ such that $B_\epsilon(v) \subseteq B_N$.

If w is a point in V such that $\|w\|_V < \epsilon$, then the point $v + w$ belongs to $B_\epsilon(v)$ and so to B_N. This implies that $\|T\alpha(v+w)\|_W \leq N$ for all α. Therefore

$$\|T\alpha(w)\|_W = \|T_\alpha(v+w) - T\alpha(v)\|_W \leq \|T_\alpha(v+w)\|_W + \|T\alpha(v)\|_W \leq 2N.$$

This implies that for any w in V and for any α we have that

$$\|T_\alpha(w)\| \leq \frac{2N\|w\|_V}{\epsilon'}$$

where ϵ' is any number such that $0 < \epsilon' < \epsilon$. This shows that $\{T_\alpha\}$ is bounded by $\frac{2N}{\epsilon'}$. So ends the proof of The Uniform Boundedness Principle.

33. MOST CONTINUOUS FUNCTIONS ARE NOT DIFFERENTIABLE ANYWHERE

Recall The Baire Category Theorem.

The Baire Category Theorem: Let X be a metric space. The intersection of countably many dense open subsets of X is nonempty.

If you don't remember a dense subset of X is a set whose intersection with any nonempty subset of X is nonempty.

Therefore the complement of a dense subset of X has the property that it does not contain any nonempty open subsets of X and conversely if a subset of X has the property that it does not contain any nonempty open subsets of X then its complement must be a dense subset of X.

A set is called nowhere dense if its closure does not contain any nonempty open subsets of X. That is, a subset of X is nowhere dense if the complement of its closure is a dense subset of X. This implies that an open set is dense if and only if its complement is closed and nowhere dense.

Using this we can state a version of The Baire Category theorem: Let X be a metric space. The union of countably many closed nowhere dense subsets of X is not equal to X.

Let's now make a definition of smallness for a subset of X. A subset Y of X is called meagre if it is contained in the union of countably many closed nowhere dense subsets of X.

Note that the interior of a nowhere dense subset of X is nowhere dense so if Y is contained in the union of countably many nowhere dense subsets of X then it is also contained in the the union of the closure of these sets and the closure of each of these sets is nowhere dense. The reason for demanding in the definition that the nowhere dense sets be closed I believe is so that we can easily apply our version of The Baire Category Theorem.

The complement of a meagre set is called comeagre. A meagre set is small when compared to X. Therefore a comeagre set is large when compared to X.

In $\mathbf{R}^n$ we also have a notion of a set being small. That notion is that the set has Lebesgue measure zero.

Here is an example of a set with measure zero that is not meagre. The rational numbers as a subset of $\mathbf{R}$ has Lebesgue measure zero but is not meagre. It is not meagre because it is a dense subset of $\mathbf{R}$ and therefore its closure is $\mathbf{R}$ and $\mathbf{R}$ certainly contains open subsets of $\mathbf{R}$.

Here is an example of a set with positive measure that is meagre. Consider the sequence $a_1, a_2, ...$ of real numbers such that the sum $a_1 + a_2 + ...$ converges. Let $r_1, r_2, ...$ be an enumeration of the rational numbers. Let

$$A = (r_1 - a_1, r_1 + a_1) \cup (r_2 - a_2, r_2 + a_2) \cup ...$$

. By subadditivity of the Lebesgue measure the Lebesgue measure of A is no more than $\frac{1}{2}$. That is:

$$\mu(A) \leq \mu(r_1 - a_1, r_1 + a_1) + \mu(r_2 - a_2, r_2 + a_2) + ... = 2(a_1 + a_2 + ...) < \infty.$$

Therefore $\mathbf{R} - A$ has infinite Lebesgue measure. The set A is an open and dense subset. Therefore $\mathbf{R} - A$ is closed and nowhere dense. Therefore $\mathbf{R} - A$ is meagre.

Even though a set being meagre does not imply that it has Lebesgue measure zero and a set having Lebesgue measure zero does not imply it is meagre both ideas still describe that a set is small. Therefore if we are given a set that satisfies one condition it probably satisfies the other.

In 1872 Weierstrass published an example of a continuous function $f : \mathbf{R} \to \mathbf{R}$ that is nowhere differentiable. Before Weierstrass's discovery it was often assumed that a continuous function is differentiable at most of the points in its domain.

Such a function is given by $f(x) = \sum_{n=1}^{\infty} \frac{1}{2^n} \cos(99^n x)$. Let's show that this function is continuous and nowhere differentiable. Note for each n and x that $\left|\frac{1}{2^n} \cos(99^n x)\right| \leq \frac{1}{2^n}$. Therefore $\sum_{n=1}^{\infty} \left|\frac{1}{2^n} \cos(99^n x)\right|$ converges. Since the real numbers are complete it is possible to show that this implies that the series $\sum_{n=1}^{\infty} \frac{1}{2^n} \cos(99^n x)$ converges to some number which we will call $f(x)$. It turns out that this convergence happens uniformly. That is, let $\epsilon > 0$. Then there exists an integer K such that for any x we have

$$\left|f(x) - \sum_{n=1}^{k} \frac{1}{2^n} \cos(99^n x)\right| < \epsilon \text{ whenever } k \geq K.$$

Let's show this. Since $\sum_{n=1}^{\infty} \frac{1}{2^n}$ converges to 1 there exists an integer K such that $k \geq K$ implies that

$$\left|1 - \sum_{n=1}^{k} \frac{1}{2^n}\right| < \epsilon.$$

We then have that

$$\left|f(x) - \sum_{n=1}^{k} \frac{1}{2^n} \cos(99^n x)\right| = \left|\sum_{n=k+1}^{\infty} \frac{1}{2^n} \cos(99^n x)\right|$$

$$\le \sum_{n=k+1}^{\infty} \left| \frac{1}{2^n} \cos(99^n x) \right|$$
$$\le \left| \sum_{n=k+1}^{\infty} \frac{1}{2^n} \right| =$$
$$\left| 1 - \sum_{n=1}^{k} \frac{1}{2^n} \right| < \epsilon.$$

This shows that the sequence of functions $f_1, f_2, ...$ defined by $f_k(x) = \sum_{n=1}^{k} \frac{1}{2^n} \cos(99^n x)$ converges uniformly to $f(x)$. Since each function f_k is a linear combination of continuous functions it is itself continuous. It turns out that if a sequence of continuous functions converges uniformly to another function then that other function must be continuous. Therefore f is continuous.

I am not completely sure how to show that f is not differentiable at any point in $\mathbf{R}$. I think the strategy is the following: let x be a point in $\mathbf{R}$. Because the function f oscillates very very fast it is possible to find a sequences $y_1, y_2, ...$ and $z_1, z_2, ...$ both in $\mathbf{R}$ both converging to x such that

$$\liminf \frac{f(y_n) - f(x_n)}{|y_n - x|} > \limsup \frac{f(z_n) - f(z)}{|z_n - z|}.$$

This implies that the sequences cannot have the same limit which means that f is not differentiable at the point x.

It turns out that Wierstrass's example is not a special case. Most continuous functions are not differentiable at any point in their domain. The way we will show this is the following. Let $C([0,1])$ be the set of real valued continuous functions on $[0,1]$. Let $\|\cdot\|$ be the supremum norm on this set. That is $\|f\| = \sup\{f(x) : x \in [0,1]\}$. It is possible to show that $C([0,1])$ with this norm is a Banach space. Therefore we can think of $C([0,1])$ as a complete metric space. What we will show is that the set of functions in $C([0,1])$ that are not differentiable at any point in $(0,1)$ is comeagre. That is, the set of functions in $C([0,1])$ that are differentiable at at least one point of $(0,1)$ is a meagre set.

Note that this is much stronger than saying that the set of functions in $C([0,1])$ that are differentiable at each point in $(0,1)$ is a meagre set. It is saying that the set of nowhere differentiable functions (functions like the one Wierstrauss discovered) is large in the sense that it is comeagre (its complement is a meagre set).

To do this let's derive an implication of differentiability at a point in $(0,1)$ and then show that the set of functions that have this property at a point in $(0,1)$ is a closed nowhere dense set. This shows that the set of continuous functions that are differentiable at at least one point in $(0,1)$ is a subset of a meagre set and so is itself a meagre set.

So let $f : [0,1] \to \mathbf{R}$ be a continuous function (i.e. an element of $C([0,1])$) that is differentiable at a point y in $(0,1)$. Then there exists an integer N such that $f'(y) \leq N$. This implies that if $z > y$ is sufficiently close to y then $\frac{f(z)-f(y)}{z-y} \leq N$ and this implies that $f(z) - f(y) \leq N(z-y)$. Likewise, if $x < y$ is sufficiently close to y then $\frac{f(y)-f(x)}{y-x} \leq N$ and this implies that $f(y) - f(x) \leq N(y-x)$. If we choose N sufficiently large then we can say that if z and x are such that

$$y - \frac{1}{N} < x < y < z < y + \frac{1}{N} and \frac{1}{N} \leq y \leq 1 - \frac{1}{N}$$

, then

$$f(z) - f(y) \leq N(z-y) and f(y) - f(x) \leq N(y-x)$$

. Combining these two inequalities gives

$$f(z) - f(x) \leq N(z-x).$$

So what we have show is that if f is differentiable at the point y in $(0,1)$, then there exists an integer N such that

$$f(z) - f(x) \leq N(z-x) whenever y - \frac{1}{N} < x < y < z < y + \frac{1}{N} and \frac{1}{N} \leq y \leq 1 - \frac{1}{N}.$$

Let D_N denote the set of functions f in $C([0,1])$ such that there exists a point $y \in (0,1)$ such that $f(z) - f(x) \leq n(z-x)$ whenever $y - \frac{1}{N} < x < y < z < y + \frac{1}{N}$ and $\frac{1}{N} \leq y \leq 1 - \frac{1}{N}$. Then the union $D_1 \cup D_2 \cup ...$ contains the set of functions that are differentiable at some point y in $(0,1)$. Denote this set by D. We will show that each set D_N is closed and nowhere dense. Therefore the set of functions in $C([0,1])$ that are differentiable at at least one point y in $(0,1)$ is meagre. This shows that the set of functions in $C([0,1])$ that are nowhere differentiable is comeagre.

Let's first show that D_N is closed. Let $f_1, f_2, ...$ be a sequence of functions in D_N converging to f. There corresponds to this sequence a sequence of points $y_1, y_2, ...$ in $\left[\frac{1}{N}, 1 - \frac{1}{N}\right]$ such that for each $n = 1, 2, ...$ we have $f_n(z) - f_n(x) \leq N(z-x)$ whenever $y_n - \frac{1}{N} < z < y_n < x < y_n + \frac{1}{N}$. Since $\left[\frac{1}{N}, 1 - \frac{1}{N}\right]$ is compact there exists a subsequence $y_{k_1}, y_{k_2}, ...$ of $y_1, y_2, ...$ converging to a point y in $\left[\frac{1}{N}, 1 - \frac{1}{N}\right]$. Therefore $f(z) - f(x) \leq N(z-x)$ whenever $y - \frac{1}{N} < z < y < x < y + \frac{1}{N}$. This implies that f belongs to D_N. Therefore D_N is a closed set.

Let's show that D_N is nowhere dense. What does this mean? It means that D_N does not contain any open subsets of $C([0,1])$. Any open subset contains an open ball around any of its points so it is enough to show that D_N does not contain any open balls. Now let's use another theorem of Weierstrass called the Weierstrass approximation theorem. This theorem says that the set of polynomial function is a dense subset of the metric space $C([0,1])$. So if D_N does contain an open set then that open set contains a polynomial function f_0. And then centered at that polynomial function f_0 must be an open ball contained in the open set and so contained in D_N. So to show that D_N is nowhere dense we can show that if f_0 is a polynomial function and ϵ is a positive real number, then $B_\epsilon(f_0)$ contains points not in D_N (that is, contains a point f such that for all $y \in (0,1)$ we have $f(z) - f(x) > N(z-x)$ whenever

$$y - \frac{1}{N} \leq z < y < x < y + \frac{1}{N} and \frac{1}{N} \leq y \leq 1 - \frac{1}{N}).$$

Now recall The Mean Value Theorem.

The Mean Value Theorem: Let $f : [a,b] \to \mathbf{R}$ be a continuous function on the closed interval $[a,b]$ and differentiable on the open interval (a,b) where $a < b$. Then there exists a point c in (a,b) such that

$$f'(c) = \frac{f(b) - f(a)}{b-a}.$$

Since f_0 is a polynomial function it is continuously differentiable on $[0,1]$. Therefore there exists a number C such that $C \leq f_0'(c)$ for all c in $(0,1)$. By The Mean Value Theorem this implies that $C \leq \frac{f_0(z)-f_0(x)}{z-x}$ whenever $1 > z > x > 0$. That is, $f_0(z) - f_0(x) \geq C(z-x)$ whenever $0 < x < z < 1$.

Let's try to modify f_0 to get a nowhere differentiable function while staying close to f_0. Let $f : [0,1] \to \mathbf{R}$ be the element of $C([0,1])$ defined by the equation $f(x) = f_0(x) + \frac{1}{M}\sin(M^2 x)$. Note that two things happen as M gets large. First for a fixed x the number $\frac{1}{M}\sin(M^2x)$ gets very small. So if we choose M large enough we can make f belong to the ball

$$B_\epsilon(f_0) = \{g \in C([0,1]) : \sup_{x \in [0,1]} |g(x) - f_0(x)| < \epsilon\}.$$

The second thing that happens is that $\frac{1}{M}\sin(M^2x)$ becomes a function that oscillates faster and faster. This will be what allows us to make f not belong to D_N (that is, for all $\frac{1}{N} \leq y \leq 1 - \frac{1}{N}$ there exists

$$y - \frac{1}{N} < x < y < z < y + \frac{1}{N}$$

such that $f(z) - f(x) \geq N(z-x)$)

Why is this so? Let y be such that $\frac{1}{N} \leq y \leq 1 - \frac{1}{N}$ and choose z and x such that

$$y - \frac{1}{N} < x < y < z < y + \frac{1}{N}$$

. Then we have that

$$f(z) - f(x) = f_0(z) - f_0(x) + \frac{1}{M}\sin(M^2z) - \frac{1}{M}\sin(M^2x) \geq C(z-x) + \frac{1}{M}(\sin(M^2z) - \sin(M^2x))$$

Let's use The Mean Value Theorem again. The derivative of $\frac{1}{M}\sin(M^2x) is M\cos(M^2x)$. By The Mean Value Theorem there exists a point c in (x,z) such that

$$M\cos(M^2c) = \frac{\frac{1}{M}\sin(M^2z) - \frac{1}{M}\sin(M^2x)}{z-x}.$$

Therefore

$$f(z) - f(x) \geq (C + M\cos(M^2c))(z-x).$$

What I need to show now is the following: Let r, K, and J be positive numbers. Then there exists points z and x such that $y - r < x < y < z < y + r$ and a number $M > J$ such that

$$\frac{\frac{1}{M}\sin(M^2z) - \frac{1}{M}\sin(M^2x)}{z-x} \geq K.$$

If this is true then for $r = \frac{1}{N}$ we can choose points x and z and a number $M > 0$ so that

$$\frac{\frac{1}{M}\sin(M^2z) - \frac{1}{M}\sin(M^2x)}{z-x}$$

is bigger than $N - C$ and such that f belongs to $B_\epsilon(f_0)$. We will then have that

$$f(z) - f(x) > N(z-x).$$

And this will imply that f does not belong to D_N. This will show that D_N is nowhere dense. I'll leave this claim to you because I can't figure out how to do it.

Then we will have shown that each the set of functions in $C([0,1])$ that are differentiable at some point y in $(0,1)$ is a meagre set because it is contained in the countable union of the closed and nowhere dense sets $D_1, D_2, ...$ This means that the set of functions in $C([0,1])$ that are nowhere differentiable is comeagre.

34. HILBERT SPACE

Let V be a Banach space and let V_0 be a closed subspace of V. The set V/V_0 is a set of equivalence classes of V such that if $w \in V/V_0$, then $v_1, v_2 \in w$ if and only if there exists $v_0 \in V_0$ such that $v_1 + v_0 = v_2$. From this construction we can define a map $\pi : V \to V/V_0$ that sends $v \in V$ to the equivalence class to which it belongs $\pi(v) \in V/V_0$. We have shown that V/V_0 with the norm given by $\|w\|_{V/V_0} = \inf\{\|v\|_V : \pi(v) = w\}$ is a Banach space. A question we could ask about this Banach space is if $w \in V/V_0$, then does there exist a $v \in V$ such that $\|w\|_{V/V_0} = \|v\|_V$? That is, is the infimum in the equation $\|w\|_{V/V_0} = \inf\{\|v\|_V : \pi(v) = w\}$ achieved?

Consider the following counterexample. Let

$$V = \{f : [0,1] \to \mathbf{R} : f \text{ is continuous and } f(0) = 0\}.$$

The set V with the norm given by $\|f\|_V = \max\{|f(x)| : x \in [0,1]\}$ (the "sup norm") is a Banach space. Let $V_0 = \{f \in V : \int f = 0\}$. Define $F : V \to \mathbf{R}$ by the formula $F(f) = \int f$. This is a bounded linear operator. Since F is bounded the set $F^{-1}(\{0\})$ is a closed subset of V. It is now easy to see that V_0 is a closed subspace of V. Therefore V/V_0 with the norm given by

$$\|w\|_{V/V_0} = \inf\{\|f\|_V : \pi(f) = w\}$$

is a Banach space. Is the infimum achieved?

What is V/V_0? If $w \in V/V_0$, then $f_1, f_2 \in w$ if and only if there exists $f \in V_0$ such that $f_1 + f_0 = f_2$. This implies that $\int f_1 = \int f_2$. Now suppose that $f_1, f_2 \in V$ are such that $\int f_1 = \int f_2$. Let $f_0 = f_2 - f_1$. Then $f_0 \in V_0$ and $f_1 + f_0 = f_2$. That is, $f_1, f_2 \in \pi(f_2 - f_1)$. This shows that two functions $f_1, f_2 \in V$ belong to the same element $w \in V/V_0$ if and only if $\int f_1 = \int f_2$.

Let's consider the element w in V/V_0 that is the set of functions in V whose integral is equal to one half. Does there exist a function f in V such that $\|w\|_{V/V_0} = \|f\|_V$? Firstly, what is $\|w\|_{V/V_0}$? It is equal to $\frac{1}{2}$. Any function $f \in V$ that that takes values no greater than $\frac{1}{2}$ cannot be in w because for such a function we have $\int f < \frac{1}{2}$. Therefore if $f \in w$, then $\|f\|_V > \frac{1}{2}$. However consider the sequence of functions $f_1, f_2, ...$ defined by $f_n(x) = nyx$ for $0 \leq x \leq \frac{1}{n}$ and $f_n(x) = y$ for $\frac{1}{n} \leq x \leq 1$ where y is the number that solves the equation $\int f = \frac{1}{2}$. That is, the equation

$$\frac{y}{2n} + y\left(1 - \frac{1}{n}\right) = \frac{1}{2}$$

. The sequence $f_1, f_2, ...$ belongs to w and the sequence of numbers $\|f_1\|_V, \|f_2\|_V, ...$ converges to $\frac{1}{2}$. This shows that $\|w\|_{V/V_0} = \frac{1}{2}$ and that there exists no $f \in w$ such that $\|w\|_{V/V_0} = \|f\|_V$.

When is the answer yes? Let $V = \mathbf{R}^2$. Let V_0 be the line defined by the equation $y = x$ (which is a closed subspace of $\mathbf{R}^2$). Then V/V_0 with the norm $\|w\|_{V/V_0} = \inf\{\|v\|_{\mathbf{R}^2} : \pi(v) = w\}$ is a Banach space. An element of V/V_0 is simply a translation of V_0. That is, a line defined by the equation $y = x + c$ where c is a number. So asking whether there exists a point v such that $\|w\|_{V/V_0} = \|v\|$ is the same as asking if there is a point $v \in w$ that is closest to the origin. It must be that this point should the vector $v \in w$ that is orthogonal to V_0. That is, if $v = (v_1, v_2)$ is the solution then it should satisfy the equation $(v_1, v_2) \cdot (1, 1) = 0$. That is $v_1 = -v_2$. And if w is the line defined by the equation $y = x + c$ then we also need that $v_2 = v_1 + c$. Therefore $v_2 = \frac{c}{2}$ and $v_1 = -\frac{c}{2}$. So there exists a point $v \in V$ such that $\|w\|_{V/V_0} = \|v\|_V$ and this point is unique.

This suggests that the answer to our question is "yes" when there is a notion of orthogonality in the Banach space V.

So let V be a Banach space. Let's try to construct a function $< \cdot, \cdot >: V \times V \to \mathbf{R}$ such that two vectors $v_1, v_2 \in V$ are orthogonal if and only if $< v_1, v_2 >= 0$. Now for a general Banach space it is not completely clear what orthogonality should mean. For example, consider the set of continuos functions $f : [0,1] \to \mathbf{R}$ with the sup norm. This is a Banach space. What does it mean for two continuous functions to be orthogonal? To get around this problem let's try to write down the properties we expect $< \cdot, \cdot >$ to have. We can do this by thinking about the set $\mathbf{R}^2$. Let's ask what are the defining properties of $< \cdot, \cdot >$ on $\mathbf{R}^2$ and then apply this as our definition of orthogonality in an arbitrary Banach space.

In $\mathbf{R}^2$ the dot product of two vectors gives an idea of the angle between the two vectors. Let $x = (x_1, x_2)$ and $y = (y_1, y_2)$ be vectors in $\mathbf{R}^2$ and let θ be the angle between these vectors. It is possible to show that $\|x\|\|y\| \cos(\theta) =< x, y >$ where $< x, y >= x_1 y_1 + x_2 y_2$. We also have that $\sqrt{< x, x >} = \|x\|$ for any vector x in $\mathbf{R}^2$.

Let's start with the idea that if v, w are elements of a Banach space V then there exists an angle θ between v and w. Let's require that the function $< \cdot, \cdot >$ tells us (for example, through an equation like the one above for $\mathbf{R}^2$) what θ is.

A property we would like $< \cdot, \cdot >$ to have is symmetry. That is the angle between v and w is the same as the angle between w and v. That is $< v, w >=< w, v >$ for all $v, w \in V$.

Let's just demand that the equation $\|x\|\|y\| \cos(\theta) =< x, y >$ is the defining property of $< \cdot, \cdot >$ for a Banach space. What I would like to do is to write down some properties for $< \cdot, \cdot >$ that satisfy this definition and then show that in a Banach space we can derive the equation $\|x\|\|y\| \cos(\theta) =< x, y >$ from these properties. My guess is that this can be done.

The function $< \cdot, \cdot >: V \times V \to \mathbf{R}$ where V is a vector space is called an inner product and I'm told its defining properties are

(1) Symmetry: $< v, w >=< w, v >$ for all $v, w \in V$. (2) Bilinearity: $< v + v', w >=< v, w > + < v', w >$ for all $v, v', w \in V$ and $< \lambda v, w >= \lambda < v, w >$ for all $v, w \in V$ and $\lambda \in \mathbf{R}$. (3) Positive definite: $< v, v >\geq 0$ with equality if and only if $v = 0$.

I'm not sure if starting from these properties one can show that if V is a Banach space then it makes sense to say that the angle between points $x, y \in V$ is given by the equation $\|x\|\|y\| \cos(\theta) =< x, y >$. Doing this certainly gives a $< \cdot, \cdot >$ which has the symmetry property. I think the other properties will follow if we agree that the angle between a vector and itself is zero and that the angle between two vectors in the same "quadrant" is between zero and ninety degrees.

One thing we can say is that if $< \cdot, \cdot >$ is an inner product on a vector space V, then the function $\| \cdot \|_V : V \to \mathbf{R}$ defined by the formula $\|v\|_V = \sqrt{< \cdot, \cdot >}$ is a norm.

Let's prove this. We get nonnegativity by the positive definite property of $<\cdot,\cdot>$. We get linearity by the bilinear and symmetry property of $<\cdot,\cdot>$:

$$\|\lambda v\|_V = \sqrt{<\lambda v, \lambda v>} = \sqrt{\lambda < v, \lambda v>} = \sqrt{\lambda < \lambda v, v>}$$
$$= \sqrt{\lambda^2 < v, v>} = |\lambda|\sqrt{<v,v>} = |\lambda|\|v\|_V.$$

For the triangle inequality we want to show that $\sqrt{<v+w, v+w>} \leq \sqrt{<v,v>} + \sqrt{<v,v>}$. Since each side is positive this is equivalent to showing the inequality holds when we square each side. That is, it is equivalent to show that

$$<v+w, v+w> \leq <v,v> + 2\sqrt{<v,v><w,w>} + <w,w>$$

.Using the bilinear and symmetry property of $<\cdot,\cdot>$ this is equivalent to showing that

$$<v,w>^2 \leq <v,v><w,w>.$$

By the bilinear property of $<\cdot,\cdot>$ we can assume that $\|v\|_V = \|w\|_V = 1$ (just divide both sides by $\|v\|_V^2\|w\|_V^2$ and relabel v and w). So the inequality we want to show is $<v,w>^2 \leq 1$ which is the same as showing that

$$<v,w> \leq 1.$$

Let's try to show this inequality. The positive definite, symmetry, and bilinear property of $<\cdot,\cdot>$ implies that

$$<v,v> -2 <v,w> + <w,w> = <v-w, v> + <w-v, w>$$
$$= <v-w, v> - <v-w, w> = <v, v-w> - <w, v-w>$$
$$= <v, v-w> + <-w, v-w> = <v-w, v-w> \geq 0.$$

In our case this inequality implies that

$$1 - 2 <v,w> + 1 \geq 0.$$

Therefore $<v,w> \leq 1$ which is the thing we wanted to show.

Let's go back to our original problem. Let V be a Banach space and let V_0 be a closed subspace of V. Then V/V_0 with the norm

$$\|w\|_{V\mathbb{V}_0} = \inf\{\|v\|_V : \pi(v) = w\}$$

is a Banach space. It will turn out that if V is an inner product space such that $\|v\|_V = <v,v>$ for all v in V then for each $w \in V/V_0$ there exists a point $v \in w$ such that $\|w\|_{V/V_0} = \|v\|_V$. This means that v is the element of w that is at minimum distance from the origin. This will be a useful property so let's distinguish Banach spaces V that are also inner product spaces with the property that $\|v\|_V = \sqrt{<v,v>}$ for all v in V by calling them Hilbert spaces.

35. THE CLOSEST POINT TO A CLOSED SUBSPACE OF A HILBERT SPACE

Recall that a Hilbert space is a Banach space V whose norm is obtained from an inner product $< \cdot, \cdot >: V \times V \to \mathbf{R}$ such that $\|v\| = \sqrt{< v.v >}$ for all v in V.

That is, a Hilbert space is a Banach space in which we have a coherent way to measure the angle between any two vectors.

Our motivation for defining Hilbert spaces was the following minimization problem. Let V be a Banach space and let V_0 be a closed subspace of V. Let v be a point in V. Is there a point in $\{v\} + V_0$ that is closest to the origin?

Recall that V/V_0 with the norm $\| \cdot \|_{V/V_0} : V/V_0 \to \mathbf{R}$ defined by the equation $\|w\|_{V/V_0} = \inf\{\|v\|_V : \pi(v) = w\}$ is a Banach space and that an element w of V/V_0 is defined by the property that $v_1, v_2 \in V$ belong to w if and only if there exists a point v_0 in V_0 such that $v_1 + v_0 = v_2$, that is, if and only if $v_2 - v_1$ belongs to V_0 and this implies that w can be represented by the set $V_0 + \{v_1\}$. That is, if v belongs to w then $v - v_1$ belongs to V_0 so that v belongs to $V_0 + \{v_1\}$. And if v belongs to $V_0 + \{v_1\}$ then $v - v_1$ belongs to V_0 which implies that v belongs to w.

We showed last time that the minimization problem:

"Let V be a Banach space and let V_0 be a closed subspace of V. Let v be a point in V. Is there a point in $\{v\} + V_0$ that is closest to the origin?"

does not always have a solution. Our intuition suggests that the solution to this problem must be the point x in $\{v\} + V_0$ that is orthogonal to V_0. Intuition also suggests that the solution will be unique – there will not be more than one in $\{v\} + V_0$ that is orthogonal to V_0. But orthogonality requires that we can measure the angle between vectors. This is why we introduced the idea of a Hilbert space. (If V is a Hilbert space and v_1, v_2 are points in V then we define the angle θ between x and y to be the solution to the equation $\|x\|\|y\| \cos(\theta) =< x, y >$.)

Note that if V is a Hilbert space then the inner product on V is determined by the norm because

$$< v, w >= \frac{< v + w, v + w > - < v, v > - < w, w >}{2} = \frac{\|v + w\|^2 - \|v\|^2 - \|w\|^2}{2}.$$

Note also that given a Banach space you can try to define an inner product using this formula. You can see that it will be positive definite and symmetric. It may not be bilinear. It is possible to show that a Banach space is a Hilbert space if and only if

$$\|u + v + w\|^2 = \|u + v\|^2 + \|u + w\|^2 + \|v + w\|^2 - \|v\|^2 - \|u\|^2 - \|w\|^2.$$

The usual norm on Euclidian space is better than other norms because it satisfies this. (How many inner products are there on Euclidian space?)

So one example of a Hilbert space is the set $\mathbf{R}^n$ with the usual norm.

Another example of a Hilbert space is $L^2(E)$ where E is a measurable subset of $\mathbf{R}^n$ and the inner product is defined by the equation $< f, g >= \int fg$ (its easy to see that this is positive definite, symmetric, and bilinear). One concern is that $< f, g >= \infty$ for some functions f and g but this is ruled out by Holder's inequality which implies that $\int fg \leq \|f\|_{L^2}\|g\|_{L^2}$. We have that $\|f\|_{L^2} = \sqrt{< f, f >}$ and previously shown that $L^2(E)$ with the L^2-norm is complete. Therefore $L^2(E)$ with this inner product is a Hilbert space.

Let's prove the following proposition.

Proposition: Let V be a Hilbert space, $v \in V$ a vector and $V_0 \subseteq V$ a closed subspace. Then there exists a unique "closest point" to v contained in V_0.

Consider the map $\pi : V \to V/V_0$. We have

$$\|\pi(v)\|_{V/V_0} = \inf\{\|u\|_V : \pi(u) = \pi(v)\}$$

$$= \inf\{\|u\|_V : \text{there exists } v_0 \in V_0 \text{ such that } u + v_0 = v\}$$

$$= \inf\{\|v - v_0\|_V : v_0 \in V_0\}.$$

We will prove the Proposition if we can show that there exists a unique point y in V_0 such that $\|v - y\|_V = \|\pi(v)\|_{V/V_0}$. Let C denote $\|\pi(v)\|_{V/V_0}$. For any $\epsilon > 0$ we can choose a point v_ϵ in V_0 such that $C \leq \|v - v_\epsilon\|_V < C + \epsilon$. Consider the sequence of points $v_1, v_{\frac{1}{2}}, v_{\frac{1}{3}}, \cdots in V_0$. Let's try to show this is a Cauchy sequence. If it is a Cauchy sequence then it converges to some point y in V_0. We then have that $\|v - y\|_V = C$. So showing that the sequence $v_1, v_{\frac{1}{2}}, v_{\frac{1}{3}}, \cdots$ is a Cauchy sequence shows that there exists a closest point to v in V_0.

Consider two points a, b in V_0 such that $\|v - a\|_V$ and $\|v - b\|_V$ are less than $C + \epsilon$. What can we say about $\|a - b\|_V$? Because V is a Hilbert space we have that $C^2 \leq < v - a, v - a > < (C + \epsilon)^2$ and $C^2 \leq < v - b, v - b > < (C + \epsilon)^2$.

Consider the midpoint of a and b. Let $w = v - \frac{a+b}{2}$ and $d = \frac{a-b}{2}$. Then we have that $v - a = w - d$ and $v - b = w + d$. We then have that

$$(C + \epsilon)^2 > < w - d, w - d >=< w, w > -2 < w, d > + < d, d >,$$

$$(C^2 + \epsilon)^2 > < w + d, w + d >=< w, w > +2 < w, d > + < d, d > .$$

Adding these gives $< w, w > + < d, d > < (C + \epsilon)^2$. Since $C \leq < w, w >$ we have that $< d, d > < 2C\epsilon + \epsilon^2$. Therefore $\|a - b\|_V < 2\sqrt{2C\epsilon + \epsilon^2}$.

Let $\epsilon > 0$. Choose N such that

$$2\sqrt{2C\frac{1}{N} + \left(\frac{1}{N}\right)^2} < \epsilon$$

. Then

$$\|v_{\frac{1}{n}} - v_{\frac{1}{m}}\|_V < \epsilon$$

whenever $n, m \geq N$. This shows that the sequence $v_1, v_{\frac{1}{2}}, \cdots$ is a Cauchy sequence which shows that there exists a closest point to v in V_0. Call one such closest point y.

Let's show that the point y is the only closest point. Let u denote the point $v - y$.

For any vector $v_0 \in V_0$ and any real number λ we have that $\|u\|_V \leq \|u - \lambda y\|$. This implies that

$$< u, u > \leq < u - \lambda y, u - \lambda y > = < u, u > -2\lambda < u, y > +\lambda^2 < y, y > .$$

Therefore

$$2\lambda < u, y > \leq \lambda^2 < y, y > .$$

If $< y, y > = 0$ then $y = 0$ so that if we choose a nonzero λ it must be that $< u, y > = 0$. If $< y, y > > 0$ then it also must be that $< u, y > = 0$ because otherwise we could choose a small positive λ and violate the inequality. Let $V_0^\perp$ denote the set

$$\{v \in V : < v, v_0 > = 0 \text{ for all } v_0 \in V_0\}$$

. What we have shown is that u belongs to $V_0^\perp$.

Now let y' denote a point in V_0 such that $v - y'$ belongs to $V_0^\perp$. Then $y - y'$ belongs to V_0 and $(v - y') - (v - y)$, which equals $y - y'$, belongs to $V_0^\perp$. That is, $< y - y', y - y' > = 0$ and this implies that $y = y'$. Therefore the closest point is unique.

36. THE RIESZ REPRESENTATION THEOREM

Let S be a set. Define the set $l^2(S)$ to be the set of functions $f : S \to \mathbf{R}$ such that $\sum_{s \in S} f(s)^2 < \infty$. If S is infinite this sum is meant to mean the supremum over all finite sums.

Define the function $< f, g >: l^2(S) \times l^2(S) \to \mathbf{R}$ by the formula

$$< f, g > = \sum_{s \in S} f(s) g(s).$$

I claim that this is an inner product. It is easy to see that this function is symmetric, bilinear, and positive definite. We only need to check that $< f, g > | < \infty$

for all f, g in $l^2(S)$. Note that

$$(f(s) - g(s))^2 = f(s)^2 - 2f(s)g(s) + g(s)^2 \geq 0$$

and this implies that

$$\sum_{s \in S} f(s) g(s) \leq \sum_{s \in S} \frac{f(s)^2 + g(s)^2}{2}.$$

This shows that $< f, g >< \infty$. We also have that if f is in $l^2(S)$, then $-f$ is in $l^2(S)$. And by bilinearity $< -f, g >= - < f, g >> -\infty$. Therefore $< \cdot, \cdot >$ is an inner product. So $l^2(S)$ is a normed vector space. Note also that the sum is absolutely convergent so summing the terms of the series in a different order will not change the value to which it converges.

Let $\lambda : l^2(S) \to \mathbf{R}$ be a bounded linear function. Let e_s be the element of $l^2(S)$ defined by $e_s(t) = 1$ if $t = s$ otherwise $e_s(t) = 0$. Now define a function $f : S \to \mathbf{R}$ by $f(s) = \lambda(e_s)$. We want to show that f belongs to $l^2(S)$ and that for any function g in $l^2(S)$ we have that $< f, g >= \lambda(g)$. So far we have that $< f, e_s >= \lambda(e_s)$ for all s in S.

We want to show that f is a square summable function. That is, that f belongs to $l^2(S)$.

We also want to show for each function g in $l^2(S)$ that $< f, g >= \lambda(g)$. If g has finite support then that $< f, g >= \lambda(g)$ follows from the bilinearity of $< \cdot, \cdot >$. I now claim that each function g in $l^2(S)$ is the limit of some sequence of functions in $l^2(S)$ that have finite support.

What does it mean for a sequence of functions $g_1, g_2, \cdots$ in $l^2(S)$ to converge to g. It means that $\lim_n \|g - g_n\|_{\ell^2} = 0$. That is, it means that

$$\lim_n \sqrt{\sum_{s \in S} (g(s) - g_n(s))^2} = 0.$$

I claim that if a function g belongs to $l^2(S)$ then it does not have uncountable support. That is, the set $\{s \in S : g(s) \neq 0\}$ is not uncountable. I claim that if a function $g : S \to \mathbf{R}$ has uncountable support then the sum $\sum_s \in Sg(s)^2$ (the supremum over all finite sums) diverges and therefore g does not belong to $l^2(S)$. How can I show this? Let K be an uncountable set of positive numbers. Consider the set $K_n = \{k \in K : k \geq \frac{1}{n}\}$. I claim that there exists an integer n such that the set K_n is infinite. If this is not true then K is countable because it can be written as $K_1 \cup K_2 \cup \cdots$ which is a countable union of finite sets and so is a countable set. So there exists such an integer n and for this n we have that $\sum_{k \in K} k > \sum_{k \in K_n} \frac{1}{n} = \infty$. This proves the claim.

So if g belongs to $l^2(S)$ then g has countable support $s_1, s_2, \cdots$. Define g_n by the formula

$$g_n = g(s_1)e_{s_1} + g(s_2)e_{s_2} + \cdots + g(s_n)e_{s_n}.$$

Then

$$g - g_n = g(s_{n+1})e_{s_{n+1}} + g(s_{n+2})e_{s_{n+2}} + \cdots$$

which converges to the zero function as n tends to infinity.Therefore

$$\lim_n \|g - g_n\|_{\ell^2} = \| \lim_n g - g_n\| = 0$$

by the continuity of the norm. That is, g_n converges to g in the ℓ^2 norm.

Now let g belong to $l^2(S)$ and let $g_1, g_2, \cdots$ be a sequence of functions with finite support in $l^2(S)$ which converge to g. Then by the continuity of λ and the ℓ^2 inner product we have that

$$< g, f >= \lim_n < g_n, f_n >= \lim_n \lambda(g_n) = \lambda(g).$$

Now we would like to show that f belongs to $l^2(S)$. That is we would like to show that $\sum_{s\in S} f(s)^2 < \infty$. Let's recall the definition of f. We defined $f : S \to \mathbf{R}$ by the formula $f(s) = \lambda(e_s)$ where $e_s(t) = 1$ if $t = s$ and 0 otherwise.

Recall the operator norm. We have that $\|\lambda\|$ is the smallest number C such that $|\lambda(g)| \leq C\|g\|$ for all g in $l^2(S)$. Let S_0 be a finite subset of S and define the function $g : S \to \mathbf{R}$ by the formula $g(s) = f(s)$ if $s \in S_0$ and $g(s) = 0$ otherwise. Because g has finite support it is an element of $l^2(S)$. We then have that

$$\|g\|^2 = \sum_{s\in S_0} f(s)^2 =< f, g >= \lambda(g) \leq \|\lambda\|\|g\|.$$

Dividing both sides by $\|g\|$ gives

$$\|g\| \leq \|\lambda\| and \sum_{s\in S_0} f(s)^2 \leq \|\lambda\|.$$

We know that $\|\lambda\|$ is finite because we assumed λ to be bounded. Since this inequality holds for all finite subsets S_0 of S the supremum over all finite subsets is no greater than λ. In symbols this means that $\sum_{s\in S} f(s)^2 \leq \|\lambda\|$. Therefore f belongs to $l^2(S)$.

What we have shown is the following:

Theorem: If $\lambda : l^2(S) \to \mathbf{R}$ is a bounded linear function, then there exists an element $f \in \ell^2(S)$ (defined by the formula $f(s) = \lambda(e_s)$) such that $\lambda(g) =< f, g >$ for each g in $l^2(S)$.

Now let f belong to $l^2(S)$. Define $\lambda : l^2(S) \to \mathbf{R}$ by the formula $\lambda(g) =< g, f >$. By bilinearity of $< \cdot, \cdot >$ the function λ is linear. Recall The Cauchy Schwartz Inequality`http://en.wikipedia.org/wiki/Cauchy-Schwarz_inequality`

$$| < g, f > | \leq \|f\|\|g\|$$

with equality if and only if $f = \alpha g$ for some $\alpha \in \mathbf{R}$. This implies that λ is bounded and that $\|\lambda\| = \|f\|$. This shows that the map $f \mapsto < \cdot, f >$ is an isometry. The previous theorem shows that is is surjective. Therefore we have that $l^2(S)$

and its dual are isometrically isomorphic via this map. This is called The Riesz Representation Theorem.

37. EVERY HILBERT SPACE IS ISOMETRICALLY ISOMORPHIC TO $l^2(S)$ FOR SOME S

Last time I proved The Riesz-Fischer Theorem.

The Riesz-Fischer Theorem: The Hilbert space $l^2(S)$ is isometrically isomorphic to its dual via the map $f \mapsto < \cdot, f >$.

I will now show that if V is a Hilbert space then V is isomorphically isometric to $l^2(S)$ for some set S.

Because V is a Hilbert space it has a Hilbert basis $\{v_s\}_{s\in S}$ where S is an index set for the Hilbert basis. We will use this as the set S when talking about $l^2(S)$.

In the last email I showed that any function in $l^2(S)$ can be approximated as the limit of finite linear combinations of functions from the orthonormal set $\{e_s\}_{s\in S}$ where $e_s(t) = 1$ if $t = s$ and 0 otherwise.

Define the function $\phi : l^2(S) \to V$ by saying that it is continuous linear map such that $\phi(e_s) = v_s$. This map is well defined because if g is an element of $l^2(S)$ with finite support then g can be written as a finite linear combination of elements of $\{e_s\}_{s\in S}$ and then $\phi(g)$ is given by the linearity of ϕ. Otherwise the support of g is countable (we showed before that it cannot be uncountable) in which case g can be written as the limit of functions which have finite support. Then $\phi(g)$ is given by the continuity of g.

I claim that the function ϕ is an isometric isomorphism. It is isometric because iff, g belong to $l^2(S)$, then f can be written as $\lim_n \sum_{s\in S_n} f(s)e_s$ and g can be written as $g = \lim_n \sum_{t\in S_n} g(s)e_s$ where $S_1, S_2, \cdots$ is a sequence of finite sets increasing to a set that contains the support of f and the support of g. Using the continuity of ϕ and the continuity of the inner product on V we get

$$< \phi(f), \phi(g) >_V = \lim_n < \sum_{s\in S_n} f(s)v_s, \sum_{t\in S_n} g(t)v_t >_V$$

$$= \lim_n \sum_{s\in S_n} f(s)g(s) = < f, g >_{l^2} .$$

Therefore $\|\phi(f)\|_V = \|f\|_{l^2}$.

Why is the function ϕ a bijection?

It is surjective because each point v in V can be written as the limit of finite linear combinations of element of the set $\{v_s\}_{s\in S}$. From this it is easy to construct the function f in $l^2(S)$ such that $\phi(f) = v$.

Why is the function ϕ injective? Any isometry is injective because if $f \neq g$, then $\|f - g\|_{l^2} \neq 0$ so that $\|\phi(f - g)\|_V = \|\phi(f) - \phi(g)\|_V \neq 0$ which implies that $\phi(f) \neq \phi(g)$. Since ϕ is an isometry it is injective.

So we have proved the following theorem.

Theorem: If V is a Hilbert space, then there exists a set S such that $l^2(S)$ is isometrically isomorphic to V.

The cardinality of the set S is called the Hilbert dimension of V.

Note that if the sets S and T have the same cardinality then $l^2(S)$ is isometrically isomorphic to $l^2(T)$. This is true by the above construction using $\{e_t\}_{t\in T}$ as the Hilbert basis of $l^2(T)$.

38. A HILBERT BASIS FOR L^2

Let E be a measurable subset of $\mathbf{R}^n$. Recall that $L^2(E)$ is the set of measurable functions (modulo functions that are the same almost everywhere) $f : E \to \mathbf{R}$ such that $\left(\int f^2\right)\frac{1}{2} < \infty$. With the inner product $< \cdot, \cdot >: L^2(E) \times L^2(E) \to \mathbf{R}$ defined by $< f, g >= \left(\int fg\right)^{\frac{1}{2}}$ this is a Hilbert space.

What is a Hilbert basis for $L^2(E)$? If E has measure zero then the only element of $L^2(E)$ is the zero function. Therefore $\{0\}$ is an orthonormal basis for $L^2(E)$ because it is a maximal orthonormal subset of $L^2(E)$.

Suppose that $E = [0, 2\pi]$. We have that $\{\sin(nx), \cos(nx)\}_{n=0}^{\infty}$ is an orthogonal subset of $L^2(E)$. Let's try to get an orthonormal subset of $L^2(E)$. We have $\int_0^{2\pi} sin(nx)^2 = \pi$ for $n = 1, \cdots$ and 0 for $n = 0$. We have $\int_0^{2\pi} cos(nx)^2 = \pi$ for $n = 1, 2, \cdots$ and 2π for $n = 0$. Therefore

$$\{\frac{\sin(nx)}{\sqrt{\pi}}, \frac{\cos(nx)}{\sqrt{\pi}}, \frac{1}{\sqrt{2\pi}}\}_{n=1}^{\infty}$$

is an orthonormal subset of $L^2(E)$. Is it maximal? That is, is each element of $L^2(E)$ the limit of finite linear combinations of elements of this set. Another way to say this is that any function in $L^2(E)$ can be approximated by a finite linear combination of sines and cosines.

We will prove that the answer is yes. Let A be the set of functions that are finite linear combinations of sines and cosines on $[0, 2\pi]$. The set A is much smaller than $L^2([0, 2\pi])$.

Let B be the set of continuous functions on $[0, 2\pi]$ having the same values at their endpoints. We have that A is contained in B and that B is contained in $L^2([0, 2\pi])$.

Let S denote the set of simple functions on $[0, 2\pi]$. The set S is contained in $L^2([0, 2\pi])$. We also know that any function in $L^2([0, 2\pi])$ is equal to the limit of simple functions.

Let T denote the set of step functions on $[0, 2\pi]$. A step function is a linear combination of characteristic functions on intervals. I claim that any simple function is equal to the limit in the L^2 norm of step functions. To see this, let K be a measurable subset of $[0, 2\pi]$. Let $\epsilon > 0$. By the definition of outer measure there exists a sequence of intervals $I_1, I_2, \cdots$ which we can assume to be disjoint whose union contains K and such that

$$\mu(K) \leq \sum_{i=1}^{\infty} \mu(I_i) \leq \mu(K) + \epsilon.$$

We can choose finitely many of these intervals $I_1, I_2, \cdots, I_n$ such that

$$\mu(K) - \epsilon < \sum_{i=1}^{n} \mu(I_i) < \mu(K) + \epsilon.$$

Then we have that

$$\|\chi_E - \sum_{i=1}^{n} \chi_{I_i}\|_{L^2} = \sqrt{\int (\chi_E - \sum_{i=1}^{n} \chi_{I_i})^2}$$

$$= \sqrt{\mu(E \triangle \bigcup_{i=1}^{n} I_i)} \leq 2\sqrt{2\epsilon}.$$

Since we can approximate characteristic functions arbitrarily well we can approximate simple functions arbitrarily well. Note that this is convergence in the L^2 norm. Such convergence would not hold in the sup norm for example because the symmetric difference is never empty.

Since we can write any function in $L^2([0, 2\pi])$ as a limit of simple functions and we can write any simple function as a limit of step functions we can also write any function in $L^2([0, 2\pi])$ as a limit of step functions.

So far we have that A which is the set of finite linear combinations of sines and cosines on $[0, 2\pi]$ is a subset B which is the set of continuous functions on $[0, 2\pi]$ with equal values at its endpoints. We also have that the set D which is the set of step functions on $[0, 2\pi]$ is a dense subset of $L^2(E)$.

I now claim that any step function is a limit in the L^2 norm of continuous functions on $[0, 2\pi]$ with equal values at its endpoints. This is fairly obvious. Just mollify the step function you want to get a piecewise continuous function. This piecewise continuous function can be made arbitrarily close to the step function in the L^2 norm.

Now recall The Stone-Weierstrass Theorem.

The Stone-Weierstrass Theorem: Let $X \subset \mathbf{R}^n$ be a compact set. Then any continuous function on X can be written as the limit of a uniformly convergent sequence of polynomials.

Now suppose $n = 2$. Note that the set B which is the set of continuous functions on $[0, 2\pi]$ which have the same value at their endpoints is in bijection the set D of continuous functions on the unit circle $S^1 \subseteq \mathbf{R}^2$. The bijection is $f(\theta) =$

$g(\sin(\theta), \cos(\theta))$. The Stone-Weierstrass Theorem implies that any element of D can written as the limit of a uniformly convergent sequence of polynomials and thus any element of A can be written as the limit of a uniformly convergent sequence whose elements are finite linear combinations of sines and cosines. Uniform convergence implies convergence in L^2.

Therefore any function in $L^2([0, 2\pi])$ can be approximated in the L^2 norm by a simple function and each simple function can be approximated in the L^2 norm by a step function and each step function can be approximated in the L^2 norm by a continuous function with the same value at its endpoints and each continuous function with the same value at its endpoints can be uniformly approximated by a polynomial of sines and cosines. Since uniform convergence implies convergence in L^2 we have that any function in $L^2([0, 2\pi])$ can be approximated in the L^2 norm by a polynomial of sines and cosines.

Hopefully a polynomial of sines and cosines can be approximated by our orthonormal set $\{\frac{\sin(nx)}{\sqrt{\pi}}, \frac{\cos(nx)}{\sqrt{\pi}}, \frac{1}{\sqrt{2\pi}}\}_{n=1}^{\infty}$. If this is true then this orthonormal set is a Hilbert basis for $L^2([0, 2\pi])$.