SPACES: FROM ANALYSIS TO GEOMETRY AND BACK

MASS 2011 LECTURE NOTES

1. Lecture 1 (8/22/11): Introduction

There are many problems in analysis which involve constructing a function with desirable properties or understanding the properties of a function without completely precise information about its structure that cannot be easily tackled using direct "hands on" methods. A fruitful strategy for dealing with such problems is to recast it as a problem concerning the geometry of a well-chosen space of functions, thereby making available the many techniques of geometry. For example, one can construct solutions for a large class of ordinary differential equations by applying the "contraction mapping principle" from the theory of metric spaces to an appropriate space of continuous functions.

The application of geometric techniques to spaces of functions proved so successful that it led to the birth of an independent area of mathematical research known as functional analysis. This subject has taken on a life of its own, but the deep interplay between geometry and analysis is still very relevant. The goal of this course is to investigate some of the basic ideas and techniques which drive this interplay.

1.1. Metric Spaces.

The notion of a complete metric space is the most fundamental geometric abstraction relevant to our tour of functional analysis. The formalism attempts to capture the essential features of the intuitive notion of distance.

Definition 1.1. A metric space is a set X equipped with a function $d : X \times X \to \mathbb{R}$ which satisfies the following axioms:

- Positive Definiteness: $d(x, y) \geq 0$ for every $x, y \in X$, and $d(x, y) = 0$ if and only if $x = y$
- Symmetry: $d(x, y) = d(y, x)$ for every $x, y \in X$
- Triangle Inequality: $d(x, y) \leq d(x, z) + d(y, z)$ for every $x, y, z \in X$

Example 1.2. The set $\mathbb{Q}$ of all rational numbers equipped with the distance function $d(x, y) = |x - y|$ is a metric space.

Example 1.3. The set $\mathbb{R}$ of all real numbers equippped with the distance function $d(x, y) = |x - y|$ is a metric space.

Example 1.4. The plane $\mathbb{R}^2$ equipped with the distance function

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

is a metric space.

Example 1.5. The plane $\mathbb{R}^2$ equipped with the distance function

$$d(\mathbf{x}, \mathbf{y}) = \max \{|x_1 - x_2|, |y_1 - y_2|\}$$

is a metric space.

There is an important intrinsic difference between Example 1.2 and the other examples. Intuitively, the set of all rational numbers with the distance function above has "holes" where the irrational numbers should be. Capturing this intuition precisely is a somewhat subtle matter involving the idea of convergent sequences.

Definition 1.6. Let (X, d) be a metric space and let $\{x_n\}$ be a sequence of points in X. Declare that $\{x_n\}$ converges to a limit x in X if for every real number $\varepsilon > 0$ there exists N such that $d(x, x_n) < \varepsilon$ whenever $n \geq N$.

Exercise 1.7. Prove that if $\{x_n\}$ converges to both x and y in X then $x = y$.

To see that the notion of convergence helps to distinguish between Example 1.2 and Example 1.3, observe that there is a sequence of the form $x_1 = 3$, $x_2 = 3.1$, $x_3 = 3.14$, $x_4 = 3.141, \ldots$ converges to the real number π in $\mathbb{R}$ but has no limit in $\mathbb{Q}$. Thus the "holes" in $\mathbb{Q}$ correspond to sequences which converge in $\mathbb{R}$ but whose limit is not in $\mathbb{Q}$. This can be captured intrinsically in $\mathbb{Q}$ by considering sequences which "accumulate on themselves" in the sense of the following definition.

Definition 1.8. Let (X, d) be a metric space and let $\{x_n\}$ be a sequence of points in X. Declare that $\{x_n\}$ is a Cauchy sequence if for every real number $\varepsilon > 0$ there exists N such that $d(x_m, x_n) < \varepsilon$ whenever $n, m \geq N$.

Exercise 1.9. Prove that every convergent sequence is Cauchy.

Definition 1.10. A metric space (X, d) is complete if every Cauchy sequence converges to a point in X.

Thus Example 1.3 is a complete metric space while Example 1.2 is not.

Exercise 1.11. Prove that the metric spaces of Example 1.4 and 1.5 are both complete.

1.2. **Banach Spaces.** The metric spaces which most frequently arise in functional analysis come equipped with a linear structure which is compatible with the metric. To be precise, we assume that V is a vector space over $\mathbb{R}$ and that V has a metric d which is compatible with its linear structure in the sense that $d(u + w, v + w) = d(u, v)$ and $d(\lambda u, \lambda v) = |\lambda|\, d(u, v)$. By the first compatibility property d is determined by the values $d(u, 0)$ where u ranges over V, and so we are led to the following definition.

Definition 1.12. Let V be a vector space over $\mathbb{R}$. A norm on V is a function $\|\cdot\| : V \to \mathbb{R}$ which satisfies the following axioms:

- Positive Definiteness: $\|u\| \geq 0$ for every $u \in V$ and $\|u\| = 0$ if and only if $u = 0$
- Homogeneity: $\|\lambda u\| = |\lambda|\, \|u\|$ for every $u \in V$ and every $\lambda \in \mathbb{R}$
- Convexity: $\|u + v\| \leq \|u\| + \|v\|$ for every $u, v \in V$

Exercise 1.13. Show that if V is a vector space and d is a metric on V which is compatible with its linear structure in the sense above then $\|u\| := d(u, 0)$ is a norm on V. Conversely, show that if $(V, \|\cdot\|)$ is a normed vector space then $d(u, v) := \|u - v\|$ is a metric on V which is compatible with its linear structure.

Thanks to Exercise 1.13, every normed vector space $(V, \|\cdot\|)$ has the structure of a metric space, and unless otherwise specified the metric will be understood to be $d(u, v) = \|u - v\|$. Examples 1.3, 1.4, and 1.5 each consist of a vector space over $\mathbb{R}$

equipped with a compatible metric, so those three examples correspond to normed vector spaces.

Definition 1.14. A Banach space is a normed vector space which is complete as a metric space.

It can be shown that any finite dimensional normed vector space over $\mathbb{R}$ is a Banach space. (This will be discussed in later lectures and in the homework.) However, the spaces of functions which arise in functional analysis are very often infinite dimensional, and it can be quite difficult to prove that an infinite dimensional normed vector space is complete. Fortunately there are many useful examples of infinite dimenaional Banach spaces among spaces of functions.

One of the most fundamental examples is the vector space $C[0,1]$ of continuous real valued functions on the unit interval in the real line. It carries the following norm, often called the "uniform" norm:

$$\|f\| = \sup_{x\in[0,1]} |f(x)|$$

Exercise 1.15. Prove that this really is a norm on $C[0,1]$.

Proving that $C[0,1]$ with the uniform norm really is a Banach space is not too difficult, but it requires many of the tools of basic real analysis. Note that the fact that the norm is well-defined already uses the extreme value theorem!

Theorem 1.16. $C[0,1]$ *with the uniform norm is a Banach space.*

Proof. All that remains is to prove completeness. Let f_n be a Cauchy sequence in $C[0,1]$ relative to the uniform norm. This means that for every $\varepsilon > 0$ there exists N_ε such that if $n, m \geq N_\varepsilon$ then $|f_n(x) - f_m(x)| < \varepsilon$ for every $x \in [0,1]$. In particular, $f_n(x)$ is a Cauchy sequence of real numbers for each x, and hence it has a limit. Call the limit $f(x)$, and so define a function $f : [0,1] \to \mathbb{R}$.

First, we show that f_n converges to f in the uniform norm (a norm defined on the space of bounded functions on $[0,1]$). Given $\varepsilon > 0$, let N_ε be as above so that $|f_n(x) - f_m(x)| < \varepsilon$ whenever $n, m \geq N_\varepsilon$. Fix n and pass to the limit as m tends to infinity to conclude that $|f_n(x) - f(x)| < \varepsilon$ whenever $n \geq N_\varepsilon$ (independently of x). This implies that $\sup_{x\in[0,1]} |f_n(x) - f(x)| \leq \varepsilon$, and thus $\|f_n - f\| \leq \varepsilon$. Thus $f_n \to f$ in norm. We need only show that $f \in C[0,1]$, i.e. that f is continuous.

To that end, let $\varepsilon > 0$ be given, and set $N = N_\varepsilon$ as above. Fix a point $x_0 \in [0,1]$, and choose δ_N so that if $|x - x_0| < \delta_N$ then $|f_N(x) - f_N(x_0)| < \varepsilon$; δ_N exists by continuity of f_N. Thus if $|x - x_0| < \delta_N$ then $|f(x) - f(x_0)| \leq |f(x) - f_N(x)| + |f_N(x) - f_N(x_0)| + |f_N(x_0) - f(x_0)| < 3\varepsilon$. Thus f is continuous, and the proof is complete. $\square$

Exercise 1.17. Show that the space $P[0,1]$ of polynomial functions defined on $[0,1]$ is an infinite dimensional subspace of $C[0,1]$ which is not complete in the uniform norm. (This will appear as a homework problem.)

We feel that in a sense metrics in the plane (and corresponding structures of $\mathbb{R}^2$ as a Banach space) defined in examples 1.4 and 1.5 are different. To make this rigorous we define an *isometry* between metric spaces to be an invertible map that preserves distances. Then we can say that there is no isometry between $\mathbb{R}^2$ with the Euclidean metric from Example 1.4 and the metric from Example 1.5. The main idea is that if two metric spaces are isometric then the sets (in fact they are groups) of their self-isometries are isomorphic. In both cases translations are isometries by definition. But if one considers only isometries that fix the origin there are infinitely many of those for the Euclidean metric and only finitely many for the other metric. This will be discussed in detail later.

2. Lecture 2 (8/24/11): Examples of Banach Spaces

So far we have three finite dimensional examples of Banach spaces (examples 1.2, 1.2, and 1.2) and one infinite dimensional example (via Theorem 1.16). Before continuing to develop the theory of Banach spaces, we will enlarge our supply of examples.

We begin with some finite dimensional examples. We have already remarked (deferring the proof) that every finite dimensional normed vector space over $\mathbb{R}$ is a Banach space, but it will be useful to investigate a few specific examples all the same.

First, recall that every vector space V over $\mathbb{R}$ of finite dimension n is linearly isomorphic to $\mathbb{R}^n$. To see this, fix a basis $\{v_1, \dots, v_n\}$ for V and fix the standard basis $\{e_1, \dots, e_n\}$ for $\mathbb{R}^n$. There is a unique linear map $T : V \to \mathbb{R}^n$ which satisfies $T(v_i) = e_i$ for each i, and it is straightforward to check that T is an isomorphism. Thus when constructing examples of finite dimensional Banach spaces there is no loss of generality in assuming that the underlying vector space is $\mathbb{R}^n$. The problem then reduces to constructing interesting norms on $\mathbb{R}^n$.

Given a positive real number p, define a function $\|\cdot\|_p : \mathbb{R}^n \to \mathbb{R}$ by the formula:

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

This function is positive definite and homogeneous in the sense of Definition 1.12, but it can fail to be convex. When it is convex it defines a norm on $\mathbb{R}^n$ called the ℓ^p norm.

Exercise 2.1. Show that $\|\cdot\|_p$ is a norm when $p = 1$ and $p = 2$ (hint: use the Cauchy-Schwarz inequality to handle $p = 2$). For which other values of p is $\|\cdot\|_p$ a norm? (This will appear as a homework problem.)

There is a related norm which, informally, corresponds to the case "$p = \infty$". It is defined by the formula:

$$\|\mathbf{x}\|_\infty = \max_{i=1,\dots,n} \{|x_i|\}$$

It is quite straightforward to check that this really does define a norm on $\mathbb{R}^n$, called the ℓ^∞ norm. Note that examples 1.4 and 1.5 correspond to $\mathbb{R}^2$ equipped with the ℓ^2 and ℓ^∞ norms, respectively.

We now turn to analogues of the ℓ^p norms defined on infinite dimensional spaces. There are a number of challenges which arise when passing from finite dimensions to infinite dimensions; one such challenge is that infinite dimensional normed vector spaces are not automatically complete (see Exercise 1.17). An even more basic problem is that two infinite dimensional vector spaces need not be isomorphic to each other, and even when they are there is not always a natural way to identify them. So when we construct infinite dimensional Banach spaces we have to be careful to specify the underlying vector space as well as the norm with which we want to equip it.

In the case of the ℓ^p norms, the problem arises because of convergence issues. It would be natural to replace $\mathbb{R}^n$ with the vector space of infinite sequences of real numbers and to replace the finite sum which defines the ℓ^p norm on $\mathbb{R}^n$ with an infinite sum, but for arbitrary sequences of real numbers the sum will not converge. The solution to this problem is to define it away:

Definition 2.2. For each positive real number p, ℓ^p is the vector space of all countably infinite sequences $\mathbf{x} = \{x_n\}$ of real numbers with the property that $\sum_{n=1}^{\infty} |x_n|^p < \infty$. Correspondingly, ℓ^∞ is the vector space of all countably infinite sequences $\mathbf{x}$ of real numbers with the property that $\sup_{n\in\mathbb{N}}\{|x_n|\} < \infty$.

For $p < \infty$ the formula $\|\mathbf{x}\|_p = \left(\sum_{n=1}^{\infty} |x_n|^p\right)^{1/p}$ yields a well-defined function on ℓ^p. Similarly, $\|\mathbf{x}\|_\infty = \sup_{n\in\mathbb{N}}\{|x_n|\}$ is a well defined function on ℓ^∞. It is straightforward to check that all of these functions are positive definite and homogeneous, but it is again not immediately obvious when they are convex.

Lemma 2.3. *For each positive real number p (together with $p = \infty$), $\|\cdot\|_p$ is a norm on ℓ^p if it is a norm on $\mathbb{R}^n$ for each n.*

Proof. Given $\mathbf{x} \in \ell^p$, define $\mathbf{x}^{(n)}$ to be the sequence $(x_1, \ldots, x_n, 0, 0, \ldots)$. We have that $\|\mathbf{x}\|_p = \lim_n \left\|\mathbf{x}^{(n)}\right\|_p$; this is clear for $p = \infty$ and it follows from the definition of convergent series for $p < \infty$. If $\|\cdot\|_p$ is a norm on $\mathbb{R}^n$ then we have $\left\|(x+y)^{(n)}\right\|_p \leq \left\|x^{(n)}\right\|_p + \left\|y^{(n)}\right\|_p$, and this inequality is preserved upon passing to the limit as $n \to \infty$. This proves that $\|\cdot\|_p$ is convex. $\square$

Later in the course we shall prove that ℓ^p equipped with $\|\cdot\|_p$ is a Banach space whenever $\|\cdot\|_p$ is a norm.

2.1. Metric Topology.

We shall now take a break from our discussion of Banach spaces to build some machinery for working with complete metric spaces in the abstract. We will find that for many arguments and constructions we are not primarily interested in the precise details of a metric, but rather in the notions of convergence and continuity (the *topology*) that it provides for the ambient space. Here we will explore the rudiments of the topology of metric spaces.

For what follows, let (X, d) be a metric space. Given a point $x \in X$ and a positive real number r, let $B_r(x)$ denote the open unit ball of radius r centered at x; thus $B_r(x) = \{y \in X : \ d(x, y) < r\}$.

Definition 2.4. A subset $U \subseteq X$ is said to be open if for each $x \in U$ there is a positive real number $r(x)$ such that $B_{r(x)}(x) \subseteq U$. A subset $C \subseteq X$ is said to be closed if its complement is open.

Exercise 2.5. Prove the following:

- X and the empty set are open sets.
- The union of any collection of open sets is an open set.
- The intersection of any finite collection of open sets is an open set.

These are the axioms for the abstract notion of a topology on a set X.

Exercise 2.6. Given a sequence $\{x_n\}$ in X, show that x_n converges to a point x if and only if for every open set U which contains x there exists N such that $x_n \in U$ for every $n \geq N$.

Exercise 2.6 shows that convergence of sequences in X depends on the metric only insofar as it depends on which sets are open. In other words, convergence of sequences is a topological notion.

Exercise 2.7. Show that a closed subspace $C \subseteq X$ contains the limit of every convergent sequence of points in C. Is the converse true?

One of the most important and fundamental concepts in topology is compactness. The reader is likely to have encountered properties of compact subspaces of the real line in a real analysis course, but we will need to deal with compactness in more general metric spaces.

Definition 2.8. X is compact if every sequence in X has a convergent subsequence.

The Heine-Borel theorem asserts that the compact subspaces of $\mathbb{R}^n$ are precisely the closed and bounded subspaces. However, it is a bit harder to characterize more general compact metric spaces. One problem is that closedness is is not intrinsic: the set of all rational numbers x such that $0 \leq x^2 < 2$ is a closed subset of $\mathbb{Q}$, but there are sequences of rational numbers in this set which converge to $\sqrt{2}$ in $\mathbb{R}$ that have no convergent subsequences in $\mathbb{Q}$. Replacing closedness with completeness solves this problem, but complete and bounded is still not equivalent to compact. In fact, it can be shown that the closed unit ball in an infinite dimensional Banach space is *never* compact! Nevertheless, complete and bounded are useful necessary conditions:

Lemma 2.9. *Every compact metric space X is complete and bounded.*

Proof. Suppose X is compact. Given a Cauchy sequence $\{x_n\}$ and $\varepsilon > 0$, choose N_ε so that $n, m \geq N_\varepsilon$ implies that $|x_n - x_m| < \varepsilon$. By compactness we can extract a convergent subsequence $\{x_{n_k}\}$, meaning there exists K_ε such that $k \geq K_\varepsilon$ implies $|x_{n_k} - x| < \varepsilon$ where x is the limit of the subsequence. Pick any $n \geq N_\varepsilon$, and choose k large enough so that $k \geq K_\varepsilon$ and $n_k \geq N_\varepsilon$. We get $|x_n - x| \leq |x_n - x_{n_k}| + |x_{n_k} - x| < 2\varepsilon$, so x_n converges to x. Thus we have shown that any Cauchy sequence in X converges, i.e. X is complete.

Now, suppose X is not bounded. We will use the fact that for any finite set S there is a point x in S such that the distance from x to every point in S is at least 1 - if this were not the case then the diameter of X would be no larger than two plus the diameter of S, contradicting the assumption that X is unbounded. This allows us to recursively define a sequence $\{x_n\}$ so that each x_n has distance at least 1 from each of $x_1, \ldots, x_{n-1}$. Such a sequence has no convergent subsequence, so X cannot be compact. $\square$

3. Lecture 3 (8/26/11):The Completion of a Metric Space

Lemma 2.9 suggests that there are two different reasons why a metric space might fail to be compact: a "local" reason (the set is not complete) and a "global" reason (the set is not bounded). For example, the failure of the set of all rational numbers in $[0, 1]$ to be compact is local in the sense that there are sequences with no convergent subsequences in a tiny neighborhood of any point, while the failure of $\mathbb{R}$ to be compact is global in the sense that the only sequences with no convergent subsequences are those which tend to infinity. Note, however, that Lemma 2.9 is not the end of the story: the unit ball in $C[0, 1]$ is a complete and bounded set which is not compact (see the homework). As we shall see, a criterion which is a bit stronger than boundedness is needed to obtain a necessary and sufficient condition for a metric space to be compact.

In this lecture we will consider the local aspect of compactness, i.e. completeness. We introduce a weaker notion of compactness which emphasizes this local behavior:

Definition 3.1. A metric space X is precompact if every sequence in X has a Cauchy subsequence.

Note that a precompact metric space is compact if and only if it is complete.

Exercise 3.2. Show that a closed subset of a complete metric space is complete. Deduce that if X is a complete metric space and $K \subseteq X$ is a precompact subspace then the closure of K (that is, the intersection of all closed subsets of X which contain K) is compact.

According to Exercise 3.2, we can enlarge a precompact space into a compact space if we can embed it in a complete space. This raises a question of independent interest: can we embed an arbitary metric space into a complete metric space? We shall show that every metric space can be realized as a dense subspace of a unique complete metric space. The proofs are a little technical and may appear intimidating, but in fact the only substantial conceptual step is the construction of the completion as a space of Cauchy sequences. Everything else follows directly, though perhaps not effortlessly, from the definitions.

Definition 3.3. Let (X, d) be a metric space. A completion for (X, d) is a complete metric space $(\widetilde{X}, \widetilde{d})$ equipped with a map $i : X \to \widetilde{X}$ with the following properties:

- i is an isometry, i.e. $d(x, y) = \widetilde{d}(i(x), i(y))$ for every $x, y \in X$
- $i(X)$ is dense in $\widetilde{X}$

We begin by proving that the completion of a metric space is unique if it exists, in the following sense:

Lemma 3.4. *Let $(\widetilde{X}, \widetilde{d}, i)$ and $(\widetilde{X}', \widetilde{d}', i')$ be two completions of a metric space (X, d). Then there is a unique bijective isometry $\phi : \widetilde{X} \to \widetilde{X}'$ with the property that $\phi \circ i = i'$*

Proof. First we define ϕ. Given any point $x \in \widetilde{X}$ there is a sequence $\{x_n\}$ in X such that $x = \lim i(x_n)$ since $i(X)$ is dense in $\widetilde{X}$. Since $\{i(x_n)\}$ is convergent in $\widetilde{X}$ and i is an isometry, $\{x_n\}$ is a Cauchy sequence in X. But i' is also an isometry, so $\{i'(x_n)\}$ is a Cauchy sequence in $\widetilde{X}'$. Since $\widetilde{X}'$ is complete, $\{i'(x_n)\}$ has a limit $x' \in \widetilde{X}'$. Define $\phi : \widetilde{X} \to \widetilde{X}'$ by $\phi(x) = x'$.

We must check that ϕ is well-defined, i.e. that x' is independent of the sequence $\{i(x_n)\}$ which converges to x. This follows from the fact that two sequences $\{a_n\}$ and $\{b_n\}$ in a metric space converge to the same point p if and only if $\lim d(a_n, b_n) = 0$.

Next, we check that ϕ is an isometry. Observe that if $\{i(x_n)\}$ converges to x and $\{i(y_n)\}$ converges to y in $\widetilde{X}$ then $\widetilde{d}(x, y) = \lim \widetilde{d}(i(x_n), i(y_n))$. From the calculation

$$\widetilde{d}(i(x_n), i(y_n)) = d(x_n, y_n) = \widetilde{d}'(i'(x_n), i'(y_n))$$

we deduce that

$$\widetilde{d}'(\phi(x), \phi(y)) = \lim \widetilde{d}'(i'(x_n), i'(y_n)) = \lim \widetilde{d}(i(x_n), i(y_n)) = d(x, y)$$

as desired.

That ϕ is injective follows from the fact that it is an isometry, and it is surjective since $i'(X)$ is dense in $\widetilde{X}'$. So ϕ is bijective. For any $x \in X$ we can write $i(x)$ as the limit of the constant sequence $\{i(x)\}$, and thus by definition $\phi(i(x))$ is the limit of the constant sequence $\{i'(x)\}$ which is of course just $i'(x)$. So $\phi \circ i = i'$.

Finally, to see that ϕ is unique note that the condition $\phi \circ i = i'$ defines ϕ unambiguously on the dense set $i(x) \subseteq \widetilde{X}$. Any continuous map (in particular, any isometry) is determined by its values on a dense set, so ϕ is unique. $\square$

We now have to show that the completion of a metric space actually exists. For specific examples it is usually possible to identify a more concrete model of the completion than the abstract construction described here, and the previous lemma guarantees that such concrete realizations are, up to isometry, just as good. Indeed, the details of our construction are rarely needed in practice.

Proposition 3.5. *Every metric space (X, d) has a completion.*

Proof. **Step 1**: Construct $(\widetilde{X}, \widetilde{d})$.

Let $\mathcal{S}$ denote the set of all Cauchy sequences in X. Define a relation $\sim$ on $\mathcal{S}$ as follows: given two Cauchy sequences $\mathbf{x} = \{x_n\}$ and $\mathbf{y} = \{y_n\}$, declare that $\mathbf{x} \sim \mathbf{y}$ if $\lim d(x_n, y_n) = 0$. This relation is clearly symmetric and reflexive, and it is transitive by the triangle inequality: if $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ are Cauchy sequences such that $\mathbf{x} \sim \mathbf{y}$ and $\mathbf{y} \sim \mathbf{z}$ then $\lim d(x_n, z_n) \leq \lim d(x_n, y_n) + \lim d(y_n, z_n) = 0$. Thus it is an equivalence relation, and we may define $\widetilde{X}$ to be the set of all equivalence classes of $\mathcal{S}$ for this relation.

We must equip $\widetilde{X}$ with a metric. Given two equivalence classes $\mathbf{x}, \mathbf{y} \in \widetilde{X}$, define $\widetilde{d}(\mathbf{x}, \mathbf{y}) = \lim d(x_n, y_n)$ where $\{x_n\}$ and $\{y_n\}$ are Cauchy sequences representing the classes $\mathbf{x}$ and $\mathbf{y}$, respectively. We need to check that this limit exists and that it is independent of the chosen representatives for $\mathbf{x}$ and $\mathbf{y}$. Given $\varepsilon > 0$, choose N large enough so that if $m, n \geq N$ then $d(x_m, x_n) < \varepsilon$ and $d(y_m, y_n) < \varepsilon$. For such m and n we have:

$$d(x_m, y_m) \leq d(x_m, x_n) + d(x_n, y_n) + d(y_n, y_m) < d(x_n, y_n) + 2\varepsilon$$

Thus $|d(x_m, y_m) - d(x_n, y_n)| < 2\varepsilon$ for $n, m \geq N$ which proves that $\{d(x_n, y_n)\}$ is a Cauchy sequence of real numbers. Thus this sequence has a limit by the completeness of $\mathbb{R}$, as desired. To see that the limit does not depend on the equivalence classes chosen, take two Cauchy sequences $\{x_n\}$ and $\{x'_n\}$ which both represent $\mathbf{x}$ (meaning that $\lim d(x_n, x'_n) = 0$). By the triangle inequality $|d(x_n, y_n) - d(x'_n, y_n)| \leq d(x_n, x'_n)$, so $|d(x_n, y_n) - d(x'_n, y_n)|$ converges to 0 and hence the two sequences have the same limit.

Now we must check that $\widetilde{d}$ is a metric. Symmetry and positivity of $\widetilde{d}$ follow immediately from the corresponding properties for d. It is also immediate that $\widetilde{d}(\mathbf{x}, \mathbf{x}) = 0$ for any $\mathbf{x} \in \widetilde{X}$. If $\widetilde{d}(\mathbf{x}, \mathbf{y}) = 0$ then $\lim d(x_n, y_n) = 0$ for representatives $\{x_n\}$ and $\{y_n\}$ of $\mathbf{x}$ and $\mathbf{y}$, respectively, and this is precisely the condition that $\{x_n\}$ and $\{y_n\}$ define the same equivalence class in $\tilde{X}$. Finally if $\{x_n\}$, $\{y_n\}$, and $\{z_n\}$ are Cauchy sequences then $d(x_n, z_n) \leq d(x_n, y_n) + d(y_n, z_n)$ by the triangle inequality for d, and this inequality persists upon taking a limit. This proves the triangle inequality for $\widetilde{d}$.

Step 2: Construct the isometry $i : X \to \widetilde{X}$.

The most natural idea is to define $i(x)$ to be the equivalence class of the constant sequence $\{x\}$. We must check that this map is an isometry and that $i(X)$ is dense in $\widetilde{X}$. To see that i is an isometry, simply write $\widetilde{d}(\{x\},\{y\}) = \lim d(x,y) = d(x,y)$. To show that i has dense image, we must show that every Cauchy sequence in X can be written as the limit in $\widetilde{X}$ of constant Cauchy sequences. Given a Cauchy sequence $\{x_n\}$ representing an equivalence class $\mathbf{x} \in \widetilde{X}$, define $\mathbf{x}^n \in i(X)$ to be the equivalence class of the constant sequence $x_n, x_n, x_n, \ldots$. We have $\widetilde{d}(\mathbf{x}, \mathbf{x}^n) = \lim_{m\to\infty} d(x_m, x_n)$, and this tends to zero as n tends to infinity since $\{x_n\}$ is Cauchy.

Step 3: Prove that $(\widetilde{X}, \widetilde{d})$ is complete.

We must show that every Cauchy sequence $\{\mathbf{x}^n\}$ in $\widetilde{X}$ has a limit. Each $\mathbf{x}^n$ is an equivalence class of Cauchy sequences in X, so choose a Cauchy sequence $\{x^n_m\}$ representing $\mathbf{x}^n$. We just proved that the set $i(X)$ of equivalence classes which have a constant sequence as a representative is dense in $\widetilde{X}$, so for every n there exists $\mathbf{y}^n$ with the property that $\widetilde{d}(\mathbf{x}^n, \mathbf{y}^n) < 1/n$. Let $y_n, y_n, y_n, \ldots$ be a constant sequence representing $\mathbf{y}^n$; by definition of $\widetilde{d}$ there exists M such that $d(x^n_m, y_n) < 1/n$ if $m \geq M(n)$.

Consider the sequence $y_1, y_2, y_3, \ldots$. We shall prove that this sequence is Cauchy and that its equivalence class $\mathbf{y}$ in $\widetilde{X}$ is the limit of $\{\mathbf{x}^n\}$. Unravelling the definitions, the statement that $\{\mathbf{x}^n\}$ is Cauchy means that for any ε there exists N such that if $n_1, n_2 \geq N$ then $d(x^{n_1}_m, x^{n_2}_m) < \varepsilon$ for m sufficiently large - say, larger than $M'(n_1, n_2)$. So let us choose n_1, n_2 greater than both N and $1/\varepsilon$, and m larger than the maximum of $M(n_1)$, $M(n_2)$, and $M'(n_1, n_2)$. We have $d(y_{n_1}, y_{n_2}) \leq d(y_{n_1}, x^{n_1}_m) + d(x^{n_1}_m, x^{n_2}_m) + d(x^{n_2}_m, y_{n_2}) < 1/n_1 + \varepsilon + 1/n_2 < 3\varepsilon$. This proves that $\{y_n\}$ is Cauchy.

It remains only to show that $\{\mathbf{x}^n\}$ converges to $\mathbf{y}$, i.e. that $\lim_{n\to\infty} \widetilde{d}(\mathbf{x}^n, \mathbf{y}) = 0$. By definition $\widetilde{d}(\mathbf{x}^n, \mathbf{y}) = \lim_{m\to\infty} d(x^n_m, y_m)$, and we know that $d(x^n_m, y_m) \leq 1/n$ for m sufficiently large by our choice of y_m above. So $\lim_{n\to\infty} \widetilde{d}(\mathbf{x}^n, \mathbf{y}) \leq \lim_{n\to\infty} 1/n = 0$, as desired. □

4. Lecture 4 (8/29/11): Non-Archimedian Completions of $\mathbb{Q}$

We noted above that it is often possible to construct the completion of a metric space X by embedding it in a complete metric space and taking its closure. Many examples of metric spaces that arise naturally in mathematics are presented as subspaces of complete metric spaces, so it is tempting to dismiss the labor required to prove Proposition 3.5 as unnecessary abstraction. However, there are important examples of metric spaces that do not come equipped with obvious complete spaces into which they embed. One family of such examples which is of fundamental importance in number theory is produced by using properties of prime numbers to equip $\mathbb{Q}$ with metrics not equivalent to the standard one.

Recall that the standard metric on $\mathbb{Q}$ is induced by a norm - the usual absolute value function. The completion of $\mathbb{Q}$ with respect to this norm is simply $\mathbb{R}$. Note that our definition of a norm in lecture 1 assumed that the ground field was $\mathbb{R}$, but there is no harm in regarding $\mathbb{Q}$ as a $\mathbb{Q}$-vector space and defining norms on $\mathbb{Q}$-vector spaces in an analogous way.

We now associate to each prime number p a norm on $\mathbb{Q}$. Given a nonzero rational number r, there is a unique way to write $r = p^n \frac{k}{l}$ where k and l are integers coprime to p. Define the *p-adic norm* of r to be $\|r\|_p = p^{-n}$, and define $\|0\|_p = 0$.

Lemma 4.1. $\|\cdot\|_p$ *is a norm on* $\mathbb{Q}$

Proof. It is clear from the definition that $\|\cdot\|_p$ is positive definite. If $r = p^n \frac{k}{l}$ and $r' = p^{n'} \frac{k'}{l'}$ where k, l, k', and l' are relatively prime to p then $rr' = p^{n+n'} \frac{kk'}{ll'}$, and kk' and ll' are both relatively prime to p. Thus $\|rr'\|_p = p^{-(n+n')} = p^{-n}p^{-n'} = \|r\|_p \|r'\|_p$, and homogeneity is established. Now take r and r' as above and assume $0 \leq n \leq n'$. We have:

$$r + r' = \frac{p^n kl' + p^{n'} k'l}{ll'}$$

It is clear that ll' is relatively prime to p, and the largest power of p which divides the numerator is at least n. It follows that $\|r + r'\|_p \leq p^{-n} = \max\{\|r\|_p, \|r'\|_p\}$. Similar calculations prove this inequality for any configuration of signs of n and n', so we have $\|r + r'\|_p \leq \max\{\|r\|_p, \|r'\|_p\} \leq \|r\|_p + \|r'\|_p$. $\square$

Exercise 4.2. Show that if p is replaced by any integer m in the definition of of the p-adic norm then the resulting function is positive definite and convex but not homogeneous. (This will be assigned as a homework problem.)

Note that in the proof of Lemma 4.1 we proved that $\|\cdot\|_p$ satisfies an inequality stronger than the triangle inequality: $\|r + r'\|_p \leq \max\{\|r\|_p, \|r'\|_p\}$. A norm with this property is said to be "non-Archimedean" and the corresponding metric is often called an "ultrametric".

$\mathbb{Q}$ equipped with the p-adic norm is a metric space, and by Proposition 3.5 it has a completion which we will denote by $\mathbb{Q}_p$.

Exercise 4.3. Show that the usual algebraic operations of addition, subtraction, multiplication, and division extend to $\mathbb{Q}_p$ and that $\mathbb{Q}_p$ equipped with these operations is a field.

Thus $\mathbb{Q}_p$ is often called the field of p-adic numbers. It has a wildly different algebraic and geometric structure compared to $\mathbb{R}$; we will now investigate some of these differences.

Exercise 4.4. Show that the sequnce $\{p^n\}$ converges to 0 while $\{p^{-n}\}$ diverges to infinity in $\mathbb{Q}_p$.

Exercise 4.5. Show that $\mathbb{Z}$ is a subset of the closed unit ball centered at 0 in $\mathbb{Q}_p$. Show that the closed unit ball centered at 0 is the same as the open unit ball centered at 0.

Proposition 4.6. $\mathbb{Z}$ *is precompact in the metric topology on* $\mathbb{Q}_p$.

Proof. Let $\{a_n\}$ be a sequence in $\mathbb{Z}$; we must extract a Cauchy subsequence. We will recursively define such a subsequence as follows. There are infinitely many integers a_n but only finitely many residue classes in $\mathbb{Z}/p\mathbb{Z}$, so there is a residue class c_1 and infinitely many values of n with the property that $a_n \equiv c_1 \bmod p$. Similarly, there is a residue class c_2 in $\mathbb{Z}/p^2\mathbb{Z}$ and infinitely many values of n with the property that $a_n \equiv c_1 \bmod p$ and $a_n \equiv c_2 \bmod p^2$. By the principle of recursive definition, there is a subsequence $\{a_{n_k}\}$ with the property that $a_{n_k} \equiv a_{n_l} \bmod p^i$ for every $i \leq k$ and every $l \geq k$. In other words, the residue classes of a_{n_l} for every l larger than k are determined by the residue classes of $a_{n_k} \bmod p, p^2, \ldots, p^k$.

Let us prove that $\{a_{n_k}\}$ is Cauchy. If i and j are larger than K then by construction $a_{n_i} - a_{n_j} = p^K m$ for some integer m. It follows that $\left\|a_{n_i} - a_{n_j}\right\|_p \leq p^{-K}$.

Given any $\varepsilon > 0$, choose K large enough so that $p^{-K} < \varepsilon$; this will guarantee that $\left\|a_{n_i} - a_{n_j}\right\|_p < \varepsilon$ for i, j larger than K. □

Our next aim is to show that in fact $\mathbb{Z}$ is dense in the closed unit ball centered at 0 in $\mathbb{Q}_p$. The closed unit ball in a complete metric space is automatically complete, so this will prove that the closed unit ball (and hence every other closed ball) is compact. The fact that the closed balls in $\mathbb{R}$ are compact is crucial in analysis, and the fact that the same statement is true in $\mathbb{Q}_p$ makes it possible to develop "p-adic analysis".

Proposition 4.7. *The closure of* $\mathbb{Z}$ *in* $\mathbb{Q}_p$ *is* $B_1(0)$.

Proof. $\mathbb{Z} \subseteq B_1(0)$ by Exercise 4.5, so it suffices to show that $\mathbb{Z}$ is dense in $B_1(0)$. Note that the only possible values that the norm of a p-adic number can take are p^n with n an integer and 0, so $B_1(0)$ consists of those p-adic numbers whose norm is p^n for $n \leq 0$ together with 0. Thus the set $B_{\mathbb{Q}}$ of all rational numbers whose denominators (written in lowest terms) are coprime with p is a dense subset of $B_1(0)$. We will show that $\mathbb{Z}$ is dense in $B_{\mathbb{Q}}$.

Take $\frac{1}{k} \in B_{\mathbb{Q}}$, so that k is coprime to p. We want to show that for every n there exists m such that $\left\|m - \frac{1}{k}\right\|_p \leq p^{-n}$. Unravelling the definitions, this equation is equivalent to the statement that there are integers a and b such that b is coprime to p and $m - \frac{1}{k} = p^n \frac{a}{b}$, or in other words $\frac{km-1}{k} = \frac{2^n a}{b}$. We are forced to choose $b = k$, so we have reduced the problem to showing that there are integers a and m such $km - 1 = p^n a$. In other words, we want to solve the equation $kx + p^n y = 1$ for x and y. By the division algorithm this equation has a solution if and only if k and p^n are coprime, and this is true by assumption.

This shows that $\frac{1}{k}$ is in the closure of $\mathbb{Z}$ if k is coprime to p, and this easily implies that $\frac{l}{k}$ is also in the closure of $\mathbb{Z}$ for any integer l. Thus $\mathbb{Z}$ is dense in $B_{\mathbb{Q}}$ and therefore in $B_1(0)$. □

Exercise 4.8. Given any negative integer n, find a sequence of positive integers which converges to n in $\mathbb{Q}_p$. Deduce that $B_1(0)$ is the closure of $\mathbb{N}$.

5. Lecture 5 (8/31/11): Examples of Closures of Function Spaces

We now return to a context more relevant to functional analysis: normed vector spaces of functions. We have already seen one of the most important examples, namely the space $C[0,1]$ of continuous functions on the unit interval equipped with the uniform norm. There are many other interesting spaces of functions and many interesting norms that one can put on them; in this sectino we will consider some examples.

Example 5.1. Recall that a function f on $[0,1]$ is piecewise linear if $[0,1]$ can be partitioned into a finite collection of subintervals such that f restricts to a linear function on each subinterval. Let $PL[0,1]$ denote the set of all continuous piecewise linear functions on $[0,1]$; this is a linear subspace of $C[0,1]$ and we can equip it with the uniform norm. What is the completion of $PL[0,1]$ with this norm?

By Proposition 1.16 and Lemma 3.4, we can naturally identify the completion of $PL[0,1]$ with its closure in $C[0,1]$. We shall prove that in fact $PL[0,1]$ is dense in $C[0,1]$.

Given any $f \in C[0,1]$ and any $\varepsilon > 0$ we must show that there is a continuous piecewise linear function g such that $\|f - g\| < \varepsilon$. Every continuous function on

$[0,1]$ is uniformly continuous, so there exists $\delta > 0$ such that $|f(x) - f(y)| < \frac{\varepsilon}{2}$ whenever $|x - y| < \delta$. Choose N large enough so that $\frac{1}{N} < \delta$, and let g be the unique function in $PL[0,1]$ with the property that g restricts to a linear function on $[\frac{k}{N}, \frac{k+1}{N}]$ for every integer k from 0 to $N-1$ and $g(\frac{k}{N}) = f(\frac{k}{N})$ for every integer k from 0 to N.

Observe that for any k from 0 to $N-1$ and any $x \in [\frac{k}{N}, \frac{k+1}{N}]$ we have that $|f(x) - f(\frac{k}{N})| < \frac{\varepsilon}{2}$ since $|x - \frac{k}{N}| \leq \frac{1}{N} < \delta$. Moreover since g is linear on $[\frac{k}{N}, \frac{k+1}{N}]$ we have

$$\left|g(x) - g\left(\frac{k}{N}\right)\right| \leq \left|g\left(\frac{k+1}{N}\right) - g\left(\frac{k}{N}\right)\right| = \left|f\left(\frac{k+1}{N}\right) - f\left(\frac{k}{N}\right)\right| < \frac{\varepsilon}{2}$$

Now for any $x \in [0,1]$ there exists k from 0 to N such that $x \in [\frac{k}{N}, \frac{k+1}{N}]$. For such k we have $|f(x) - g(x)| \leq \left|f(x) - f(\frac{k}{N})\right| + \left|f(\frac{k}{N}) - g(\frac{k}{N})\right| + \left|g(\frac{k}{N}) - g(x)\right| < \varepsilon$. Thus $|f(x) - g(x)| < \varepsilon$ for every $x \in [0,1]$ and hence $\|f - g\| < \varepsilon$. This completes the proof that $PL[0,1]$ is dense in $C[0,1]$.

Example 5.2. Equip $C[0,1]$ with the function $\|f\|_1 = \int_0^1 |f(x)|\, dx$. It is clear that this function is positive definite and homogeneous, and it is convex since $\|f + g\| = \int_0^1 |f(x) + g(x)|\, dx \leq \int_0^1 |f(x)|\, dx + \int_0^1 |g(x)|\, dx$. Thus $\|\cdot\|_1$ is a norm on $C[0,1]$.

It is a bit difficult to understand the topology of $C[0,1]$ with this norm. Unlike with the uniform norm, wherein convergence is equivalent to uniform convergence of functions, convergence with respect to $\|\cdot\|_1$ does not even imply pointwise convergence.

Consider the sequence of functions $f_n(x) = x^n$ in $C[0,1]$. The pointwise limit of this sequence is the discontinuous function which takes the value 0 at $x \in [0,1)$ and the value 1 at $x = 1$. But $\|f_n\|_1 = \int_0^1 x^n dx = \frac{1}{n}$, so f_n converges to the constant function 0 with respect to $\|\cdot\|_1$.

There are even more dramatic examples. Given any interval $[a,b]$ in the real line, let $T_{a,b}$ be the triangular wave of height 1 supported on $[a,b]$. Thus $T_{a,b}$ restricts to 0 outside $[a,b]$, it restricts to the linear function joining $(a,0)$ and $(\frac{a+b}{2}, 1)$ on $[a, \frac{a+b}{2}]$, and it restricts to the linear function joining $(\frac{a+b}{2}, 1)$ and $(b,0)$ on $[\frac{a+b}{2}, b]$. Now define a sequence f_n in $C[0,1]$ by $f_n = T_{\frac{j}{2^k}, \frac{j+1}{2^k}}$ where $n = 2^k + j$ with $0 \leq j < 2^k$. For every $x \in [0,1]$ there are infinitely many n such that $f_n(x) > \frac{1}{2}$ (in fact n can be chosen to make $f_n(x)$ arbitrarily close to 1), and there are also infinitely many n such that $f_n(x) = 0$. So $f_n(x)$ does not converge at any point x. However if $n = 2^k + j$ with $0 \leq j < 2^k$ then $\|f_n\|_1 = \frac{1}{2^k}$, so f_n converges to 0 with respect to $\|\cdot\|_1$.

Thus $(C[0,1], \|\cdot\|_1)$ is a metric space with a rather confusing notion of convergence. It is not a complete metric space. For n large define $f_n \in C[0,1]$ to be the function which restricts to 0 on the interval $[0, \frac{1}{2} - \frac{1}{n}]$, to the linear function joining $(\frac{1}{2} - \frac{1}{n}, 0)$ and $(\frac{1}{2}, 1)$ on the interval $[\frac{1}{2} - \frac{1}{n}, \frac{1}{2}]$, and to 1 on the interval $[\frac{1}{2}, 1]$. One can check by direct calculation that $\|f_n - f_m\|_1 \leq \frac{1}{2}\left|\frac{1}{n} - \frac{1}{m}\right|$, so the sequence $\{f_n\}$ is Cauchy in norm. However, we can show that $\{f_n\}$ cannot have a limit in $C[0,1]$. Let $PC[0,1]$ be the space of all functions which are piecewise continuous and continuous from the left, and observe that $\|\cdot\|_1$ extends to a norm on $PC[0,1]$. The Cauchy sequence $\{f_n\}$ has a limit in $PC[0,1]$, namely the function f which restricts to 0 on $[0, \frac{1}{2}]$ and 1 on $(\frac{1}{2}, 1]$. If $\{f_n\}$ had a limit in $C[0,1]$ then it would have to be f since $(C[0,1], \|\cdot\|_1)$ is naturally a subspace of $(PC[0,1], \|\cdot\|_1)$, but f is not continuous. So $C[0,1]$ is not complete.

Unfortunately there are examples which show that $PC[0,1]$ is not complete either. The completion of $(C[0,1], \|\cdot\|_1)$ is a rather exotic object which cannot readily be identified with a space of functions. Rather, it is convenient to think of its completion as a set of equivalence classes of potentially exotic functions. We shall investigate this space further in future lectures.

6. Lecture 6 (9/2/11): Euclidean Spaces

So far we have considered vector spaces equipped with a compatible notion of distance, i.e. a norm. We have seen that this structure is enough to build machinery relevant to topology - such as convergence, continuity, and compactness - but it is not enough to capture classical Euclidean geometry. Specifically, we are missing the notion of angle. By the cosine law for triangles one can compute angles using only distances, but as we shall see it is not the case that every notion of distance will give rise to a sensible notion of angle.

The goal of this section is to investigate the algebraic structure needed to discuss angles. We shall restrict our attention to $\mathbb{R}^n$, but the tools we develop will be very important in the infinite dimensional setting and thus we shall try to give proofs which generalize to infinite dimensions.

Definition 6.1. An *inner product* on $\mathbb{R}^n$ is a function $\langle\cdot\rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ which satisfies the following axioms:

- Bilinearity:

$$\langle a\mathbf{x} + a'\mathbf{x}', \mathbf{y}\rangle = a\langle \mathbf{x}, \mathbf{y}\rangle + a'\langle \mathbf{x}', \mathbf{y}\rangle$$
$$\langle \mathbf{x}, b\mathbf{y} + b'\mathbf{y}'\rangle = b\langle \mathbf{x}, \mathbf{y}\rangle + b'\langle \mathbf{x}, \mathbf{y}'\rangle$$

- Symmetry: $\langle \mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{y}, \mathbf{x}\rangle$
- Positive Definiteness: $\sqrt{\langle \mathbf{x}, \mathbf{x}\rangle}$ is nonnegative for every $\mathbf{x}$, and it is 0 if and only if $\mathbf{x} = \mathbf{0}$

A *Euclidean space* is $\mathbb{R}^n$ equipped with an inner product.

Example 6.2. Define $\langle \mathbf{x}, \mathbf{y}\rangle = \sum_{i=1}^n x_i y_i$ where $\mathbf{x} = \{x_i\}$ and $\mathbf{y} = \{y_i\}$. This is an example of an inner product on $\mathbb{R}^n$, often called the *standard inner product* or simply the *dot product*.

Exercise 6.3. Show that if $\langle\cdot,\cdot\rangle$ is an inner product then the function $\mathbf{x} \mapsto \sqrt{\langle \mathbf{x}, \mathbf{x}\rangle}$ defines a norm on $\mathbb{R}^n$. This is referred to as the norm *associated to* or *induced by* the inner product. Show that $\|\cdot\|_2$ is the norm associated to the standard inner product.

The previous exercise asserts that every inner product gives rise to a norm. In fact, an inner product can be recovered from the norm that it induces via the so-called *polarization identity*:

Lemma 6.4. *Let $\langle\cdot,\cdot\rangle$ be an inner product and let $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x}\rangle}$ be its associated norm. Then $\langle \mathbf{x}, \mathbf{y}\rangle = \frac{1}{2}(\|\mathbf{x}+\mathbf{y}\|^2 - \|\mathbf{x}\|^2 - \|\mathbf{y}\|^2)$.*

Proof. Calculate:

$$\begin{aligned}\|\mathbf{x}+\mathbf{y}\|^2 &= \langle \mathbf{x}+\mathbf{y}, \mathbf{x}+\mathbf{y}\rangle \\ &= \langle \mathbf{x},\mathbf{x}\rangle + 2\langle \mathbf{x},\mathbf{y}\rangle + \langle \mathbf{y},\mathbf{y}\rangle \\ &= \|\mathbf{x}\|^2 + 2\langle \mathbf{x},\mathbf{y}\rangle + \|\mathbf{y}\|^2\end{aligned}$$

The result follows by solving for $\langle \mathbf{x},\mathbf{y}\rangle$ □

Lemma 6.4 yields a necessary and sufficient condition for a norm $\|\cdot\|$ to be induced by an inner product, namely that $\frac{1}{2}(\|\mathbf{x}+\mathbf{y}\|^2 - \|\mathbf{x}\|^2 - \|\mathbf{y}\|^2)$ is bilinear in $\mathbf{x}$ and $\mathbf{y}$. The following exercise gives a more straightforward condition:

Exercise 6.5. Prove that a norm $\|\cdot\|$ is induced by an inner product if and only if it satisfies the *parallelogram law*: $\|\mathbf{x}+\mathbf{y}\|^2 + \|\mathbf{x}-\mathbf{y}\|^2 = 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2$. Deduce that the ℓ^p norm on $\mathbb{R}^n$ is induced by an inner product if and only if $p = 2$.

Exercise 6.6. Prove that if $\mathbf{x}$ and $\mathbf{y}$ are vectors in $\mathbb{R}^n$ then the angle α between them satisfies $\cos\alpha = \frac{\langle \mathbf{x},\mathbf{y}\rangle}{\|x\|\|y\|}$ where $\langle\cdot,\cdot\rangle$ is the standard inner product and $\|\cdot\|$ is its associated norm.

We can take the formula appearing in Exercise 6.6 as a definition of angle in any Euclidean space. In practice we will not need to be able to calculate the angles between arbitrary pairs of vectors; it is primarily important to be able to determine when two vectors are perpendicular.

Definition 6.7. Let $(\mathbb{R}^n, \langle\cdot,\cdot\rangle)$ be a Euclidean space. Two vectors $\mathbf{x}$ and $\mathbf{y}$ are said to be *orthogonal* if $\langle \mathbf{x},\mathbf{y}\rangle = 0$.

Exercise 6.8. Prove that $\mathbf{x}$ and $\mathbf{y}$ are orthogonal if and only if $\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$. This is a generalization of the Pythagorean theorem.

7. Lecture 7 (9/7/11): Euclidean Spaces, Continued

In this lecture we will discuss the problem of classifying inner products on $\mathbb{R}^n$. We begin by proving that for every inner product $\langle\cdot,\cdot\rangle$ there is a coordinate system on $\mathbb{R}^n$ for which $\langle\cdot,\cdot\rangle$ is simply the standard inner product of Example 6.2. This requires us to introduce a definition which is crucial for working with inner product spaces.

Definition 7.1. An *orthogonal basis* for a Euclidean space $(\mathbb{R}^n, \langle\cdot,\cdot\rangle)$ is a basis $\{\mathbf{u}_1, \dots \mathbf{u}_n\}$ for $\mathbb{R}^n$ with the property that $\langle \mathbf{u}_i, \mathbf{u}_j\rangle = 0$ if $i \neq j$. An orthogonal basis is *orthonormal* if additionally $\langle \mathbf{u}_i, \mathbf{u}_i\rangle = 1$.

Example 7.2. The standard basis for $\mathbb{R}^n$ is an orthonormal basis for $\mathbb{R}^n$ equipped with the standard inner product.

We will now show that orthonormal bases always exist.

Proposition 7.3. *Every Euclidean space* $(\mathbb{R}^n, \langle\cdot,\cdot\rangle)$ *has an orthonormal basis.*

Proof. We will construct an orthonormal basis by appealing to the principle of recursive definition. Let $\mathbf{u}_1$ be any vector of norm 1. Assume that we have constructed an orthonormal set $S_k = \{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ where $1 \leq k < n$; we must construct a unit vector $\mathbf{u}_{k+1}$ which is orthogonal to each $\mathbf{u}_i$ in S_k. Since S_k has fewer than n vectors span$\{S_k\}$ cannot be all of $\mathbb{R}^n$, so take $\mathbf{x}$ not in span$\{S_k\}$. Define a_i to be the number $\frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle}$ for $1 \leq i \leq k$ and let $\mathbf{v}_{k+1} = \mathbf{x} - \sum_i a_i \mathbf{u}_i$. Note that $\mathbf{v}_{k+1} \neq 0$ since $\mathbf{x}$ is not in the span of the $\mathbf{u}_i$'s. I claim that $\mathbf{u}_j$ is orthogonal to $\mathbf{v}_{k+1}$ for $1 \leq j \leq k$. Indeed, since $\langle \mathbf{u}_i, \mathbf{u}_j \rangle$ is 1 when $i = j$ and 0 otherwise, we have:

$$\langle \mathbf{v}_{k+1}, \mathbf{u}_j \rangle = \langle \mathbf{x}, \mathbf{u}_j \rangle - \frac{\langle \mathbf{x}, \mathbf{u}_j \rangle}{\langle \mathbf{u}_j, \mathbf{u}_j \rangle} \langle \mathbf{u}_j, \mathbf{u}_j \rangle = 0$$

It follows that the vector $\mathbf{u}_{k+1} := \frac{\mathbf{v}_{k+1}}{\|\mathbf{v}_{k+1}\|}$ is a unit vector which is orthogonal to each other $\mathbf{u}_i$, as desired. □

This proof used an important geometric device which appears frequently in the theory of inner product spaces. The vector $\sum_i \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i$ is called the *orthogonal projection* of $\mathbf{x}$ onto the subspace $\operatorname{span}\{S_k\}$; one can show that among all vectors in $\operatorname{span}\{S_k\}$ this is the unique vector which minimizes the distance to $\mathbf{x}$. The proof illustrates the general principle that if $\mathbf{x}$ is a vector in a Euclidean space and V is a subspace then the difference between $\mathbf{x}$ and its orthogonal projection onto V is orthogonal to every vector in V.

We can interpret the existence of orthonormal bases as a statement about the structure of an inner product. Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be an orthonormal basis for $(\mathbb{R}^n, \langle \cdot, \cdot \rangle)$ and write $\mathbf{x} = \sum_i x_i \mathbf{u}_i$, $\mathbf{y} = \sum_j y_j \mathbf{u}_j$. Using bilinearity, we get

$$\begin{aligned} \langle \mathbf{x}, \mathbf{y} \rangle &= \left\langle \sum_i x_i \mathbf{u}_i, \sum_j y_j \mathbf{u}_j \right\rangle \\ &= \sum_{i,j} x_i y_j \langle \mathbf{u}_i, \mathbf{u}_j \rangle \\ &= \sum_i x_i y_i \end{aligned}$$

So any inner product takes the form of the standard inner product relative to an orthonormal basis. In other words, the standard inner product is the only inner product up to linear changes of coordinates.

This already goes a long way toward classifying inner products, but a stronger statement is possible. Indeed, it is possible to classify inner products up to *orthogonal* changes of coordinates, i.e. isometries of $\mathbb{R}^n$. The precise statement is as follows.

Theorem 7.4. *Let $\langle \cdot, \cdot \rangle$ and $\langle \cdot, \cdot \rangle'$ be two inner products on $\mathbb{R}^n$. Then there is an orthonormal basis for $\langle \cdot, \cdot \rangle$ which is an orthogonal basis for $\langle \cdot, \cdot \rangle'$.*

Thus, up to isometry, every inner product is equivalent to one of the form $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_i x_i y_i c_i$ where $c_1, \ldots, c_n$ are positive constants (by positive definiteness). The proof of this theorem uses important facts about the structure theory of linear transformations on $\mathbb{R}^n$, and we will deferr it until future lectures.

As we discussed above there are many norms which are not induced by inner products, and we will be unable to classify norms on $\mathbb{R}^n$ up to isometry the way we could classify inner products. However, for many purposes a weaker notion of equivalence is good enough.

Definition 7.5. Let V be a vector space. Two norms $\|\cdot\|$ and $\|\cdot\|'$ on V are said to be equivalent if there is a constant $C \geq 1$ such that $\frac{1}{C} \|\cdot\|' \leq \|\cdot\| \leq C \|\cdot\|'$.

Exercise 7.6. Show that norm equivalence is an equivalence relation on the set of all norms on a fixed vector space V.

Exercise 7.7. Show that the ℓ^1 norm is equivalent to the ℓ^2 norm.

As we shall explain, many useful statements about a normed space remain true of the norm is replaced by an equivalent one. So it is convenient that there is only one norm on $\mathbb{R}^n$ up to equivalence, as we shall now prove.

Theorem 7.8. *All norms on $\mathbb{R}^n$ are equivalent.*

Proof. Let $\|\cdot\|$ be any norm on $\mathbb{R}^n$; it suffices to show that $\|\cdot\|$ is equivalent to the ℓ^2 norm. Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be the standard basis on $\mathbb{R}^n$ and let M denote the largest value of $\|\mathbf{e}_i\|$ for $1 \leq i \leq n$. For any $\mathbf{x} \in \mathbb{R}^n$, we have:

$$\begin{aligned} \|\mathbf{x}\| &= \left\| \sum_i x_i \mathbf{e}_i \right\| \\ &\leq \sum_i |x_i| \, \|\mathbf{e}_i\| \\ &\leq M \sum_i |x_i| \\ &= M \, \|\mathbf{x}\|_1 \end{aligned}$$

By Exercise 7.7, we have that $\|\mathbf{x}\| \leq M' \, \|\mathbf{x}\|_2$ for some constant M'. It follows that $\|\cdot\|$ is a continuous function on $\mathbb{R}^n$ (equipped with the standard metric). Let S^{n-1} denote the unit sphere in $\mathbb{R}^n$, i.e. the set of all $\mathbf{x} \in \mathbb{R}^n$ with $\|\mathbf{x}\|_2 = 1$. S^{n-1} is a compact subset of $\mathbb{R}^n$, so $\|\cdot\|$ attains a minimum value m on S^{n-1}. Thus for any $\mathbf{x} \in \mathbb{R}^n$ we have $\left\| \frac{\mathbf{x}}{\|\mathbf{x}\|_2} \right\| \geq m$ and thus $\|\mathbf{x}\| \geq m \, \|\mathbf{x}\|_2$. Taking C to be any number bigger than $\frac{1}{m}$, M', and 1, we conclude that $\frac{1}{C} \|\mathbf{x}\|_2 \leq \|\mathbf{x}\| \leq C \, \|\mathbf{x}\|_2$. Thus $\|\cdot\|$ is equivalent to $\|\cdot\|_2$, as desired. $\square$

Corollary 7.9. *If $\|\cdot\|$ and $\|\cdot\|'$ are norms on $\mathbb{R}^n$ then a sequence converges with respect to $\|\cdot\|$ if and only if it converges with respect to $\|\cdot\|'$.*

Proof. By Theorem 7.8 there is a constant $C \geq 1$ such that $\frac{1}{C} \|\cdot\|' \leq \|\cdot\| \leq C \, \|\cdot\|$. Let $\{\mathbf{x}_n\}$ be a sequence in $\mathbb{R}^n$ which converges to $\mathbf{x}$ in the norm $\|\cdot\|$. Given any $\varepsilon > 0$ there exists N such that $\|\mathbf{x} - \mathbf{x}_n\| < \frac{\varepsilon}{C}$ whenever $n \geq N$. Thus for all such n we have $\|\mathbf{x} - \mathbf{x}_n\|' \leq C \, \|\mathbf{x} - \mathbf{x}_n\| < \varepsilon$, and hence $\{\mathbf{x}_n\}$ converges to $\mathbf{x}$ in the norm $\|\cdot\|'$. The converse follows from a similar argument using the inequality $\|\cdot\| \leq C \, \|\cdot\|'$. $\square$

A similar argument shows that a sequence is Cauchy in $\|\cdot\|$ if and only if it is Cauchy in $\|\cdot\|'$. It immediately follows that $\mathbb{R}^n$ is complete with respect to any norm, and that a function on $\mathbb{R}^n$ is continuous with respect to one norm if and only if it is continuous with respect to every other norm. Indeed, there is only one norm topology on $\mathbb{R}^n$. It is useful to contrast $\mathbb{R}^n$ with $\mathbb{Q}$, which has infinitely many inequivalent p-adic norms, and with $C[0,1]$, which has infinitely many inequivalent ℓ^p norms.

8. Lecture 8 (9/9/11): The Geometry of Finite Dimensional Banach Spaces

We have just seen that inner products on R^n admit a complete algebraic classification, and we remarked that it is very difficult to arrive at a similar classification of norms on $\mathbb{R}^n$. However, we can get some idea of how large the universe of norms is by characterizing them geometrically. Specifically, we will provide necessary and sufficient conditions for a subset B of $\mathbb{R}^n$ to be the closed unit ball for some norm.

Two necessary conditions are immediately apparent: $\mathbf{0} \in B$ since $\mathbf{0}$ always has norm 0, and B is centrally symmetric (meaning $-\mathbf{x} \in B$ whenever $\mathbf{x} \in B$) since every norm is homogeneous. Additionally, Theorem 7.8 implies that B is compact:

Lemma 8.1. *Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$ and let $B = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| \leq 1\}$ be the closed unit ball for this norm. Then B is compact.*

Proof. First, we show that B is compact relative to the standard Euclidean norm $\|\cdot\|_2$. By the Heine-Borel theorem, it is necessary and sufficient to show that B is a closed and bounded subset of $(\mathbb{R}^n, \|\cdot\|_2)$. Appealing to Theorem 7.8, there is a constant $C \geq 1$ such that $\frac{1}{C}\|\cdot\| \leq \|\cdot\|_2 \leq C\|\cdot\|$. Thus $\|\mathbf{x}\|_2 \leq C$ for every $\mathbf{x} \in B$, and hence B is bounded relative to $\|\cdot\|_2$. To prove that B is closed, let $\{\mathbf{x}_n\}$ be a sequence in B which converges to $\mathbf{x} \in \mathbb{R}^n$ relative to $\|\cdot\|_2$. By Corollary 7.9 we have that $\{\mathbf{x}_n\}$ converges to $\mathbf{x}$ relative to $\|\cdot\|$ as well. Since $\|\mathbf{x}\| \leq \|\mathbf{x} - \mathbf{x}_n\| + \|\mathbf{x}_n\|$ for every n, it follows that $\|\mathbf{x}\| < \varepsilon + 1$ for every $\varepsilon > 0$. Thus $\|\mathbf{x}\| \leq 1$ and we conclude that B is closed relative to $\|\cdot\|_2$. This completes the proof that B is compact relative to $\|\cdot\|_2$.

To prove that B is compact relative to $\|\cdot\|$, we use the sequential formulation of compactness. In other words we show that every sequence in B has a $\|\cdot\|$-convergent subsequence. Every sequence in B has a $\|\cdot\|_2$-convergent subsequence since B is compact relative to $\|\cdot\|_2$, and that subsequence converges relative to $\|\cdot\|$ by Corollary 7.9. This completes the proof. □

Remark 8.2. The reader may find it strange that the standard Euclidean norm plays such a prominent role in the proof of this lemma. What the proof really shows is that the unit ball in a normed space is compact relative to one norm if and only if it is compact relative to any equivalent norm. Thus the lemma is equivalent to the statement that the unit ball is compact for some norm on $\mathbb{R}^n$, and this is well-known for the standard Euclidean norm. We already saw the strategy of reducing a statement about a general norm to a statement about the standard Euclidean norm in the proof of Theorem 7.8, and it will continue to be useful.

So far we have not used the convexity of $\|\cdot\|$ to place any constraints on B. In fact it imposes a very strong geometric constraint which we now discuss.

Definition 8.3. A subset $S \subseteq \mathbb{R}^n$ is said to be convex if for every $\mathbf{x}, \mathbf{y} \in S$ and every $\alpha \in [0, 1]$ we have $\alpha\mathbf{x} + (1 - \alpha)\mathbf{y} \in S$.

Exercise 8.4. Show that the intersection of an arbitrary family of convex sets is convex.

In other words, a set is convex if it contains the line segment joining any two of its points. The following lemma gives a useful criterion for determining whether or not a set is convex.

Lemma 8.5. *A closed set $S \subseteq \mathbb{R}^n$ is convex if and only if $\frac{\mathbf{x}+\mathbf{y}}{2} \in S$ whenever $\mathbf{x}$ and $\mathbf{y}$ are in S.*

Proof. If S is convex and $\mathbf{x}$ and $\mathbf{y}$ are in S then we can set $\alpha = \frac{1}{2}$ in the definition of convexity to deduce that $\frac{1}{2}\mathbf{x} + \frac{1}{2}\mathbf{y} \in S$. Conversely, the condition that $\frac{\mathbf{x}+\mathbf{y}}{2} \in S$ whenever $\mathbf{x}$ and $\mathbf{y}$ are in S together with induction guarantees that $q\mathbf{x}+(1-q)\mathbf{y} \in S$ for every diadic rational number q in $[0,1]$ (i.e. every number in $[0,1]$ of the form $\frac{p}{2^n}$ where n and p are integers). The set of all diadic rational numbers in $[0,1]$ is dense, so $\alpha\mathbf{x} + (1-\alpha)\mathbf{y} \in S$ for every $\alpha \in [0,1]$. ☐

Remark 8.6. The subsete $\mathbb{Q} \subseteq \mathbb{R}$ has the property that $\frac{\mathbf{x}+\mathbf{y}}{2} \in \mathbb{Q}$ whenever $\mathbf{x}$ and $\mathbf{y}$ are in $\mathbb{Q}$, but $\mathbb{Q}$ is not convex. So it is crucial that S in the statement of the lemma be closed.

Exercise 8.7. Show that if S is convex and $\mathbf{x}_1, \ldots, \mathbf{x}_k \in S$ then $\sum_{i=1}^k \alpha_i \mathbf{x}_i \in S$ whenever $\alpha_i \in [0,1]$ and $\sum_{i=1}^k \alpha_i = 1$. This is called a *convex combination* of the $\mathbf{x}_i$'s.

It follows immediately from the convexity of $\|\cdot\|$ that the closed unit ball B is convex. Indeed, if $\mathbf{x}$ and $\mathbf{y}$ are in B and $\alpha \in [0,1]$ then $\|\alpha\mathbf{x} + (1-\alpha)\mathbf{y}\| \leq \alpha\|\mathbf{x}\| + (1-\alpha)\|\mathbf{y}\| \leq \alpha + 1 - \alpha = 1$.

So B must contain $\mathbf{0}$ and it must be compact, convex, and centrally symmetric. However, there are sets which meet these four requirements but cannot be the unit ball for any norm; take the line segment $\{(t,0) : t \in [-1,1]\}$, for example. Indeed, this example can be eliminated via the following result.

Lemma 8.8. *If $\|\cdot\|$ is a norm on $\mathbb{R}^n$ then its unit ball B contains a basis for $\mathbb{R}^n$.*

Proof. Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_k\}$ be a linearly independent subset of B of maximal cardinality, and assume for contradiction that $k < n$. Then there is a vector $\mathbf{v}_{k+1} \in \mathbb{R}^n$ with the property that $\{\mathbf{e}_1, \ldots, \mathbf{e}_k, \mathbf{v}_{k+1}\}$ is a linearly independent set. Define $\mathbf{e}_{k+1} = \frac{\mathbf{v}_{k+1}}{\|\mathbf{v}_{k+1}\|}$, a vector in B. Observe that $\{\mathbf{e}_1, \ldots, \mathbf{e}_{k+1}\}$ is a linearly independent set in B of cardinality $k+1$, contradicting the maximality of k. ☐

We have now listed enough necessary conditions to completely characterize the subsets of $\mathbb{R}^n$ which can be realized as the unit ball for some norm. Before proving this, we need the following technical lemma:

Lemma 8.9. *Let $B \subseteq \mathbb{R}^n$ be a convex centrally symmetric set which contains a basis for $\mathbb{R}^n$. Then B contains an open Euclidean ball centered at the origin.*

Proof. Assume B contains the basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ for $\mathbb{R}^n$. Given any nonzero vector $\mathbf{x} \in \mathbb{R}^n$ we may write $\mathbf{x} = \sum_i a_i \mathbf{e}_i$ where the a_i's are real numbers not all zero. This can be rewritten uniquely as $\sum_i |a_i| (\pm\mathbf{e}_i)$; set $a = \sum_i |a_i|$ and observe that $\frac{1}{a}\mathbf{x}$ is a convex combination of vectors in the set $\{\pm\mathbf{e}_1, \ldots, \pm\mathbf{e}_n\}$. These vectors are all in B since B is centrally symmetric, so $\frac{1}{a}\mathbf{x} \in B$ since B is convex.

Let S^{n-1} be the Euclidean unit sphere and define a function $\phi : S^{n-1} \to \mathbb{R}$ by the formula $\phi(\sum_i a_i \mathbf{e}_i) = \sum_i |a_i|$. Thus $\frac{1}{\phi(\mathbf{x})}\mathbf{x}$ is in B whenever $x \in S^{n-1}$. ϕ is continuous and S^{n-1} is compact, so ϕ is bounded by some constant M. Now, since $\mathbf{0} = \frac{1}{2}\mathbf{e}_1 + \frac{1}{2}(-\mathbf{e}_1)$ is in B, we have that $\lambda\mathbf{x} \in B$ whenever $\mathbf{x} \in B$ and $0 \leq \lambda \leq 1$. Thus for any $\lambda \leq \frac{1}{M}$ and any $\mathbf{x} \in S^{n-1}$ we have that $\lambda\mathbf{x} \in B$ since $\lambda \leq \frac{1}{\phi(\mathbf{x})}$ and $\frac{1}{\phi(\mathbf{x})} \in B$. Therefore the open Euclidean ball of radius $\frac{1}{M}$ centered at the origin is contained in B. ☐

Theorem 8.10. *Let $B \subseteq \mathbb{R}^n$ be a convex, compact, centrally symmetric set which contains a basis for $\mathbb{R}^n$. Then there exists a norm $\|\cdot\|$ on $\mathbb{R}^n$ with the property that $B = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| \leq 1\}$.*

Proof. For each $x \in \mathbb{R}^n$, let $S(\mathbf{x}) = \{\alpha\mathbf{x} : 0 \leq \alpha < \infty\}$; observe that $S(x) \cap B$ is a compact convex subset of $S(\mathbf{x})$ since the intersection of two convex sets is convex and the intersection of a closed set with a compact set is compact. $S(\mathbf{x}) \cap B$ contains $\mathbf{0}$, and the only compact convex subsets of $S(\mathbf{x})$ which contain $\mathbf{0}$ are the sets of the form $\{\alpha\mathbf{x} : 0 \leq \alpha \leq m\}$. So define a function $m : \mathbb{R}^n \to \mathbb{R}$ by $m(\mathbf{x}) = \sup\{\alpha : \alpha\mathbf{x} \in S(\mathbf{x}) \cap B\}$. Note that m takes strictly positive values by Lemma 8.9. Thus we may define $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ by $\|\mathbf{x}\| = \frac{1}{m(\mathbf{x})}$, with the convention $\frac{1}{\infty} = 0$. We shall prove that $\|\cdot\|$ is a norm on $\mathbb{R}^n$ whose unit ball is B.

$\|\cdot\|$ is clearly nonnegative, and $\|\mathbf{0}\| = 0$ since $m(\mathbf{0}) = \infty$. B is compact and therefore bounded, so $m(\mathbf{x}) = \infty$ only if $\mathbf{x} = 0$. Thus $\|\cdot\|$ is positive definite.

For $\lambda > 0$ we have

$$\sup\{\alpha : \alpha\lambda\mathbf{x} \in S(\mathbf{x}) \cap B\} = \sup\{\frac{1}{\lambda}\alpha : \alpha\mathbf{x} \in S(\mathbf{x}) \cap B\}$$

Thus $m(\lambda\mathbf{x}) = \frac{1}{\lambda}m(\mathbf{x})$. $m(-\mathbf{x}) = m(\mathbf{x})$ since B is centrally symmetric, so for $\lambda < 0$ we have $m(\lambda\mathbf{x}) = -\frac{1}{\lambda}m(\mathbf{x})$. This proves homogeneity.

To prove convexity of $\|\cdot\|$, we must show that $m(\mathbf{x}+\mathbf{y}) \geq \frac{m(\mathbf{x})m(\mathbf{y})}{m(\mathbf{x})+m(\mathbf{y})}$ for every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Suppose that α and β are positive numbers with the property that $\alpha\mathbf{x} \in B$ and $\beta\mathbf{y} \in B$. Setting $\lambda = \frac{\beta}{\alpha+\beta}$, note that $\lambda\alpha\mathbf{x} + (1-\lambda)\beta\mathbf{y} \in B$ since B is convex. This gives that $\frac{\alpha\beta}{\alpha+\beta}(\mathbf{x}+\mathbf{y}) \in B$, and hence $m(\mathbf{x}+\mathbf{y}) \geq \frac{\alpha\beta}{\alpha+\beta}$ for every pair α, β such that $\alpha\mathbf{x} \in B$ and $\beta\mathbf{y} \in B$. Taking the supremum over all such α and β, the desired inequality follows.

We have now proven that $\|\cdot\|$ is a norm; it remains only to show that B is its unit ball. We have that $\mathbf{x}$ is in B if and only if $1 \in S(\mathbf{x})$, and this is equivalent to the statement that $m(\mathbf{x}) \geq 1$, i.e. $\|\mathbf{x}\| \leq 1$. This completes the proof. □

Remark 8.11. The function $\|\cdot\|$ defined in the proof of this theorem makes sense for more general sets B than those which satisfy the hypotheses above, and it is often called the *Minkowski functional.* However, the Minkowski functional is only a norm in the situation of the theorem.

9. Lecture 9 (9/12/11)

9.1. **Duality for Finite Dimensional Banach Spaces.** One of the most important concepts in functional analysis - and mathematics in general - is the concept of duality. Duality is based on the general principle that an algebraic object X can be investigated by providing a space X^* of functions on X with a compatible structure. In a wide variety of circumstances, X can be recovered from X^* by embedding it in X^{**}, the dual of its dual.

One of the most elementary examples of this principle occurs in the theory of Banach spaces. Duality for Banach spaces is most interesting and useful in the infinite dimensional case, but we will begin by focusing on the finite dimensional case to suppress some subtleties.

Definition 9.1. Let $(V, \|\cdot\|)$ be a normed space over $\mathbb{R}$ (of any dimension).

- A *linear functional* on V is a linear map $\ell : V \to \mathbb{R}$.

- The *dual space* of V is the set V^* of all linear functionals on V which are continuous relative to $\|\cdot\|$.

Let us now restrict to the finite dimensional case. Thus we may assume that $V = \mathbb{R}^n$ equipped with some norm $\|\cdot\|$. Any linear functional $\ell \in \mathbb{R}^n$ can be expressed as $\ell(\mathbf{x}) = \sum_{i=1}^n a_i x_i$ in standard coordinates, and thus ℓ is obviously continuous with respect to the standard norm on $\mathbb{R}^n$. But recall that continuity for functions on $\mathbb{R}^n$ is independent of the norm chosen by Theorem 7.8, so we see that every linear functional on a finite dimensional normed space is automatically continuous.

This characterization of linear functionals on $\mathbb{R}^n$ also yields an identification $(\mathbb{R}^n)^* \to \mathbb{R}^n$ given by $\ell \mapsto (a_1, \ldots, a_n)$. Note, however, that this identification depends on our initial choice of coordinates on $\mathbb{R}^n$, and thus it is not *canonical*.

Remark 9.2. It is known to be impossible to explicitly construct a discontinuous linear functional on an infinite dimensional Banach space even though such functionals exist. This is our first example of many to come of infinite dimensional construction which depends on the "axiom of choice" in set theory. This axiom is independent of the other axioms of set theory which lie at the foundation of modern mathematics, meaning one can choose to adopt or not adopt the axiom without introducing new contradictions into mathematics. It is standard practice in functional analysis to freely employ the axiom of choice.

While we cannot canonically identify V and V^* when V is finite dimensional (even though they are isomorphic), it turns out that there is a canonical isomorphism $V \xrightarrow{\cong} V^{**}$ where V^{**} is the dual space of V^*. The careful reader may notice that V^* does not a priori have a dual in the sense of Definition 9.1 because we have not equipped V^* with a norm, but if V isfinite dimensional then V^{**} is well-defined independently of the norm chosen on V^* since V^* is also finite dimensional. For the infinite dimensional case we will develop a natural way to equip the dual of a normed vector space with a norm of its own.

We now define a map $i : V \to V^{**}$ which will turn out to be an isomorphism if V is finite dimensional. A map from V to V^{**} assigns to each $\mathbf{x} \in V$ a linar functional on V^*; in other words it is family of linear maps $V^* \to \mathbb{R}$ - one map for each $\mathbf{x} \in V$. There is a particularly natural way to define such a family.

Definition 9.3. Define the *canonical embedding* of V into $V^*\ast$ to be the map $i : V \to V^{**}$ which sends $\mathbf{x} \in V$ to the linear functional $i_{\mathbf{x}} : V^* \to \mathbb{R}$ given by $i_{\mathbf{x}}(\ell) = \ell(\mathbf{x})$.

Note that the definition of i did not require any auxiliary choices to be made - it uses only the structure of V, V^*, and V^{**}. This is the sense in which it is canonical. It turns out to be injective (although this is not obvious and the proof in the general case requires the axiom of choice!) but it is not in general an isomorphism. However, it is an isomorphism in the finite dimensional case.

Lemma 9.4. *The canonical embedding is a linear map, and it is an isomorphism if V is finite dimensional.*

Proof. First we show that it is linear. Let $\mathbf{x}, \mathbf{y} \in V$ and $a, b \in \mathbb{R}$, and take any $\ell \in V^*$.

$$\begin{aligned} i_{a\mathbf{x}+b\mathbf{y}}(\ell) &= \ell(a\mathbf{x} + b\mathbf{y}) \\ &= a\ell(\mathbf{x}) + b\ell(\mathbf{y}) \\ &= a i_{\mathbf{x}}(\ell) + b i_{\mathbf{y}}(\ell) \\ &= (a i_{\mathbf{x}} + b i_{\mathbf{y}})(\ell) \end{aligned}$$

Thus $i_{a\mathbf{x}+b\mathbf{y}} = a i_{\mathbf{x}} + b i_{\mathbf{y}}$ as linear functionals on V^*. This means $i : V \to V^*$ is linear.

Now we show that it is injective. Suppose that $i_{\mathbf{x}} = 0$ as a linear functional on V^*. This means that $i_{\mathbf{x}}(\ell) = 0$ for every $\ell \in V^*$, so that $\ell(\mathbf{x}) = 0$. We want to show that $\mathbf{x} = 0$, so it suffices to show that for every nonzero $\mathbf{x} \in V$ there exists $\ell \in V^*$ such that $\ell(\mathbf{x}) \neq 0$. Given $\mathbf{x} \neq 0$, choose vectors $\mathbf{e}_2, \dots, \mathbf{e}_n$ so that $\{\mathbf{x}, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ is a basis for V. Any function defined on a basis for V can be extended to a linear functional on all of V, so there is a linear functional $\ell \in V^*$ which satisfies $\ell(\mathbf{x}) = 1$ and $\ell(\mathbf{e}_i) = 0$ (the values of ℓ on the $\mathbf{e}_i$'s are unimportant). This shows that $\mathbf{x} = 0$ if $\ell(\mathbf{x}) = 0$ for every $\ell \in V^*$, and hence that i is injective. Since V and $V^{*}*$ have the same dimension, we conclude that i is an isomorphism. □

10. Lecture 10 (9/14/11): The Dual Norm

In the last section we defined the dual of a normed vector space V as the space of all continuous linear functionals on V and proved some basic statements about dual spaces in the finite dimensional case. We remarked that the dual of a finite dimensional normed space is independent of the norm, but that this is not the case in infinite dimensions. In this section we will equip the dual of a normed space with a norm of its own and prove some results that hold for duals of normed spaces of arbitrary dimension. Our first step is to give an alternative characterization of continuity for linear functionals.

Definition 10.1. Let $(V, \|\cdot\|)$ be a normed space. A linear functional $\ell : V \to \mathbb{R}$ is *bounded* if $\sup_{\|\mathbf{x}\|=1} |\ell(\mathbf{x})| < \infty$.

Lemma 10.2. *A linear functional on a normed space is continuous if and only if it is bounded.*

Proof. Suppose $\ell : V \to \mathbb{R}$ is continuous. By the definition of continuity there exists $\delta > 0$ such that $|\ell(\mathbf{x})| < 1$ if $\|\mathbf{x}\| = \delta$. Thus if $\|\mathbf{x}\| = 1$ then $|\ell(\delta\mathbf{x})| < 1$, so that $|\ell(\mathbf{x})| < \frac{1}{\delta}$. So ℓ is bounded.

Conversely, suppose ℓ is bounded and let $M = \sup_{\|\mathbf{x}\|=1} |\ell(\mathbf{x})|$. Given $\varepsilon > 0$, set $\delta = \frac{\varepsilon}{M}$ and assume $\mathbf{x}$ is such that $\|\mathbf{x}\| = N < \delta$. Then $\left|\ell(\frac{\mathbf{x}}{N})\right| \leq M$, so that $|\ell(\mathbf{x})| \leq MN < M\delta = \varepsilon$. Thus ℓ is continuous at 0. The translation maps $\mathbf{x} \mapsto \mathbf{x} + \mathbf{x}_0$ are homeomorphisms, so ℓ is continuous at every point of V. □

This lemma allows us to define a norm on V^*.

Definition 10.3. Let $(V, \|\cdot\|)$ be a normed space. The *dual norm* on V^* is defined to be $\|\ell\|^* = \sup_{\|\mathbf{x}\|=1} |\ell(\mathbf{x})|$.

Lemma 10.4. $\|\cdot\|^*$ *is a norm on* V^*.

Proof. It is clear that $\|0\|^* = 0$. If $\|\ell\|^* = 0$ then for every $\mathbf{x} \in V$ we have $\ell(\frac{\mathbf{x}}{\|\mathbf{x}\|}) = 0$ and thus $\ell(\mathbf{x}) = 0$. So $\ell = 0$, and we conclude that $\|\cdot\|^*$ is positive definite.

Take $\ell \in V^*$ and $\lambda \in \mathbb{R}$. We have $\sup_{\|\mathbf{x}\|=1} |\lambda\ell(\mathbf{x})| = \lambda \sup_{\|\mathbf{x}\|=1} |\ell(\mathbf{x})|$, so $\|\lambda\ell\|^* = |\lambda| \|\ell\|^*$. Thus $\|\cdot\|^*$ is homogeneous.

Finally, $\sup_{\|\mathbf{x}\|=1} |\ell(\mathbf{x}) + \ell'(\mathbf{x})| \leq \sup_{\|\mathbf{x}\|=1} |\ell(\mathbf{x})| + |\ell'(\mathbf{x})| \leq \sup_{\|\mathbf{x}\|=1} |\ell(\mathbf{x})| + \sup_{\|\mathbf{y}\|=1} |\ell(\mathbf{y})|$. So $\|\cdot\|^*$ is convex. This completes the proof. □

Example 10.5. Let us explicitly calculate the dual norm in the case where $V = \mathbb{R}^n$ equipped with the standard Euclidean norm $\|\cdot\|_2$. Recall that every linear functional $\ell \in (\mathbb{R}^n)^*$ has the form $\ell(\mathbf{x}) = \sum_{i=1}^n a_i x_i$ in standard coordinates where $a_1, \ldots, a_i$ are constants determined by ℓ. If $\mathbf{x}$ has norm 1 then by the Cauchy-Schwarz inequality we have $|\ell(\mathbf{x})| \leq \sum_i |a_i| |x_i| \leq \|\mathbf{x}\|_2 (\sum_i |a_i|^2)^{1/2} = \|(a_1 \ldots a_n)\|_2$. Moreover $\ell(a_1 \ldots a_n) = \|(a_1 \ldots a_n)\|_2$, so under the identification $(\mathbb{R}^n)^* = \mathbb{R}^n$ determined by the standard coordinate system we see that $\|\cdot\|_2^*$ is naturally identified with $\|\cdot\|_2$.

Note that we did not assume at the outset that V is complete with respect to its norm. It is quite convenient that $(V^*, \|\cdot\|^*)$ is complete whether or not V is.

Proposition 10.6. *$(V^*, \|\cdot\|^*)$ is a Banach space.*

Proof. Assume $\{\ell_n\}$ is a Cauchy sequence in V^*. Simple estimates show that $\{\ell_n(\mathbf{x})\}$ is a Cauchy sequence of real numbers for each $\mathbf{x} \in V$, so by completeness of $\mathbb{R}$ it converges to a real number $\ell(\mathbf{x})$. Passing to the limit as n tends to infinity in the equation $\ell_n(\mathbf{x}) + \ell_n(\mathbf{y}) = \ell_n(\mathbf{x} + \mathbf{y})$, we obtain that ℓ is linear. ℓ is bounded since $|\ell(\mathbf{x})| \leq |\ell(\mathbf{x}) - \ell_n(\mathbf{x})| + |\ell_n(\mathbf{x})|$ and ℓ_n is bounded, so $\ell \in V^*$. To complete the proof we need to show that ℓ_n converges to ℓ in norm.

For every $\varepsilon > 0$ there exists n such that $\|\ell_n - \ell_m\|^* < \varepsilon$ if $n, m \geq N$. Equivalently, $|\ell_n(\mathbf{x}) - \ell_m(\mathbf{x})| < \varepsilon$ for every $\mathbf{x}$ such that $\|\mathbf{x}\| = 1$. Fixing n and passing to the limit as m tends to infinity, it follows that $|\ell_n(\mathbf{x}) - \ell(\mathbf{x})| \leq \varepsilon$ whenever $n \geq N$ and $\|\mathbf{x}\| = 1$. Thus $\|\ell_n - \ell\|^* \leq \varepsilon$ for $n \geq N$, as desired. □

Remark 10.7. It may be useful to compare this proof with the proof that $C[0, 1]$ is a Banach space (Theorem 1.16).

11. Lecture 11 (9/16/11): The Double Dual

Now that V^* is equipped with a norm, the double dual V^{**} is well-defined for any V. Moreover the norm on V^* induces a norm $\|\cdot\|^{**}$ on V^{**}. Recall that in the finite dimensional case we defined an embedding $i : V \to V^{**}$ by the formula $i_{\mathbf{x}}(\ell) = \ell(\mathbf{x})$; to extend this definition to the infinite dimensional case, we need to verify that $i_{\mathbf{x}}$ is a bounded (and therefore continuous) linear functional for every $\mathbf{x}$. This will follow from the following calculation.

Lemma 11.1. *$\|i_{\mathbf{x}}\|^{**} \leq \|\mathbf{x}\|$ for every $\mathbf{x} \in V$.*

Proof. Fix $\mathbf{x}$ and take $\ell \in V^*$ such that $\|\ell\|^* = 1$. By definition this implies that $|\ell(\mathbf{x})| \leq \|\mathbf{x}\|$ and thus $|i_{\mathbf{x}}(\ell)| = |\ell(\mathbf{x})| \leq \|\mathbf{x}\|$. We conclude that $\|i_{\mathbf{x}}\|^{**} = \sup_{\|\ell\|^*=1} |i_{\mathbf{x}}(\ell)| \leq \|\mathbf{x}\|$, as desired. □

In fact, it is the case that $\|i_{\mathbf{x}}\|^{**} = \|\mathbf{x}\|$ and thus the canonical embedding $i : V \to V^{**}$ preserves the norm on V. However, there are many examples - even among Banach spaces - where i fails to be surjective. A normed space for which i is surjective (and hence an isomorphism) is said to be *reflexive*; we will see that

ℓ^p is an infinite dimensional reflexive Banach space if $p > 1$, but $C[0,1]$ is not. A reflexive normed space is necessarily a Banach space by Lemma 10.6, but the canonical embedding is a useful tool even for incomplete normed spaces because the closure of V inside V^{**} is a model for the completion of V.

To prove that the canonical embedding preserves the norm on V, we need to show that for every $\mathbf{x} \in V$ there exists $\ell \in V^*$ of norm 1 with the property that $\ell(\mathbf{x}) = \|\mathbf{x}\|$. In infinite dimensions the proof requires an important result known as the Hahn-Banach theorem. Part of the difficulty of this theorem is that it depends crucially on the axiom of choice, and hence we cannot hope for a constructive way to produce the desired functional ℓ. Instead we will use the tools of convex geometry to infer the existence of ℓ abstractly - we will revisit this problem in a few lectures when we begin investigating convex geometry in earnest.

12. Lecture 12 (9/19/11):The Dual of $\|\cdot\|_p$

We now revisit the ℓ^p norms introduced in Lecture 2. Recall that for $p \geq 1$ the ℓ^p norm on $\mathbb{R}^n$ is given by:

$$\|\mathbf{x}\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$$

The proof that this really is a norm for $p \geq 1$ was left as an exercise in Lecture 2, but due to the importance of this fact we will include an easy proof here.

Lemma 12.1. *$\|\cdot\|_p$ is a norm on $\mathbb{R}^n$ for any $p \geq 1$.*

Proof. Positive definiteness and homogeneity are easily checked and left to the reader. We will prove that $\|\cdot\|_p$ is convex. First we'll show that $\|\mathbf{x}\|_p^p = \sum_i |x_i|^p$ is convex. Since sums of convex functions are convex, it suffices to show that the function $\mathbb{R} \to \mathbb{R}$ given by $t \mapsto |t|^p$ is convex. We have $|t+0|^p = |t|^p + |0|^p$, so it suffices to show that $|s+t|^p \leq |s|^p + |t|^p$ for s and t nonzero. This reduces to the inequality for s and t positive, and to prove it in this case we use the fact that a smooth function is convex if and only if its second derivative is everywhere nonnegative. But $\frac{d^2}{dt^2} t^p = p(p-1)t^{p-2}$, and this is nonnegative whenever t is positive since $p \geq 1$.

This shows that $\|\mathbf{x}+\mathbf{y}\|_p \leq (\|\mathbf{x}\|_p^p + \|\mathbf{y}\|_p^p)^{1/p}$ for every $\mathbf{x}$ and $\mathbf{y}$. But for any nonnegative a, b and any $p \geq 1$ we have $(a+b)^{1/p} \leq a^{1/p} + b^{1/p}$. So the proof is complete. □

Let $p > 1$ and consider $\mathbb{R}^n$ equipped with the norm $\|\cdot\|_p$. Recall that every linear functional ℓ on $\mathbb{R}^n$ can be expressed in standard coordinates as $\ell(\mathbf{x}) = \sum_{i=1}^n a_i x_i$, so that the assignment $\ell \mapsto (a_1, \ldots, a_n)$ yields a linear isomorphism $(\mathbb{R}^n)^* \to \mathbb{R}^n$. With this identification in mind the dual norm $\|\cdot\|_p^*$ can be regarded as a norm on $\mathbb{R}^n$, and thus we can hope to identify it explicitly. It turns out that $\|\ell\|_p^* = \|(a_1, \ldots, a_n)\|_q$ where q is determined by the identity $\frac{1}{p} + \frac{1}{q} = 1$. In order to prove this, we need an important result in Banach space theory known as *Holder's Inequality*.

Proposition 12.2 (Holder's Inequality)**.** *Let p be any real number strictly greater than 1 and let $q = \frac{p}{p-1}$, so that $\frac{1}{p} + \frac{1}{q} = 1$. For any vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ we have $\sum_{i=1}^n |x_i y_i| \leq \|\mathbf{x}\|_p \|\mathbf{y}\|_q$ with equality if and only if $\|\mathbf{y}\|_q^q |x_i|^p = \|\mathbf{x}\|_p^p |y_i|^q$ for $1 \leq i \leq n$.*

Proof. If the inequality is true for $\mathbf{x}$ and $\mathbf{y}$ then it also holds for $\alpha\mathbf{x}$ and $\beta\mathbf{y}$ for any pair of real numbers α, β. So we may assume without loss of generality that $\|\mathbf{x}\|_p = \|\mathbf{y}\|_q = 1$. With this assumption it suffices to prove the inequality

$$|x_i y_i| \leq \frac{|x_i|^p}{p} + \frac{|y_i|^q}{q} \tag{12.1}$$

because the sum of the right-hand side over i gives

$$\frac{\|\mathbf{x}\|_p^p}{p} + \frac{\|\mathbf{y}\|_q^q}{q} = \frac{1}{p} + \frac{1}{q} = 1 = \|\mathbf{x}\|_p \|\mathbf{y}\|_q$$

To prove (12.1), first note that it is trivial if x_i or y_i is 0. Dividing (12.1) by $|y_i|^q$ and substituting $t = \frac{|x_i|^p}{|y_i|^q}$, $\lambda = \frac{1}{p}$, we reduce the problem to proving the inequality $t^\lambda \leq \lambda t + (1-\lambda)$ for $t > 0$ and $\lambda \in (0,1)$. Set $f(t) = t^\lambda - \lambda t$ and calculate that $f'(t) = \lambda t^{\lambda-1} - \lambda$. We have $f'(t) > 0$ for $t \in (0,1)$, $f'(t) = 0$ if and only if $t = 1$, and $f'(t) > 0$ for $t > 1$, so $f(t)$ is globally maximized by $f(1) = 1 - \lambda$. This completes the proof of the inequality asserted by the proposition. Additionally, we get for free that when $\|\mathbf{x}\|_p = \|\mathbf{y}\|_q = 1$, equality holds if and only if $t = \frac{|x_i|^p}{|y_i|^q} = 1$. The general equality condition follows immediately. □

Exercise 12.3. Use Holder's inequality to give a new proof that $\|\cdot\|_p$ is convex.

Holder's inequality allows us to prove that $\|\cdot\|_p^* = \|\cdot\|_q$ in the sense described above.

Proposition 12.4. *Let $\ell \in (\mathbb{R}^n)^*$, so that ℓ has the form $\ell(\mathbf{x}) = \sum_{i=1}^n a_i x_i$. Then $\|\ell\|_p^* = \|\mathbf{a}\|_q$ where $\mathbf{a} = (a_1, \ldots, a_n)$.*

Proof. For any $\mathbf{x} \in \mathbb{R}^n$ we have

$$|\ell(\mathbf{x})| \leq \sum_{i=1}^n |a_i x_i| \leq \|\mathbf{x}\|_p \|\mathbf{a}\|_q$$

by Holder's inequality. Thus $\|\ell\|_p^* = \sup_{\|\mathbf{x}\|_p=1} |\ell(\mathbf{x})| \leq \|\mathbf{a}\|_q$. To prove equality, it suffices to show that there is some $\mathbf{x}$ such that $\ell(\mathbf{x}) = \|\mathbf{a}\|_q$. Define $\mathbf{x}$ coordinate-wise by $x_i = a_i^{q-1}$; this gives $\|\mathbf{x}\|_p^p = \sum_{i=1}^n |x_i|^p = \sum_{i=1}^n |a_i|^{p(q-1)} = \|\mathbf{a}\|_q^q$ since $p(q-1) = q$. By the equality case of Holder's inequality $\ell(\mathbf{x}) = \|\mathbf{a}\|_q$, as desired. □

The ℓ^p norm admits two different generalizations to infinite dimensions. The first is the space ℓ^p of Definition 2.2 consisting of all infinite sequences $\{x_n\}$ of real numbers such that $\sum_{n=1}^\infty |x_n|^p < \infty$. Combining Lemma 12.1 and Lemma 2.3, we now know that ℓ^p is a normed space when $p > 1$. We will now prove the analogue of Proposition 12.4 for the ℓ^p spaces.

Let $\mathbf{e}_m$ denote the infinite sequence whose nth entry is 0 if $m \neq n$ and is 1 if $m = n$. Given any linear functional ℓ on ℓ^p, we can associate to ℓ the infinite sequence $\{\ell(\mathbf{e}_1), \ell(\mathbf{e}_2), \ldots\}$; we shall prove that if ℓ is bounded then the dual norm of ℓ agrees with the ℓ^q norm of the sequence.

Theorem 12.5. *Let $\ell \in (\ell^p)^*$ for $p > 1$ and let $a_n = \ell(\mathbf{e}_n)$. Then $\mathbf{a} = \{a_n\}$ is in ℓ^q where $\frac{1}{p} + \frac{1}{q} = 1$ and $\|\ell\|_p^* = \|\mathbf{a}\|_q$.*

Proof. Consider the truncation ℓ^M given by $\ell^M(\mathbf{x}) = \sum_{n=1}^M a_n x_n$. By Proposition 12.4, $\left\|\ell^M\right\|_p^* = \left\|\mathbf{a}^M\right\|_q$ where $\mathbf{a}^M = (a_1, \ldots, a_M, 0, 0, \ldots)$. Now, $\left\|\ell - \ell^M\right\|_p^* = \sup_{\|\mathbf{x}\|_p=1} \left|\sum_{n=M+1}^\infty a_n x_n\right|$, and this tends to 0 as $M \to \infty$ by the definition of convergence for infinite series. Similarly $\mathbf{a}^M \to \mathbf{a}$ with respect to $\|\cdot\|_q$. Thus $\|\ell\|_p^* = \lim_M \left\|\ell^M\right\|_p^* = \lim_M \left\|\mathbf{a}^M\right\|_q = \|\mathbf{a}\|_q$. In particular $\mathbf{a} \in \ell^q$ since ℓ is bounded. $\square$

Corollary 12.6. *ℓ^p is a reflexive Banach space for $p > 1$.*

Proof. ℓ^p is a Banach space by Proposition 10.6 since ℓ^p is the dual of ℓ^q where $\frac{1}{p} + \frac{1}{q} = 1$. It is reflexive since the equation $\frac{1}{p} + \frac{1}{q} = 1$ is symmetric in p and q. $\square$

This gives a "discrete" infinite dimensional generalization of the norm $\|\cdot\|_p$ on $\mathbb{R}^n$. There is also a very important "continuous" generalization which we now define. It is closely related to integration of functions on the real line.

Let V_n denote the vector space of left continuous functions on $[0,1]$ which are constant on the intervals $(\frac{i}{2^n}, \frac{i+1}{2^n}]$ for $0 \leq i \leq 2^n - 1$. We have inclusion maps $i_n : V_n \to V_{n+1}$. Each V_n is isomorphic as a vector space to $\mathbb{R}^{2^n}$, so we can equip V_n with the ℓ^p norm. It is convenient, however, to normalize the ℓ^p norm by defining $\|\cdot\|_{p,n} = \|\cdot\|_p$.

The reason for this normalization is that when each V_n is equipped with $\|\cdot\|_{p,n}$ the inclusion map i_n becomes an isometry. Indeed, let $f \in V_n$ be the function which takes the value a_i on the interval $(\frac{i}{2^n}, \frac{i+1}{2^n}]$, so that f corresponds to the element $(a_0, a_1, \ldots, a_{2^n-1})$ of $\mathbb{R}^{2^n}$. Then $i_n(f)$ corresponds to the element $(a_0, a_0, a_1, a_1, \ldots, a_{2^n-1}, a_{2^n-1})$ of $\mathbb{R}^{2^{n+1}}$. Thus

$$\begin{aligned}
\|i_n(f)\|_{p,n+1} &= 2^{-(n+1)/p}(\sum_{i=0}^{2^{n-1}} 2\,|a_i|^p)^{1/p} \\
&= 2^{-n/p}(\sum_{i=0}^{2^n-1} |a_i|^p)^{1/p} \\
&= \|f\|_{p,n}
\end{aligned}$$

Thus we have an infinite chain of isometric inclusions $V_1 \overset{i_1}{\hookrightarrow} V_2 \overset{i_2}{\hookrightarrow} \ldots$. Let $V_\infty = \bigcup_{n=1}^\infty V_n$, and note that V_∞ has a natural vector space structure coming from the fact that all of its elements are functions on $[0,1]$. Moreover V_∞ has a norm $\|\cdot\|_p$ defined by setting $\|f\|_p = \|f\|_{p,n}$ where n is any number such that $f \in V_n$. Since the inclusion maps are isometries, $\|f\|_{p,n}$ is independent of n.

Define $L^p[0,1]$ to be the completion of V_∞ in the norm $\|\cdot\|_p$. It is not at all obvious how to describe $L^p[0,1]$ or its norm in any explicit way, but we shall see that $L^p[0,1]$ consists of equivalence classes of certain functions on $[0,1]$ and that the norm of the equivalence class of a continuous function f is given by $(\int_0^1 |f(x)|^p\,dx)^{1/p}$. The close relationship between the L^p-spaces and integration makes them invaluable in real analysis.

13. Lecture 13 (9/21/11): Convex Geometry

The Banach space ℓ^p has many nice properties not enjoyed by other infinite dimensional Banach spaces; we have seen, for instance, that ℓ^p is reflexive. We call attention to a particular property that is closely related to its geometric structure. Given any linear functional $\ell \in (\ell^p)^*$, recall that we can write $\ell(\mathbf{x}) = \sum_{n=1}^{\infty} a_n x_n$ where $a_n = \ell(\mathbf{e}_n)$. Recall further that $\mathbf{a} = \{a_n\}$ is in ℓ^q where $\frac{1}{p} + \frac{1}{q} = 1$, and thus we can define a specific element $\mathbf{x} \in \ell^p$ by setting $x_i = \operatorname{sign}(a_i)\,|a_i|^{q-1}$ and normalizing to guarantee $\|\mathbf{x}\|_p = 1$. By the equality case of Holder's inequality, we have $\ell(\mathbf{x}) = \|\mathbf{x}\|_p \|\mathbf{a}\|_q = \|\mathbf{a}\|_q = \|\ell\|_p^*$. Moreover the $\mathbf{x}$ we constructed is the only element of ℓ^p with this property. Thus we can construct a unique element of the unit ball in ℓ^p at which ℓ precisely achieves its supremum. For more general Banach spaces there is no guarantee that such an element exists, and even when one exists it need not be unique.

Example 13.1. Consider $\mathbb{R}^2$ equipped with $\|\cdot\|_1$ and the linear functional $\ell(x, y) = x + y$. It is not difficult to check that $\|\ell\|_1^* = 1$, and any vector of the form $(t, 1-t)$ satisfies $\ell(t, 1-t) = 1$. So uniqueness is not guaranteed even in finite dimensions.

Example 13.2. By compactness of the unit ball in a finite dimensional normed space, we must look in infinite dimensions to observe the failure of existence. Consider the Banach space $C[0, 1]$ equipped with the uniform norm and define a linear functional on this space by $\ell(f) = \int_0^{1/2} f(x)dx - \int_{1/2}^1 f(x)dx$. This linear functional is continuous because $\lim_n \int_a^b f_n(x)dx = \int_a^b f(x)dx$ whenever f_n converges to f uniformly. However the maximum value of ℓ on the unit ball, if it were attained, would have to occur at the function

$$f(x) = \begin{cases} 1 & x \in [0, \frac{1}{2}] \\ -1 & x \in (\frac{1}{2}, 1] \end{cases}$$

This function is evidently not continuous, so $\|\ell\|^*$ is not attained as a value of ℓ.

Hiding behind both of these examples are subtle issues in convex geometry. We have already seen some of the basic techniques of convex geometry in Lecture 8 when we gave a geometric characterization of the unit ball in a finite dimensional Banach space. We will now investigate some further techniques and explore their applications to functional analysis.

Recall from Definition 8.3 that a subset C of a vector space V over $\mathbb{R}$ is convex if for every $\mathbf{x}, \mathbf{y} \in C$ and every $\alpha \in [0, 1]$ we have $\alpha\mathbf{x} + (1-\alpha)\mathbf{y} \in C$.

Definition 13.3. Let C be a convex set. A point $\mathbf{x} \in C$ is an *extreme point* if for any expression $\mathbf{x} = t\mathbf{y} + (1-t)\mathbf{z}$ with $\mathbf{y}, \mathbf{z} \in C$ and $t \in [0, 1]$ we have $\mathbf{x} = \mathbf{y} = \mathbf{z}$.

Said differently, $\mathbf{x}$ is an extreme point of C if and only if $\mathbf{x}$ is an endpoint of any line segment in C which contains $\mathbf{x}$.

Exercise 13.4. Show that the extreme points of the unit ball of $\mathbb{R}^n$ equipped with the ℓ^1 norm are precisely $\pm\mathbf{e}_j$ where $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is the standard basis for $\mathbb{R}^n$. Show that every point in the boundary of the unit ball of $\mathbb{R}^n$ with the ℓ^p norm, $p > 1$, is an extreme point of the unit ball.

Definition 13.5. A *polyhedron* in $\mathbb{R}^n$ is a compact convex set with finitely many extreme points.

This definition may seem strange to the reader who is accustomed to thinking of polyhedra as objects built out of vertices, edges, faces, and so on. We will prove that our definition of polyhedron agrees with the more naive definition. Our proof will occupy the next few lectures, and it will illustrate two general principles in convex geometry.

- **General Principle 1**: A convex set is determined by its extreme points.

When this principle holds, the convex set can be realized as the "smallest" convex set containing all of the extreme points in the following sense:

Definition 13.6. Let S be any subset of $\mathbb{R}^n$. The *convex hull* of S is the intersection of all convex sets containing S. It is denoted by $C(S)$.

Remark 13.7. $C(S)$ exists because the collection of all convex sets containing S is nonempty ($\mathbb{R}^n$ is a convex set containing S) and closed under intersections (the intersection of arbitrarily many convex sets is convex).

- **General Principle 2**: Disjoint convex sets can be separated by hyperplanes.

Actually, we will really only need that a convex set can be separated from a point by a hyperplane. The family of hyperplanes which separate a convex set from the points in its complement can in favorable circumstances be used to characterize the convex set.

Both principles are illustrated in the case of polyhedra by the following key theorem:

Theorem 13.8. *The following statements about a convex compact set $C \subseteq \mathbb{R}^n$ are equivalent:*

(1) *C is a polyhedron.*
(2) *C is the convex hull of a finite set.*
(3) *C is the intersection of finitely many half-spaces in $\mathbb{R}^n$.*

14. Lecture 14 (9/23/11): Convex Geometry, Continued

Our next result will be one of our most important tools in our investigation of convex geometry. It is a statement about linear functionals on $\mathbb{R}^n$, and as an immediate byproduct it will help us complete our unfinished program in Lecture 11 of proving that the canonical embedding of a Banach space into its double dual is an isometry.

Lemma 14.1. *Let $L \subseteq \mathbb{R}^n$ be a linear subspace and let $\rho : \mathbb{R}^n \to \mathbb{R}$ be a convex function such that $\rho(t\mathbf{x}) = t\rho(\mathbf{x})$ for $t > 0$. If $f_0 : L \to \mathbb{R}$ is a linear functional with the property that $f_0(\mathbf{x}) \leq \rho(\mathbf{x})$ for every $\mathbf{x} \in L$ then there is a linear functional $f : \mathbb{R}^n \to \mathbb{R}$ such that $f|_L = f_0$ and $f(\mathbf{x}) \leq \rho(\mathbf{x})$ for every $\mathbf{x} \in \mathbb{R}^n$.*

Proof. We proceed by induction on the codimension $k = n - \dim(L)$ of L in $\mathbb{R}^n$. The result is obvious when $k = 0$ because then $L = \mathbb{R}^n$ and f_0 is its own extension. Assume the result is true for any subspace of codimension k and let L be a subspace of codimension $k+1$. Take $\mathbf{z}$ not in L and let $L' = \{\mathbf{x} + t\mathbf{z} : \mathbf{x} \in L, t \in \mathbb{R}\}$. L' is a subspace of codimension k, so by the inductive assumption it suffices to construct an extension of f_0 from L to L'.

The values of f on L are determined by the condition that $f|_L = f_0$, so we need only choose a value c for $f(\mathbf{z})$ such that f is bounded by ρ on all of L'. Our strategy is to write down the system of inequalities which express this assertion and prove that the system has a solution.

The statement that $f(\mathbf{x}+t\mathbf{z}) \leq \rho(\mathbf{x}+t\mathbf{z})$ together with the calculation $f(\mathbf{x}+t\mathbf{z}) = f_0(\mathbf{x})+ct$ gives $\rho(\mathbf{x}+t\mathbf{z})-f_0(\mathbf{x}) \geq ct$ for $t>0$. In other words $c \leq \rho(\frac{\mathbf{x}}{t}+\mathbf{z})-f_0(\frac{\mathbf{x}}{t})$ for every $\mathbf{x} \in L$ and every $t > 0$. By a similar argument we have $\rho(\mathbf{y}-cs)-f_0(\mathbf{y}) \geq -cs$ for $s < 0$, or in other words $c \geq f_0(\frac{\mathbf{y}}{s}) - \rho(\frac{\mathbf{y}}{s} - \mathbf{z})$ for every $\mathbf{y} \in L$ and every $s < 0$. So the desired value c exists if and only if

(14.1)

$$\sup\{f_0\left(\frac{\mathbf{y}}{s}\right) - \rho\left(\frac{\mathbf{y}}{s} - \mathbf{z}\right) : \mathbf{y} \in L, s < 0\} \leq \inf\{\rho\left(\frac{\mathbf{x}}{t} + \mathbf{z}\right) - f_0\left(\frac{\mathbf{x}}{t}\right) : \mathbf{x} \in L, t > 0\}$$

But since $f_0 \leq \rho$ on L and ρ is convex we have for any $\mathbf{x}, \mathbf{y} \in L$ and any s, t:

$$\begin{aligned} f_0\left(\frac{\mathbf{x}}{t} + \frac{\mathbf{y}}{s}\right) &\leq \rho\left(\frac{\mathbf{x}}{t} + \frac{\mathbf{y}}{s}\right) \\ &= \rho\left(\frac{\mathbf{x}}{t} + \mathbf{z} + \frac{\mathbf{y}}{s} - \mathbf{z}\right) \\ &\leq \rho\left(\frac{\mathbf{x}}{t} + \mathbf{z}\right) + \rho\left(\frac{\mathbf{y}}{s} - \mathbf{z}\right) \end{aligned}$$

Subtracting $f_0\left(\frac{\mathbf{x}}{t}\right) + \rho\left(\frac{\mathbf{y}}{s} - \mathbf{z}\right)$ from both sides, the inequality (14.1) follows. □

Remark 14.2. The same argument shows that if V is an infinite dimensional space and L is a subspace of finite codimension then any linear functional on L bounded by ρ extends to a linear functional on V bounded by ρ. In fact the statement for arbitrary subspaces is true, but when L has infinite codimension the proof requires some sort of transfinite induction (and hence the axiom of choice). All of the geometric details relevant to the general case are present in the proof above - only additional set theoretic machinery is required.

Corollary 14.3. *Let $(V, \|\cdot\|)$ be a finite dimensional normed space and let $i : V \to V^{**}$ be the canonical embedding. Then i is an isometry.*

Proof. Given the discussion following Lemma 9.3, it suffices to show that for every $\mathbf{x} \in V$ there exists $\ell \in V^*$ of norm 1 such that $\ell(\mathbf{x}) = \|\mathbf{x}\|$. Let L be the span of $\mathbf{x}$ (i.e. the set of all multiples of $\mathbf{x}$) and let $\ell_0 : L \to \mathbb{R}$ be the linear functional $\ell_0(\alpha\mathbf{x}) = \alpha\|\mathbf{x}\|$. Clearly $\ell_0 \leq \|\cdot\|$ on L, so by Lemma 14.1 it extends to a linear functional $\ell \in V^*$ such that $\ell \leq \|\cdot\|$. The condition that $\ell \leq \|\cdot\|$ says that $\|\ell\|^* \leq 1$, and the fact that ℓ extends ℓ_0 implies that $\ell(\mathbf{x}) = \|\mathbf{x}\|$ (in particular, $\|\ell\|^* = 1$). □

Let us examine the proof of 14.3 from a geometric point of view. Recall from Theorem 8.10 that every norm on V arises as the *Minkowski functional* of its unit ball, where the Minkowski functional of a set $B \subseteq V$ is the function $\rho : V \to R$ given by

$$\rho(\mathbf{x}) = \frac{1}{\sup\{\alpha : \alpha\mathbf{x} \in B\}}$$

We argued that ρ is a norm whenever B is a compact, convex, centrally symmetric set which contains a basis of V. However, the only point in the proof of Theorem 8.10 where we made any essential use of central symmetry was to prove that the Minkowski functional is homogeneous for arbitrary constants: $\rho(t\mathbf{x}) = |t|\,\rho(\mathbf{x})$. Since Lemma 9.3 only requires homogeneity for positive constants, we can apply it to Minkowski functionals of more general sets.

Definition 14.4. A *convex body of dimension k* is a closed, compact subset of $\mathbb{R}^n$ which contains a linearly independent set of k vectors but no linearly independent set of $k+1$ vectors. The *Minkowski functional* of a convex body B in $\mathbb{R}^n$ is the function $\rho_B : \mathbb{R}^n \to R$ defined by:

$$\rho_B(\mathbf{x}) = \frac{1}{\sup\{\alpha : \alpha\mathbf{x} \in B\}}$$

Note that a convex body of dimension k in $\mathbb{R}^n$ sits inside a hyperplane of dimension k, so when we refer to a convex body in $\mathbb{R}^n$ without specifying the dimension it will be assumed that the dimension is n. The proof of Theorem 8.10 guarantees that the Minkowski functional of an n-dimensional convex body in $\mathbb{R}^n$ which contains the origin is convex, positive definite, and homogeneous for positive constants.

Lemma 14.5. *Let $B \subseteq \mathbb{R}^n$ be a convex body and let $\rho_B : \mathbb{R}^n \to \mathbb{R}$ be the Minkowski functional for B.*

(1) *If $\mathbf{y}$ is in the boundary of B then there is a linear functional ℓ on $\mathbb{R}^n$ such that $\ell(\mathbf{x}) \leq \ell(\mathbf{y})$ for every $\mathbf{x} \in B$.*
(2) *If $\mathbf{y}$ is not in B then there is a linear functional ℓ on $\mathbb{R}^n$ such that $\ell(\mathbf{x}) \leq 1 < \ell(\mathbf{y})$ for every $\mathbf{x} \in B$.*

Proof. The proof begins the same way for both statements. The lemma is invariant under translation, so assume without loss of generality that B contains the origin. Let L be the line $L = \{\alpha\mathbf{y} :\ \alpha \in \mathbb{R}\}$ and define a linear functional $\ell_0 : L \to \mathbb{R}$ by $\ell_0(\alpha\mathbf{y}) = \alpha\rho_B(\mathbf{y})$. If $\alpha \geq 0$ then $\ell_0(\alpha\mathbf{y}) = \rho_B(\alpha\mathbf{y})$, and if $\alpha < 0$ then $\ell_0(\alpha\mathbf{y}) = -\rho_B(-\alpha\mathbf{y}) \leq 0 \leq \rho_B(\alpha\mathbf{y})$. So $\ell_0 \leq \rho_B$ on L. By Lemma 14.1, ℓ_0 can be extended to a linear functional ℓ on $\mathbb{R}^n$ with the property that $\ell \leq \rho_B$.

For any $\mathbf{x} \in B$ we have $\rho_B(\mathbf{x}) \leq 1$ with equality if and only if $\mathbf{x}$ is in the boundary of B. So if $\mathbf{y}$ is in the boundary of B then we have for every $\mathbf{x} \in B$ that $\ell(\mathbf{x}) \leq \rho_B(\mathbf{x}) \leq 1 = \rho_B(\mathbf{y}) = \ell(\mathbf{y})$. This proves the first statement.

If $\mathbf{y}$ is not in B then $\rho_B(\mathbf{y}) > 1$ since B is convex and contains the origin, so for every $\mathbf{x} \in B$ we have $\ell(\mathbf{x}) \leq \rho_B(\mathbf{x}) \leq 1 < \rho_B(\mathbf{y}) = \ell(\mathbf{y})$. This completes the proof. □

Geometrically, the lemma says that the geometry of a compact, convex set which contains a basis for $\mathbb{R}^n$ can be probed using hyperplanes in $\mathbb{R}^n$. We will introduce some language to help make this more precise.

Definition 14.6. Let $B \subseteq \mathbb{R}^n$ be any set and let ℓ be a linear functional on $\mathbb{R}^n$.

- Say that B is *supported* by the hyperplane $\ell = c$ if $\ell(\mathbf{x}) \leq c$ for every $\mathbf{x} \in B$ and $\ell(\mathbf{y}) = c$ for at least one point $\mathbf{y} \in B$.
- Say that B is *separated* from a point $\mathbf{y} \in \mathbb{R}^n$ by the hyperplane $\ell = c$ if $\ell(\mathbf{x}) \leq c < \ell(\mathbf{y})$ for every $\mathbf{x} \in B$.

Thus Lemma 14.5 asserts that any convex body is supported by a hyperplane at each point on its boundary and it is separated from every point in its complement by a hyperplane.

15. Lecture 15 (9/26/11): General Theory of Convex Bodies

Our present goal is to prove Theorem 13.8 which asserts that a polyhedron can be characterized either by its vertices (as a set with finitely many extreme points) or by its faces (as a set which is the intersection of finitely many half-spaces). First we will give some characterizations of more general convex bodies, beginning with the following:

Proposition 15.1. *Every convex body is the intersection of half-spaces.*

Proof. Let B be a convex body. For every $\mathbf{x}$ in the boundary of B, there is a linear functional $\ell_{\mathbf{x}}$ such that $\ell(\mathbf{x}) = 1$ and $\ell(\mathbf{y}) \leq 1$ for every $\mathbf{y} \in B$ by Lemma 14.5. Thus $B \subseteq \bigcap_{\mathbf{x}\in\partial B} \ell^{-1}(-\infty, 1]$, an intersection of half-spaces. Applying Lemma 14.5 again, any point not in B can be separated from B by one of the hyperplanes $\ell_{\mathbf{x}}$, so in fact B is precisely the intersection of the half-spaces $\ell^{-1}(-\infty, 1]$. $\square$

It is a little more challenging to characterize a convex body using extreme points. To start we must show that extreme points always exist.

Proposition 15.2. *Every convex body has an extreme point.*

Proof. We proceed by induction on the dimension of the convex body (i.e. the size of the largest linearly independent set that it contains). The only convex bodies of dimension 1 are closed intervals in the real line, and the extreme points of a closed interval are precisely the two endpoints. So assume that every convex body of dimension no larger than k has an extreme point and let B be a convex body of dimension $k+1$. Choose any nonzero linear functional ℓ on $\mathbb{R}^{k+1}$; by compactness, ℓ attains a maximum value $m = \ell(\mathbf{x})$ on B. The intersection of B with the hyperplane $\ell^{-1}(m)$ is a convex body of dimension at most k, so by the inductive hypothesis $B \cap \ell^{-1}(m)$ has an extreme point $\mathbf{x}$. Suppose $\mathbf{x} = t\mathbf{y} + (1-t)\mathbf{z}$ for $\mathbf{y}, \mathbf{z} \in B$ and $t \in (0, 1)$. We have $m = \ell(\mathbf{x}) = t\ell(\mathbf{y}) + (1-t)\ell(\mathbf{z}) \leq tm + (1-t)m = m$ since m is the maximum of ℓ on B. Thus $\ell(\mathbf{y}) = \ell(\mathbf{z}) = m$, i.e. $\mathbf{y}, \mathbf{z} \in \ell^{-1}(m) \cap B$. Since $\mathbf{x}$ is an extreme point of $\ell^{-1}(m) \cap B$, we conclude that $\mathbf{x} = \mathbf{y} = \mathbf{z}$ and hence $\mathbf{x}$ is an extreme point of B. $\square$

We can do even better. In each dimension there is a "simplest" convex body called a *simplex*, and any convex body B can be expressed as the union of simplices in such a way that the vertices of the simplices used are precisely the extreme points of B. Before proving this let us give a precise definition of simplex.

Recall that the convex hull of a set $S \subseteq \mathbb{R}^n$, denoted by $C(S)$, is the intersection of all convex sets which contain S. We give another description of the convex hull of a finite set.

Definition 15.3. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a finite set of vectors in $\mathbb{R}^n$. A *convex combination* of $\mathbf{v}_1, \ldots, \mathbf{v}_k$ is any vector in $\mathbb{R}^n$ of the form $\sum_{i=1}^k s_i\mathbf{v}_i$ where $0 \leq s_i \leq 1$ and $\sum_{i=1}^k s_i = 1$.

Proposition 15.4. *The convex hull of a finite set F is precisely the set of all convex combinations of the vectors in F.*

Proof. For any finite set F, let F' denote the set of all convex combinations of the vectors in F. By setting all but one of the coefficients in the definition of convex combination equal to zero, it is clear that $F \subseteq F'$; let us show that F' is convex. Take $\mathbf{x}, \mathbf{y} \in F'$ so that $\mathbf{x} = \sum_{i=1}^k s_i\mathbf{v}_i$ and $\mathbf{y} = \sum_{i=1}^k t_i\mathbf{v}_i$ where the s_i's are numbers in $[0, 1]$ which sum to 1 and likewise for the t_i's. If $\mathbf{z} = u\mathbf{x} + (1-u)\mathbf{y}$ then $\mathbf{z} = \sum_{i=1}^k (us_i + (1-u)t_i)\mathbf{v}_i$, and the numbers $(us_i + (1-u)t_i)$ lie in $[0, 1]$ and sum to 1. So $\mathbf{z} \in F'$, as desired.

It remains only to show that any convex set K which contains F also contains F'. We proceed by induction on the size of F. If F has only one element then K contains F' since $F' = F$, so assume inductively that K contains F' whenever

$F \subseteq K$ has n elements and let $F = \{\mathbf{v}_1, \ldots, \mathbf{v}_{n+1}\}$ be a subset of K with $n+1$ elements. Given $\mathbf{x} \in F'$ write $\mathbf{x} = \sum_{i=1}^{n+1} s_i \mathbf{v}_i$ where $0 \le s_i \le 1$ and $\sum_{i=1}^{n+1} s_i = 1$. We have:

$$\begin{aligned}\mathbf{x} &= \sum_{i=1}^{n} s_i \mathbf{v}_i + s_{n+1}\mathbf{v}_{n+1} \\ &= (s_1 + \ldots + s_n) \sum_{i=1}^{n} \frac{s_i}{s_1 + \ldots + s_n} \mathbf{v}_i + s_{n+1}\mathbf{v}_{n+1}\end{aligned}$$

The vector $\mathbf{w} = \sum_{i=1}^{n} \frac{s_i}{s_1+\ldots+s_n}\mathbf{v}_i$ is in K by the inductive hypothesis since it is a convex combination of n vectors in K. Thus $(s_1 + \ldots + s_n)\mathbf{w} + s_{n+1}\mathbf{v}_{n+1} \in K$ since $\mathbf{w}, \mathbf{v}_{n+1} \in K$ and $(s_1 + \ldots + s_n) + s_{n+1} = 1$. So $\mathbf{x} \in K$ and the proof is complete. □

We are now ready to discuss the notion of a simplex in detail.

Definition 15.5. An *n-simplex* is the convex hull of a set $\{\mathbf{v}_1, \ldots, \mathbf{v}_{n+1}\}$ in $\mathbb{R}^n$ which has the property that $\{\mathbf{v}_2 - \mathbf{v}_1, \ldots, \mathbf{v}_{n+1} - \mathbf{v}_n\}$ is a basis for $\mathbb{R}^n$. The points $\mathbf{v}_1, \ldots, \mathbf{v}_{n+1}$ are called the *vertices* of the simplex.

Translating so that $\mathbf{v}_1 = \mathbf{0}$, one can think of a simplex as the convex hull of the origin together with a basis for $\mathbb{R}^n$. In other words, an n-simplex is a minimal convex body.

Lemma 15.6. *If $S_n = C(\{\mathbf{v}_1, \ldots, \mathbf{v}_{n+1}\})$ is an n-simplex then every vector in S_n has a unique representation as a convex combination of $\mathbf{v}_1, \ldots, \mathbf{v}_n$.*

Proof. We only need to prove uniqueness. Suppose that $\mathbf{v} = \sum_{i=1}^{n+1} s_i\mathbf{v}_i$ and $\mathbf{v} = \sum_{i=1}^{n+1} t_i\mathbf{v}_i$ are two representations of $\mathbf{v}$ as a convex combination of $\mathbf{v}_1, \ldots, \mathbf{v}_n$. Then $\mathbf{0} = \sum_{i=1}^{n+1}(s_i - t_i)\mathbf{v}_i = \sum_{i=1}^{n}(s_i - t_i)(\mathbf{v}_i - \mathbf{v}_{i+1}) + (\sum_{i=1}^{n+1}(s_i - t_i))\mathbf{v}_{n+1}$. But $\sum_{i=1}^{n+1}(s_i - t_i) = 0$ since $\sum_{i=1}^{n+1} s_i = \sum_{i=1}^{n+1} t_i = 1$, so $\sum_{i=1}^{n}(s_i - t_i)(\mathbf{v}_i - \mathbf{v}_{i+1}) = \mathbf{0}$. It follows that $s_i - t_i = 0$ since $\{\mathbf{v}_2 - \mathbf{v}_1, \ldots, \mathbf{v}_{n+1} - \mathbf{v}_n\}$ is linearly independent. □

It follows from this lemma that the extreme points of a simplex are precisely its vertices. So an n-simplex can be characterized as an n-dimensional convex body with exactly $n+1$ extreme points. We are now ready to express an arbitrary convex body and its extreme points in terms of simplices and their vertices.

Theorem 15.7. *Every k-dimensional convex body $B \subseteq \mathbb{R}^n$ is the union of simplices whose vertices are precisely the extreme points of B.*

Proof. We use induction on the dimension of B. A 1-dimensional convex body in $\mathbb{R}^n$ is simply a line segment, and a line segment is a 1-simplex. So assume inductively that the statement is true for convex bodies of dimension no larger than k and assume B has dimension $k+1$. Let $\mathbf{p}$ be an extreme point of B ($\mathbf{p}$ exists by Proposition 15.2) and let $\mathbf{x}$ be any other point of B. It suffices to show that $\mathbf{x}$ lies in a simplex in B such that $\mathbf{p}$ is a vertex of the simplex and every vertex is an extreme point of B.

Note that the intersection of B with the line through $\mathbf{x}$ and $\mathbf{p}$ is a line segment (since it is compact and convex) and $\mathbf{p}$ is an endpoint of this segment since it is an extreme point. Let $\mathbf{q}$ be the other endpoint of the segment. By Lemma 14.5 B is supported by a hyperplane L at $\mathbf{q}$, and $L \cap B$ is a convex body of dimension at most k. So by the inductive hypothesis $L \cap B$ is the union of simplicies whose vertices are extreme points of $L \cap B$. At least one of these simplices contains $\mathbf{q}$; call it S and let $\mathbf{v}_1, \ldots, \mathbf{v}_m$ be its vertices. We have $\mathbf{q} = \sum_{i=1}^{m} s_i \mathbf{v}_i$ for some collection of numbers $s_i \in [0,1]$ such that $\sum_{i=1}^{m} s_i = 1$. Moreover since $\mathbf{q}$, $\mathbf{x}$, and $\mathbf{p}$ all lie on the same line segment we have $\mathbf{x} = t\mathbf{q} + (1-t)\mathbf{p}$ for some $t \in [0,1]$ and hence $\mathbf{x} = \sum_{i=1}^{m} t s_i \mathbf{v}_i + (1-t)\mathbf{p}$ is a convex combination of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{p}$. Thus the convex hull of these $m+1$ vectors is a simplex containing $\mathbf{x}$ such that $\mathbf{p}$ is a vertex and all vertices are extreme points. This is what we wanted to prove. □

Corollary 15.8. *Every convex body is the convex hull of its extreme points.*

Proof. Let B be a convex body of any dimension. B is the union of simplices whose vertices are extreme points by the theorem, and every point in a simplex is a convex combination of its vertices. So every point in B is a convex combination of finitely many extreme points, and hence B is the convex hull of its extreme points by Proposition 15.4. □

16. Lecture 16 (9/28/11): Convex Polyhedra

We now revisit convex polyhedra and Theorem 13.8. The results of the last section allow us to prove the equivalence of the first two conditions appearing in that theorem:

Lemma 16.1. *A subset of $\mathbb{R}^n$ has finitely many extreme points if and only if it is the convex hull of a finite set.*

Proof. Let B denote the convex hull of a finite set $F = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$, so that B is the set of all convex combinations of vectors in F by Proposition 15.4. The proof of Proposition 15.4 can be easily adapted to show that if $\mathbf{x} \in B$ is not in F then it can be written as a nontrivial convex combination of two vectors in B which means it is not an extreme point. Thus every extreme point of B is an element of F, and hence there can be only finitely many.

The converse follows immediately from Corollary 15.8. □

Thus we need only prove that the first condition in Theorem 13.8 is equivalent to the third. One direction is straightforward:

Lemma 16.2. *If a convex body $B \subseteq \mathbb{R}^n$ is the intersection of finitely many half-spaces then it has only finitely many extreme points.*

Proof. We use induction on the dimension of B. In dimension 1 the only convex bodies are closed intervals in the real line, so every convex body both has only finitely many extreme points and is the finite intersection of half-spaces and thus there is nothing to prove. Assume that the statement is true for convex bodies of dimension no larger than m and let B be a convex body of dimension $m+1$. Let $\ell_1, \ldots, \ell_k$ be linear functionals such that B is the intersection of the half-spaces $\ell_i \leq c_i$ for $i = 1, \ldots, k$ and consider the hyperplanes $L_i = \ell^{-1}(c_i)$. Every extreme point of B lies is an extreme point of one of the sets $L_i \cap B$ and there are only finitely many extreme points of $L_i \cap B$ by the inductive hypothesis, so B can have only finitely many extreme points. □

17. Lecture 17 (9/30/11): Convex Polyhedra, Continued

To finish the proof of Theorem 13.8 it remains only to prove the converse to Lemma 16.2. We will need to introduce one tool in order to give the proof. Notice that without loss of generality we nay only consider convex bodies that contain the origin inside.

Definition 17.1. Let $B \subseteq \mathbb{R}^n$ be a convex body that contains the origin inside. The *dual body* is defined to be the set B^* of all linear functionals on $\mathbb{R}^n$ whose restriction to B is bounded by 1.

Exercise 17.2. Show that the dual of a convex body is a convex body in $(\mathbb{R}^n)^*$. Show that if B is the unit ball for some norm then B^* is the unit ball for the dual norm.

The main idea of our argument is that there is a correspondence between the extreme points of B^* and the faces of B.

Lemma 17.3. *If a convex body $B \subseteq \mathbb{R}^n$ has only finitely many extreme points then it is the intersection of finitely many half-spaces.*

Proof. Suppose B has only finitely many extreme points and consider the dual body B^*. By a simple variation on the proof of Proposition 15.1, B is equal to the intersection of half-spaces of the form $\ell \leq 1$ where ℓ is an extreme point of B^*. So to prove that B is the intersection of finitely many half-spaces it suffices to show that B^* has only finitely many extreme points.

Let $\ell \in B^*$ be an extreme point. Then $\ell(\mathbf{x}) \leq 1$ for every $\mathbf{x} \in B$ by definition of B^*, and we claim that there exists $\mathbf{x} \in B$ such that $\ell(\mathbf{x}) = 1$. Indeed, if $m = \sup_{\mathbf{x} \in B} \ell(\mathbf{x}) < 1$ then ℓ is an interior point of the line segment $\{\alpha\ell : 0 \leq \alpha \leq \frac{1}{m}\}$ which lies entirely in B^*, contradicting extremality of ℓ. So B lies in the half space $\ell \leq 1$ and B is supported by the hyperplane $\ell^{-1}(1)$ at some point $\mathbf{x}$.

Consider the set $C = \ell^{-1}(1) \cap B$; C is convex and its extreme points are all extreme points of B. Let $\mathbf{x}_1, \ldots, \mathbf{x}_m$ be the extreme points of B and assume $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are the extreme points of C, so that $\ell(\mathbf{x}_1) = \ldots = \ell(\mathbf{x}_k) = 1$ and $\ell(\mathbf{x}_i) < 1$ whenever $k+1 \leq i \leq m$. We shall prove that $\ell^{-1}(1)$ is the only supporting hyperplane which contains $\mathbf{x}_1, \ldots, \mathbf{x}_k$. If $\ell'^{-1}(1)$ is another one then the difference $\widetilde{\ell} = \ell - \ell'$ sends $\mathbf{x}_i$ to 0 for i from 1 to k and hence $(\ell + \alpha\widetilde{\ell})(\mathbf{x}_i) = 1$. There exists $\varepsilon > 0$ such that $\ell(\mathbf{x}_i) \leq 1 - \varepsilon$ for i from $k+1$ to m, and we have $(\ell + \alpha\widetilde{\ell})(\mathbf{x_i}) \leq 1 - \varepsilon + \alpha\widetilde{\ell}(\mathbf{x}_i)$. This can be made strictly smaller than 1 by choosing α small, so $\ell + \alpha\widetilde{\ell}$ is in B^* for all α sufficiently small. This contradicts the extremality of ℓ.

Thus we can associate to each extreme point of B^* a collection of extreme points of B in such a way that every collection of extreme points of B is associated to at most one extreme point of B^*. In other words, we have a one-to-one (injective) map from the extreme point set of B^* to the power set of the extreme point set of B. Since the latter is finite, the former must be as well. □

Thus a convex polyhedron can be regarded as either the convex hull of its vertices or the intersection of half-spaces determined by its faces. This helps to justify the naive intuition about the structure of polyhedra, but it does not quite capture the entire structure of a high-dimensional polyhedron. For example, polyhedra in $\mathbb{R}^3$ have edges as well as vertices and faces.

Definition 17.4. Let P be a n-dimensional convex polyhedron in R^n. Define the *k-dimensional faces* of B recursively as follows. An $n-1$-face of P is defined to be any $n-1$-dimensional convex polyhedron which is the intersection of P with a supporting hyperplane. If $2 \leq k \leq n$ then an $n-k$-face is defined to be a codimension 1 face of an $n-k+1$ dimensinoal face.

Alternatively, a k-dimensional face of B is a maximal convex subset $F \subseteq B$ such that every $\mathbf{x} \in F$ has the form $\mathbf{x} = \sum_{i=1}^{k} s_i \mathbf{x}_i$ where $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are extreme points, $s_i \in [0,1]$, $\sum_{i=1}^{k} s_i = 1$, and F contains no interior points of a higher dimensional face. Thus in $\mathbb{R}^3$ the 0-faces of a polyhedron in $\mathbb{R}^3$ are the vertices, the 1-faces are the edges, and the 2-faces are what one normally refers to as simply faces.

18. Lecture 18 (10/3/11): The Hahn-Banach Theorem

The backbone of our approach to the geometry of finite dimensional convex bodies was Lemma 14.1 which asserted that any linear functional on a subspace of $\mathbb{R}^n$ which is bounded by a convex, positively homogeneous function $\rho : \mathbb{R}^n \to \mathbb{R}$ can be extended to a linear functional on all of $\mathbb{R}^n$ which is still bounded by ρ. When we took ρ to be a norm on $\mathbb{R}^n$ we were able to deduce that the canonical embedding of a finite dimensional normed space into its double dual is an isometry, and when we took ρ to be the Minkowski functinoal of a convex body we inferred the existence of supporting hyperplanes.

We mentioned after the proof of Lemma 14.1 that it generalizes to infinite dimensions but that the proof in the most general cases leads into subtle issues of logic and set theory. We will now settle these issues once and for all.

Theorem 18.1 (The Hahn-Banach Theorem)**.** *Let V be a normed vector space, let $L \subseteq V$ be a linear subspace, and let f_0 be a linear functional on L such that $\|f_0\|^* \leq 1$. Then f_0 extends to a linear functional f on V such that $\|f\|^* \leq 1$.*

Before proving this theorem in complete generality we give a proof in the special case where V is separable (i.e. V contains a countable dense set). This gives us access to the theorem for many infinite dimensional spaces, such as $C[0,1]$ and ℓ^p, while avoiding the axiom of choice.

Proof of Theorem 18.1 when V is Separable. Let $\{\mathbf{v}_n\}$ be a countable dense subset of V. Define $V_k = L + \operatorname{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ and note that f_0 extends to a linear functional f_k on V_k of norm no larger than 1 by Lemma 14.1 (note that our proof of that lemma works whenever L has finite codimension in the whole space even if L itself is infinite dimensional). Thus f_0 extends to a linear functional f on $\bigcup_{k\in\mathbb{N}} V_k$, a dense subset of V, such that $\|f\|^* \leq 1$. The condition that $\|f\|^* \leq 1$ guarantees that f is uniformly continuous, and a uniformly continuous function defined on a dense subspace of a metric space X extends uniquely to a function on X. So f extends to a linear functional $\widetilde{f}$ on V, and $\widetilde{f}$ inherits the inequality $\left\|\widetilde{f}\right\|^* \leq 1$ from f by continuity. $\square$

To prove the theorem in the general case, we need some set theoretic machinery. Recall that a subset $A \subseteq S$ of a partially ordered set is said to be *totally ordered* if for every $x, y \in A$ we have that either $x \leq y$ or $y \leq x$. An *upper bound* for a totally ordered subset A is an element $u \in S$ such that $u \geq x$ for every $x \in A$. Finally an element $M \in S$ is *maximal* if there is no element $x \in S$ other than M such that

$x \geq M$. Note that a maximal element of S need not be an upper bound for S, and there can be many different maximal elements.

Lemma 18.2 (Zorn's Lemma)**.** *Let S be any set with the property that every totally ordered subset has an upper bound. Then S has at least one maximal element.*

It is a bit of a misnomer to call this result a lemma because it is equivalent to the axiom of choice and hence is independent of the other axioms of set theory. We will not prove Zorn's lemma or comment any further on its relationship to other set-theoretic constructions - we will simply use it as a tool for proving the Hahn-Banach theorem.

Proof of Theorem 18.1 for General V. Consider the set of all pairs (W, f) where W is a linear subspace of V containing L and f is a linear functional on W which extends f_0 and satisfies $\|f\|^* \leq 1$. Define a partial ordering on the set by declaring $(W_1, f_1) \geq (W_2, f_2)$ whenever W_2 is a subspace of W_1 and f_2 is the restriction of f_1 to W_2. A totally ordered subset is a chain (W_α, f_α) indexed by any partially ordered set such that W_α is a subspace of W_β and f_β is an extension of f_α whenever $\alpha \leq \beta$. Given any such chain, the pair (W, f) defined by $W = \bigcup_\alpha W_\alpha$ and $f(\mathbf{x}) = f_\alpha(\mathbf{x})$ if $\mathbf{x} \in W_\alpha$ is an upper bound for the chain, so the hypotheses of Zorn's lemma are satisfied. Hence there is a maximal pair $(\widetilde{W}, \widetilde{f})$. If $\widetilde{W} \neq V$ then let $\mathbf{v} \in V$ be any vector not in $\widetilde{W}$ and consider the subspace $\widetilde{W} + \mathbb{R}\mathbf{v}$. By Lemma 14.1 $\widetilde{f}$ extends to a linear functional $\widetilde{f_\mathbf{v}}$ on this subspace, so $(\widetilde{W} + \mathbb{R}\mathbf{v}, \widetilde{f_\mathbf{v}}) \geq (\widetilde{W}, \widetilde{f})$, contradicting the maximality of $(\widetilde{W}, \widetilde{f})$. Thus $W = V$ and $\widetilde{f}$ is the desired extension of f_0. $\square$

Remark 18.3*.* This proof works if $\|\cdot\|$ is merely convex and positively homogeneous so long as we replace the condition $\|f_0\|^* \leq 1$ with the condition $f_0(\mathbf{x}) \leq \|\mathbf{x}\|$ for all $\mathbf{x}$ (an equivalent condition when $\|\cdot\|$ really is a norm). As in the finite dimensional case we can use this generalization to infer the existence of supporting hyperplanes to many convex sets, although this is not enough to prove all of the results about convex bodies that we proved in the finite dimensional case because we freely used induction on dimension. Still, the Hahn-Banach theorem is a very powerful tool in infinite dimensional convex geometry.

Corollary 18.4. *If V is any normed space then the canonical embedding $i : V \to V^{**}$ is an isometry.*

Proof. The proof is exactly the same as in the finite dimensional case. $\square$

We conclude our discussion of the Hahn-Banach theorem with an amusing construction that helps to illustrate its power. It is often useful in analysis to develop notions of convergence of infinite series which apply to sequences which do not converge in the usual sense. For example, a sequence $\{x_n\}$ of real numbers is said to converge in the *Cesaro* sense of the sequence of averages $\{\frac{1}{n}\sum_{k=1}^n x_k\}$ converges; thus the sequence $\{1, -1, 1, -1, \ldots\}$ converges in the Cesaro sense but not the usual sense. It is natural to ask how broadly one can generalize the notion of convergence in this manner. Any generalization $\lim'$ of the usual notion of limit should enjoy the following properties:

- $\lim'\{ax_n + by_n\} = a\lim'\{x_n\} + b\lim'\{y_n\}$
- $\lim'\{x_n\} = \lim\{x_n\}$ if $\{x_n\}$ converges in the usual sense
- $\lim' S\{x_n\} = \lim'\{x_n\}$ where $S(\{x_1, x_2, \ldots\}) = \{x_2, x_3, \ldots\}$

It turns out that a generalized limit can be defined for any bounded sequence. We will achieve this by applying the Hahn-Banach theorem to a certain linear functional on a subspace of the Banach space ℓ^∞ of bounded sequences.

Define L to be the subspace $\{S\mathbf{x}-\mathbf{x} : \mathbf{x} \in \ell^\infty\}$ of ℓ^∞ where S is the shift operator above. Let $L' = L + \mathbb{R}$ and define a linear functional f_0 on L' by $f_0(\mathbf{x}) = 0$ for $\mathbf{x} \in L$ and $f_0(1) = 1$. It is easy to see that $\|f_0\|^* \leq 1$, so by the Hahn-Banach theorem it extends to a linear functional f on ℓ^∞ such that $\|f\|^* \leq 1$. It is not difficult to show that $f(\{x_n\}) = \lim \mathbf{x_n}$ when the limit exists, and we have $f(S\mathbf{x}) - f(\mathbf{x}) = f(S\mathbf{x} - \mathbf{x}) = f_0(S\mathbf{x} - \mathbf{x}) = 0$. So f has all of the properties demanded of a generalized limit. The value of f on a sequence in ℓ^∞ is often called the *Banach limit* of the sequence.

19. Lecture 19 (10/5/11): Hilbert Spaces

19.1. **Definitions.** Having developed a substantial amount of geometric machinery relevant to the theory of Banach spaces, we turn our attention to the theory of Hilbert spaces. Hilbert spaces are infinite dimensional counterparts of the Euclidean spaces that we studied in lectures 6 and 7: a Hilbert space is simply a Banach space whose norm is induced by an inner product in the sense of Definition 6.1. As in the finite dimensional case, Hilbert spaces have a much richer and more accessible geometric structure than general Banach spaces. It is not uncommon in analysis and geometry to introduce Hilbert spaces into contexts where they do not obviously belong just to gain access to this extra structure.

Up until this point we have assumed that all vector spaces were defined over the real numbers, but many of our results remain true over the complex numbers (and with similar proofs). For reasons that will become clear in the next several lectures, it is useful to develop the theory of Hilbert spaces over the complex numbers as well rather than just the real numbers.

Definition 19.1. Let V be a vector space over $\mathbb{C}$. A (complex) *inner product* on V is a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ which satisfies the following axioms:

- Linearity in the first variable:
$$\langle a\mathbf{x} + a'\mathbf{x}', \mathbf{y} \rangle = a \langle \mathbf{x}, \mathbf{y} \rangle + a' \langle \mathbf{x}', \mathbf{y} \rangle$$
- Skew-symmetry: $\langle \mathbf{x}, \mathbf{y} \rangle = \overline{\langle \mathbf{y}, \mathbf{x} \rangle}$
- Positive Definiteness: $\sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ is a nonnegative real number for every $\mathbf{x}$, and it is 0 if and only if $\mathbf{x} = \mathbf{0}$

Example 19.2. The *standard complex inner product* on $\mathbb{C}^n$ is defined by the formula
$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} x_i \overline{y_i}$$

Note that linearity in the first variable together with skew-symmetry implies that a complex inner product is conjugate-linear in the second variable:
$$\langle \mathbf{x}, a\mathbf{y} + a'\mathbf{y}' \rangle = \overline{a} \langle \mathbf{x}, \mathbf{y} \rangle + \overline{a'} \langle \mathbf{x}, \mathbf{y} \rangle$$
A function $V \times V \to V \to \mathbb{C}$ which is linear in the first variable and conjugate linear in the second variable is sometimes called *sesquilinear*.

Exercise 19.3. Show that if V is a complex vector space equipped with a complex inner product $\langle \cdot, \cdot \rangle$ then the function $\|\cdot\| : V \to \mathbb{R}$ given by $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ is a norm on V.

Thus every complex inner product space comes equipped with a norm induced by the inner product.

Definition 19.4. A complex inner product space $(V, \langle \cdot, \cdot \rangle)$ is a *Hilbert space* if it is complete with respect to the norm $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$.

While complex Hilbert spaces exhibit some features not present for real Hilbert spaces, there are ways to pass back and forth between them.

Let $(V, \langle \cdot, \cdot \rangle_{\mathbb{C}})$ be a complex inner product space. V has the structure of a real vector space, where a real number a acts on $\mathbf{v} \in V$ as $(a + 0i)\mathbf{v}$, and the formula $\langle \mathbf{v}, \mathbf{w} \rangle_{\mathbb{R}} = \mathrm{Re}(\langle \mathbf{v}, \mathbf{w} \rangle_{\mathbb{C}})$ defines a real inner product on V (i.e. an inner product in the sense of Definition 6.1). For example, if $\mathbb{C}^n$ is equipped with the standard complex inner product then the real inner product space resulting from this construction corresponds to $\mathbb{R}^{2n}$ with the standard real inner product.

Now suppose that $(V, \langle \cdot, \cdot \rangle_{\mathbb{R}})$ is a real inner product space. Equip $V \times V$ with the structure of a complex vector space by allowing i to act as $i(\mathbf{x}, \mathbf{y}) = (-\mathbf{y}, \mathbf{x})$; intuitively, we identify $(\mathbf{x}, \mathbf{y})$ with "$\mathbf{x} + i\mathbf{y}$". Define a complex inner product on $V \times V$ by the formula:

$$\langle (\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2) \rangle_{\mathbb{C}} = (\langle \mathbf{x}_1, \mathbf{y}_1 \rangle_{\mathbb{R}} + \langle \mathbf{x}_2, \mathbf{y}_2 \rangle_{\mathbb{R}}) + i(\langle \mathbf{x}_2, \mathbf{y}_1 \rangle_{\mathbb{R}} - \langle \mathbf{x}_1, \mathbf{y}_2 \rangle_{\mathbb{R}}$$

The complex inner product space $(V \times V, \langle \cdot, \cdot \rangle_{\mathbb{C}})$ is called the *complexification* of V; for example, the complexification of $\mathbb{R}^n$ with the standard real inner product is $\mathbb{C}^n$ with the standard complex inner product.

These two constructions can be related in the following way. Let $(V, \langle \cdot, \cdot \rangle_{\mathbb{R}})$ be a real inner product space and suppose there exists a linear operator $J : V \to V$ with the property that $J^2 = -\mathrm{id}_V$ and $\langle J\mathbf{v}, J\mathbf{w} \rangle_{\mathbb{R}} = \langle \mathbf{v}, \mathbf{w} \rangle_{\mathbb{R}}$. For example, such an operator exists on a finite dimensional real inner product space V if and only if V has dimension $2n$ for some integer n. J extends to an operator on the complexification of V, and the complexification on V decomposes as the orthogonal direct sum of the i-eigenspace and $-i$-eigenspace of J. The i-eigenspace is a complex inner product space (its dimension is n if V has dimension $2n$) naturally associated to V via J. The operator J is often referred to as a *complex structure* on V for this reason.

We will not need to dwell too much on the relationship between real inner products and complex inner products, but it is worth noting that the complex numbers carry some extra subtleties that warrant caution, even in finite dimensions. So our policy will be to implicitly assume that all inner product spaces are defined over the complex numbers unless otherwise specified; the proofs and examples over the real numbers are never more difficult than their complex counterparts.

19.2. **Orthonormal Bases.** In our discussion of finite dimensional inner product spaces, we made essential use of the notion of an orthonormal basis, and we would like to have access to that tool in the infinite dimensional case. Recall that a set of vectors $\{\mathbf{e}_i\}$ in an inner product space is *orthonormal* if $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = 0$ when $i \neq j$ and $\|\mathbf{e}_i\| = 1$ for every i. We defined an orthonormal basis of a finite dimensional vector space to be an orthonormal set which spans the whole space (an orthonormal set is automatically linearly independent).

Unfortunately we need to allow the word "basis" to be a bit more flexible in infinite dimensional spaces in order to capture the power of orthonormal sets. Instead of referring simply to a linearly independent spanning set, one often uses the word "basis" to describe a linearly independent set which can approximate any vector in an appropriate sense. We will indicate some of the subtleties that arise in this context and then give the definition of basis which is most appropriate for Hilbert spaces.

The most general definition of a basis for an infinite dimensional vector space V is also the most natural one from the point of view of finite dimensional linear algebra. One simply defines a set $\mathfrak{B} \subseteq V$ to be a basis for V if it is linearly independent and every vector in V is a finite linear combination of vectors in $\mathfrak{B}$. Such a set $\mathcal{B}$ is often called a *Hammel basis*. This is generally not the appropriate definition in functional analysis, however, because infinite dimensional vector spaces are usually equipped with extra geometric structure (such as a norm) which is not reflected by a choice of Hammel basis. Additionally, Hammel bases are often far too large to be useful: no infinite dimensional Banach space has a countable Hammel basis, for example (this is not trivial!).

One can make better definitions for a normed space $(V, \|\cdot\|)$. One may define a set $\mathfrak{B} \subseteq V$ to be a basis if it is linearly independent and if its linear span is dense in V. For example, the set $\{1, x, x^2, \ldots\}$ is a basis for $C[0,1]$ in this sense because its linear span is the set of all polynomial functions on $[0,1]$ and this is well-known to be a dense set. This definition is useful but not entirely satisfactory because it is not obvious how to uniformly approximate any given continuous function by polynomial functions.

The strongest and most useful definition of a basis for $(V, \|\cdot\|)$ is a linearly independent set $\mathfrak{B} = \{\mathbf{v}_n\}_{n \in \mathbb{N}}$ such that every $\mathbf{v} \in V$ is the norm limit of some infinite series $\sum_{n=1}^{\infty} a_n \mathbf{v}_n$. Note that $\{1, x, x^2, \ldots\}$ is not a basis of $C[0,1]$ in this sense because most continuous functions cannot be written as the uniform limit of a power series. Note that it is not crucial that the set $\mathfrak{B}$ is countable so long as one generalizes the usual notion of convergence from sequences to uncountable sets. This is achieved by the notion of a *net*, but we will be able to avoid this language by restricting our attention to spaces which are separable (i.e. they contain a countable dense set).

These issues are subtle, but fortunately we do not need to worry about them too much when we discuss orthonormal bases. Here is the definition:

Definition 19.5. An *orthonormal basis* for a Hilbert space $(H, \langle \cdot, \cdot \rangle)$ is a set $\mathfrak{B}$ with the following properties:

- $\|\mathbf{v}\| = 1$ for every $\mathbf{v} \in \mathfrak{B}$
- $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = 0$ for every pair of distinct vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathfrak{B}$
- The linear span of $\mathfrak{B}$ is dense in H.

We only assume that the linear span of an orthonormal basis is dense, but in fact the stronger statement that every vector in H is the limit of an infinite series determined by vectors in $\mathfrak{B}$ automatically holds. If $\mathfrak{B} = \{\mathbf{e}_n\}_{n \in \mathbb{N}}$ is a countable orthonormal basis then for any $\mathbf{v} \in H$ we can express $\mathbf{v}$ as the norm limit of the infinite series:

$$\sum_{n=1}^{\infty} \langle \mathbf{v}, \mathbf{e}_n \rangle \, \mathbf{e}_n$$

In order to prove that this infinite series converges to $\mathbf{v}$ we will need to develop some more concepts related to the geometric structure of Hilbert spaces; these matters will be revisited in future lectures. If one is willing to make sense of infinite series with uncountably many terms (only countably many of those not equal to zero) then a similar statement holds for uncountable orthonormal bases as well, but as above we will avoid uncountable orthonormal bases.

20. Lecture 20 (10/7/11): Examples of Hilbert Spaces

So far we have only encountered finite dimensional examples of Hilbert spaces. In particular, we have seen that $\mathbb{R}^n$ equipped with the standard Euclidean inner product is a real Hilbert space, and similarly $\mathbb{C}^n$ with the standard complex inner product is a complex Hilbert space (to check completeness, note that the norm induced by the standard complex inner product agrees with the standard Euclidean norm on $\mathbb{R}^{2n} \cong \mathbb{C}^n$). In this section we will give two important infinite dimensional examples of Hilbert spaces and construct orthonormal bases on them. We will present our examples as complex Hilbert spaces, but there are obvious analogues of both examples over the real numbers.

The first infinite dimensional example is the space $\ell^2(\mathbb{C})$ of all square summable sequences of complex numbers. This space comes equipped with the complex inner product:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{n=1}^{\infty} x_n \overline{y_n}$$

For each n define $\mathbf{e}_n$ to be the sequence whose nth entry is 1 and whose kth entry is 0 if $k \neq n$. It is clear that the set $\{\mathbf{e}_n\}_{n \in \mathbb{N}}$ is an orthonormal set in $\ell^2(\mathbb{C})$; we shall argue that it is in fact an orthonormal basis. Given any element $\mathbf{x} = \{x_n\}$ in $\ell^2(\mathbb{C})$, let $\mathbf{x}^N$ denote the truncation of $\mathbf{x}$ to the first N entries, so that $\mathbf{x}^N = (x_1, \ldots, x_N, 0, 0, \ldots)$. Then $\|\mathbf{x} - \mathbf{x}_N\| = \sqrt{\sum_{n=N}^{\infty} |x_n|^2}$, and this converges to 0 as N tends to infinity since it is the tail of a convergent series. Since $\mathbf{x}^N$ is in the linear span of $\{\mathbf{e}_n\}$, we have shown that the linear span of $\{\mathbf{e}_n\}$ is dense in $\ell^2(\mathbb{C})$, as desired. In fact, we have shown directly that $\{\mathbf{e}_n\}$ is a basis for $\ell^2(\mathbb{C})$ in the strong sense that every element of $\ell^2(\mathbb{C})$ is the limit of an infinite series determined by the $\mathbf{e}_n$'s.

A more interesting example of an infinite dimensional Hilbert space is obtained by equipping the complex vector space $C_{\mathbb{C}}[0,1]$ of all continuous complex valued functions on $[0,1]$ with a complex inner product defined via Riemann integration:

$$\langle f, g \rangle = \int_0^1 f(x)\overline{g(x)}dx$$

This inner product is called the L^2 *inner product*, and $C_{\mathbb{C}}[0,1]$ equipped with the L^2 inner product is a complex inner product space. This inner product space is not complete, but its completion $L^2[0,1]$ is a Hilbert space. Intuitively one might expect that $L^2[0,1]$ is simply the space of all complex valued functions f on $[0,1]$ which are square integrable in the sense that $\int_0^1 |f(x)|^2 dx < \infty$ since any such function can be approximated by continuous functions in the L^2 norm, but this intuition is problematic for two rather subtle reasons.

The first reason is that the L^2 norm on $C[0,1]$ does not extend to a norm on the space of square integrable functions. The problem is that $\|\cdot\|$ is indefinite: the function

$$f(x) = \begin{cases} 0 & x \in [0,1) \\ 1 & x = 1 \end{cases}$$

is nonzero and square integrable, but $\|f\| = \int_0^1 |f(x)|^2 dx = 0$. The solution is to define an equivalence relation on the space of integrable functions with the property that the integral of an integrable function is 0 if and only if the function is equivalent to the zero function. To define this equivalence relation we need the following basic notion from the theory of measures:

Definition 20.1. A subset $A \subseteq [0,1]$ is said to be a *null set* if for every $\varepsilon > 0$ there exists a countable family of open intervals $\{I_n\}$ such that $A \subseteq \bigcup_n I_n$ and $\sum_n |I_n| < \varepsilon$ where $|I_n|$ denotes the length of the interval I_n.

Any set which contains only one point is null and the countable union of null sets is again a null set, so any countable set is null. There are examples of uncountable null sets, such as the Cantor set, but we will not need to worry about how exotic null sets can be. The key property of null sets which we will need is that the integral of the characteristic function of a subset of $[0,1]$ is 0 if and only if the set is null (assuming the characteristic function of the set is integrable)

Definition 20.2. Define an equivalence relation $\sim$ on the set of all complex valued functions on $[0,1]$ by declaring $f \sim g$ if the set of all $x \in [0,1]$ with the property that $f(x) \neq g(x)$ is a null set

Thus for integrable functions f and g we have that $f \sim g$ if and only if $\|f - g\| = 0$. So $\|\cdot\|$ gives rise to a well-defined norm on the set of all equivalence classes of integrable functions on $[0,1]$, and we might hope that this provides a model for $L^2[0,1]$. However, this is still not quite correct if we insist on using the Riemann integral: there exists a sequence of Riemann square integrable functions whose limit is not equivalent to any Riemann square integrable function. The solution to this dilemma is to give an alternative definition of integration - via the so-called *Lebesge integral* - which agrees with the Riemann integral in the context of continuous functions but exhibits better behavior with respect to limits. Indeed, a correct model for $L^2[0,1]$ is given by the space of equivalence classes of Lebesgue square integrable functions on $[0,1]$ (abusing notation, it is customary to express elements of $L^2[0,1]$ as functions even though they are really equivalence classes of functions). It would take us too far afield to give a completely precise definition of the Lebesgue integral, but the reader's intuition will not be too severely harmed by thinking of an element of $L^2[0,1]$ as an equivalence class of functions represented by a Riemann square integrable function on $[0,1]$. Every such equivalence class is an element of $L^2[0,1]$, but some exotic elements of $L^2[0,1]$ cannot be obtained in this way.

With this in mind, let us construct some examples of orthonormal bases for $L^2[0,1]$. The first example is given by the funtions $\chi_n(x) = e^{2\pi ix}$; let us show that the set $\{\chi_n\}_{n\in\mathbb{Z}}$ is orthonormal. For $n \neq m$ we have:

$$\begin{aligned}\langle \chi_n, \chi_m\rangle &= \int_0^1 \chi_n(x)\overline{\chi_m(x)}dx \\ &= \int_0^1 e^{2\pi i(n-m)x}dx \\ &= \frac{1}{2\pi i(n-m)}e^{2\pi i(n-m)x}|_0^1 = 0\end{aligned}$$

Thus χ_n is orthogonal to χ_m for $n \neq m$. Similarly, we have:

$$\begin{aligned}\langle \chi_n, \chi_n\rangle &= \int_0^1 \chi_n(x)\overline{\chi_n(x)}dx \\ &= \int_0^1 1dx = 1\end{aligned}$$

Thus $\|\chi_n\| = 1$, as desired. To prove that the orthonormal set $\{\chi_n\}$ is an orthonormal basis, we would need to show that the linear span of this set is dense in $L^2[0,1]$. This involves some somewhat sophisticated tools in analysis, and we postpone this discussion until the next lecture. Granting this assertion, we have that every $f \in L^2[0,1]$ is the L^2 limit of an infinite series $\sum_{n\in Z} c_n e^{2\pi inx}$. This is often called the *Fourier series* representation of f. The theory of Fourier series is a very old subject in analysis which in fact motivated the development of the abstract theory of Hilbert spaces. The original formulation of the theory used the orthonormal basis $\{\cos(2\pi nx), \sin(2\pi nx)\}_{n\in\mathbb{N}}$ for the real Hilbert space of square integrable real valued functions on $[0,1]$; the two formulations are related by the famous equation $e^{2\pi inx} = \cos(2\pi nx) + i\sin(2\pi nx)$.

Note that the Fourier series of a function a priori converges only in L^2; the problem of determining when the Fourier series converges in other senses (e.g. pointwise or uniformly) leads to extremely deep issues in classical analysis. A simple argument shows that the Fourier series of a continuously differentiable function converges uniformly, but it is very difficult to improve this result in any substantial way. To give an idea of the difficulty of this problem, it was not until 1966 that Lennart Carleson showed that the Fourier series of a function in $L^2[0,1]$ converges almost everywhere (i.e. it converges pointwise on the complement of a null set), and his proof is considered to be one of the crown jewels of classical analysis. Carleson's result is sharp: given any null set $A \subseteq [0,1]$ there is a continuous function whose Fourier series does not converge at any point in A.

There are many other examples of orthonormal bases for $L^2[0,1]$. Many difficult problems involving functions on $[0,1]$ can be solved by choosing a well-adapted orthonormal basis for $L^2[0,1]$.

Now we give an example of an orthonormal basis where the proof of completeness is straightforward and does not require sophisticated approximation results as in the case of the exponentials, Define a set of piecewise constant functions $f_{n,k}$, $n \in \mathbb{N}, k = 0, 1 \ldots, 2^k - 1$ by the formula:

$$f_{n,k}(x) = \begin{cases} 2^{(n-1)/2} & x \in (\frac{2k}{2^n}, \frac{2k+1}{2^n}], \\ -2^{(n-1)/2} & x \in (\frac{2k+1}{2^n}, \frac{2k+2}{2^n}] \\ 0, & \text{elsewhere} \end{cases}$$

Add constant function equal to one to this set. It is straightforward to check that this extended set is orthonormal. In order to show that its linear span is dense in $L^2[0,1]$ notice that linear combinations of the functions $f_{n,k}$ for $n = 1, \dots N$ and the constant are exactly all functions constant on the intervals $(\frac{k}{2^N}, \frac{k+1}{2^N}]$, $k = 0, \dots 2^N - 1$. This is easily proved by induction in N. But those functions uniformly approximate all continuous function and hence their union for all N is dense in $L^2[0,1]$, since any uniformly converging sequence converges in L^2.

21. Lecture 21 (10/17/11): Existence of Orthonormal Bases

One tool for establishing that various subsets of function spaces are dense is a generalization of the classical result due to Weierstrass that the space of polynomial functions is dense in $C[0,1]$ (equipped with the uniform norm). This tool requires a little bit of new language:

Definition 21.1. Let X be a compact metric space and let $C(X, \mathbb{R})$ (respectively, $C(X, \mathbb{C})$) denote the space of continuous real valued (respectively, complex valued) functions on X equipped with the uniform norm.

- A *subalgebra* of $C(X, \mathbb{R})$ is a linear subspace which is closed under pointwise multiplication of functions.
- A $*$-*subalgebra* of $C(X, \mathbb{C})$ is a linear subspace whcih is closed under pointwise multipliation and complex conjugation of functions.

A subalgebra A of $C(X, \mathbb{R})$ (respectively, $C(X, \mathbb{C})$) is said to *separate points* if for every pair of distinct points $p, q \in X$ there exists $f \in A$ such that $f(p) = 0$ and $f(q) \neq 0$.

Thus the set of all polynomial functions is a subalgebra of $C([0,1], \mathbb{R})$ and the set of trigonometric polynomials , i.e. finite linear combinations of $\{e^{2\pi i n x}\}_{n \in \mathbb{Z}}$ is a $*$-subalgebra of $C([0,1], \mathbb{C})$.

Theorem 21.2 (Stone-Weierstrass)**.** *Let X be a compact metric space.*

- *Any subalgebra of $C(X, \mathbb{R})$ which contains a nonzero constant function and separates points is dense.*
- *Any $*$-subalgebra of $C(X, \mathbb{C})$ which contains a nonzero constant function and separates points is dense.*

The set $P_{\mathbb{R}}[0,1]$ of all polynomial functions with real coefficients contains the constant functions (they are polynomials of degree 0) and separates points: given any pair $p, q \in [0,1]$ of distinct points the linear function $f(x) = x - p$ separates p and q. Thus the Stone-Weierstrass theorem immediately implies that $P_{\mathbb{R}}[0,1]$ is dense in $C([0,1], \mathbb{R})$. However, it is not true that the set $P_{\mathbb{C}}[0,1]$ of all polynomial functions with complex coefficients is dense in $C([0,1], \mathbb{C})$; the Stone-Weierstrass theorem doesn't apply since $P_{\mathbb{C}}[0,1]$ is not closed under complex conjugation.

We can also use the Stone-Weierstrass theorem to prove that the functions $\chi_n(x) = e^{2\pi i n x}$ have dense linear span in $L^2[0,1]$, but it takes a bit of additional work. There are two obstacles: first, the χ_n's do not separate points in $[0,1]$ because $\chi_n(0) = \chi_n(1)$ for every n; and second, the Stone-Weierstrass theorem involves uniform convergence of functions rather than L^2 convergence. Neither obstacle is particularly difficult to overcome, but these subtleties should not be ignored.

Corollary 21.3. *The set $S = \{\chi_n\}_{n \in Z}$ is an orthonormal basis for $L^2[0,1]$.*

Proof. We have already established that S is an orthonormal set, so it suffices to show that $\mathrm{Span}\{S\}$ is dense in $L^2[0,1]$. By the definition of $L^2[0,1]$, every element of $L^2[0,1]$ is the limit in the L^2 norm of a sequence of continuous functions, so it suffices to show that $\mathrm{Span}\{S\}$ is dense in $C([0,1],\mathbb{C})$ with respect to the L^2 norm.

Let X denote the circle of circumference 1 with its standard pathwise metric; we can consider X to be the unit interval $[0,1]$ with its endpoints identified. Thus a continuous function on X is simply a countinuous function f on $[0,1]$ such that $f(0) = f(1)$. In particular we can regard each χ_n as a function on X since $e^0 = e^{2\pi i n} = 1$.

Now, the set S is closed under multiplication and complex conjugation, so its span (the set of all finite linear combinations for functions in S) is a $*$-subalgebra of $C(X,\mathbb{C})$ which contains the nonzero constant function $\chi_0(x) = 1$. Moreover $\mathrm{Span}\{S\}$ separates points in X: for any pair of distinct points $p, q \in X$ we have that $\chi_1(p) \neq \chi_1(q)$ and thus the function $f(x) = \chi_1(x) - \chi_1(p)\chi_0(x)$ is an element of S which satisfies $f(p) = 0$ and $f(q) \neq 0$. Thus the Stone-Weierstrass theorem implies that $\mathrm{Span}\{S\}$ is dense in $C(X,\mathbb{C})$ with the uniform norm.

Take any $\varphi \in C(X,\mathbb{C})$ and let φ_n be a sequence in $\mathrm{Span}\{S\}$ which converges uniformly to φ, so that $|\varphi - \varphi_n|$ converges uniformly to 0. It is a well known fact from Riemann integration theory that we can intergange the order of integration and *uniform* limits, so we have $\lim_n \int_0^1 |\varphi - \varphi_n|^2 = \int_0^1 \lim_n |\varphi - \varphi_n|^2 = 0$. Thus φ_n converges to φ with respect to the L^2 norm, and hence $\mathrm{Span}\{S\}$ is L^2-dense in $C(X,\mathbb{C})$. Finally, any function $f \in C([0,1],\mathbb{C})$ can be written as the L^2-limit of the sequence $f_n \in C(X,\mathbb{C})$ obtained by setting $f_n(x) = f(x)$ on $x \in [0, 1-\frac{1}{n}]$ and by defining $f_n(x)$ to be the linear function which satisfies $f_n(1-\frac{1}{n}) = f(1-\frac{1}{n})$ and $f_n(1) = f(0)$ on $[1-\frac{1}{n}, 1]$. Thus $\mathrm{Span}\{S\}$ is L^2-dense in $C([0,1],\mathbb{C})$ and hence in $L^2[0,1]$. □

In general the Stone-Weierstrass is very useful when one wishes to prove that a specific well-chosen orthonormal subset of $C[0,1]$ (with respect to the L^2 or a similar inner product) is in fact an orthonormal basis for $L^2[0,1]$. This helps us to construct orthonormal bases for Hilbert spaces presented as spaces of functions, but it is useful to give an abstract argument which verifies that every Hilbert space has an orthonormal basis regardless of how it is presented. For arbitrary Hilbert spaces there are set theoretic issues to worry about, but these issues can be avoided when considering separable Hilbert spaces.

Proposition 21.4. *Every separable Hilbert space has a countable orthonormal basis.*

Proof. Let H be an infinite dimensional separable Hilbert space (the finite dimensional case is Proposition 7.3) and let $\{\mathbf{v}_n\}$ be a countable dense subset of H. Define $H_n = \mathrm{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and observe that $\bigcup_n H_n$ is dense in H. Clearly H_1 has an orthonormal basis (take any unit vector), and in the proof of Proposition 7.3 we showed that an orthonormal basis of any finite dimensional proper subspace of an inner product space can be extended to an orthonormal basis for a subspace of one dimension higher, so by induction there is a countable orthonormal set whose span is $\bigcup_n H_n$. This set is an orthonormal basis for H since $\bigcup_n H_n$ is dense in H. □

Remark 21.5. Note that we did not use completeness in the proof of this proposition; indeed, we have shown that any separable inner product space has an orthonormal basis.

22. Lecture 22: The Geometry of Hilbert Spaces

We saw in earlier lectures that the structure of a Banach space is most readily accessed via the tools of convex geometry. Hilbert spaces bring this interplay between algebra and geometry into even sharper focus: while one has access to linear inequalities in an arbitrary Banach space, a Hilbert space provides an environment for quadratic algebra which leads to more precise estimates and more elegant geometry. With some concrete examples in hand, our aim in this lecture is to explore the extra structure available in Hilbert space theory. Our main tool will be the existence of countable orthonormal bases discussed in the previous lecture, and because of this all of our proofs will only work for separable Hilbert spaces. We note, however, that most of the results (and their proofs) generalize to the non-separable case.

Our first concrete goal is to show that an orthonormal basis serves as a sort of infinite dimensional coordinate system for a Hilbert space. More precisely, we will fullfill the promise made in lecture 19 to prove that any vector $\mathbf{v}$ in a separable Hilbert space H with an orthonormal basis $\{\mathbf{e}_n\}_{n\in\mathbb{N}}$ can be expressed as the limit of the infinite series

$$\sum_{n=1}^{\infty} \langle \mathbf{v}, \mathbf{e}_n \rangle \, \mathbf{e}_n \tag{22.1}$$

We will achieve this using two lemmas of independent interest.

Lemma 22.1. *Let H be a Hilbert space and let $L \subseteq H$ be a finite dimensional subspace with an orthonormal basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$. Given $\mathbf{v} \in H$, if $\mathbf{v}_L = \sum_{j=1}^{n} \langle \mathbf{v}, \mathbf{e}_j \rangle \, \mathbf{e}_j$ then $\mathbf{v} - \mathbf{v}_L$ is orthogonal to every vector in L.*

Proof. It suffices to show that $\mathbf{v} - \mathbf{v}_L$ is orthogonal to $\mathbf{e}_k$ for $1 \le k \le n$. This is just a simple calculation using the fact that $\langle \mathbf{e}_j, \mathbf{e}_k \rangle$ is 0 if $j \neq k$ and is 1 if $j = k$.

$$\begin{aligned} \langle \mathbf{v} - \mathbf{v}_L, \mathbf{e}_k \rangle &= \langle \mathbf{v}, \mathbf{e}_k \rangle - \left\langle \sum_{j=1}^{n} \langle \mathbf{v}, \mathbf{e}_j \rangle \, \mathbf{e}_j, \mathbf{e}_k \right\rangle \\ &= \langle \mathbf{v}, \mathbf{e}_k \rangle - \sum_{j=1}^{n} \langle \mathbf{v}, \mathbf{e}_j \rangle \langle \mathbf{e}_j, \mathbf{e}_k \rangle \\ &= \langle \mathbf{v}, \mathbf{e}_k \rangle - \langle \mathbf{v}, \mathbf{e}_k \rangle = 0 \end{aligned}$$

□

It is natural to think of the vector $\mathbf{v}_L$ in the lemma as an "approximation" of $\mathbf{v}$ using vectors in L, though at first glance $\mathbf{v}_L$ appears to depend on the orthonormal basis for L chosen. The next lemma verifies this intuition and gives an intrinsic characterization of $\mathbf{v}_L$.

Lemma 22.2. *Let H be a Hilbert space and let $L \subseteq H$ be a finite dimensional subspace with an orthonormal basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$. Given $\mathbf{v} \in H$, define $\mathbf{v}_L = \sum_{j=1}^{n} \langle \mathbf{v}, \mathbf{e}_j \rangle \, \mathbf{e}_j$ then $\mathbf{v} - \mathbf{v}_L$ as in Lemma 22.1. Then $\mathbf{v}_L$ is the unique vector in L which satisfies $\|\mathbf{v} - \mathbf{v}_L\| = \inf_{\mathbf{w} \in L} \|\mathbf{v} - \mathbf{w}\|$.*

Proof. First we show that $\|\mathbf{v}-\mathbf{v}_L\| \le \|\mathbf{v}-\mathbf{w}\|$ for every $\mathbf{w} \in L$. Write $\mathbf{w} = \sum_{k=1}^n a_k \mathbf{e}_k$ and calculate:

$$\begin{aligned}
\|\mathbf{v}-\mathbf{w}\|^2 &= \langle \mathbf{v}-\mathbf{w}, \mathbf{v}-\mathbf{w}\rangle \\
&= \langle \mathbf{v}, \mathbf{v}\rangle - 2\mathrm{Re}\left(\langle \mathbf{v}, \mathbf{w}\rangle\right) + \langle \mathbf{w}, \mathbf{w}\rangle \\
&= \langle \mathbf{v}, \mathbf{v}\rangle - 2\mathrm{Re}\left(\sum_{k=1}^n \overline{a_k}\langle \mathbf{v}, \mathbf{e}_k\rangle\right) + \sum_{j,k} a_j \overline{a_k}\langle \mathbf{e}_j, \mathbf{e}_k\rangle
\end{aligned}$$

Replace $\mathbf{v}$ with $\mathbf{v}_L + (\mathbf{v}-\mathbf{v}_L)$ and use the fact that $\mathbf{v}-\mathbf{v}_L$ is orthogonal to L by Lemma 22.1 to obtain:

$$\|\mathbf{v}-\mathbf{w}\|^2 = \langle \mathbf{v}, \mathbf{v}\rangle - 2\mathrm{Re}\left(\sum_{k=1}^n \overline{a_k}\langle \mathbf{v}_L, \mathbf{e}_k\rangle\right) + \sum_{k=1}^n |a_k|^2$$

To finish the argument, the idea is to "complete the square" in the right-hand side of this equation by adding and subtracting $\langle \mathbf{v}_L, \mathbf{v}_L\rangle$. First, let us compute this expression explicitly:

$$\begin{aligned}
\langle \mathbf{v}_L, \mathbf{v}_L\rangle &= \left\langle \sum_{j=1}^n \langle \mathbf{v}, \mathbf{e}_j\rangle \mathbf{e}_j, \sum_{k=1}^n \langle \mathbf{v}, \mathbf{e}_k\rangle \mathbf{e}_k \right\rangle \\
&= \sum_{j,k} \langle \mathbf{v}, \mathbf{e}_j\rangle \overline{\langle \mathbf{v}, \mathbf{e}_k\rangle}\langle \mathbf{e}_j, \mathbf{e}_k\rangle \\
&= \sum_{k=1}^n |\langle \mathbf{v}, \mathbf{e}_k\rangle|^2
\end{aligned}$$

This yields:

$$\begin{aligned}
\|\mathbf{v}-\mathbf{w}\|^2 &= \langle \mathbf{v}, \mathbf{v}\rangle - \langle \mathbf{v}_L, \mathbf{v}_L\rangle + \sum_{k=1}^n \left(|\langle \mathbf{v}, \mathbf{e}_k\rangle|^2 - 2\mathrm{Re}\left(\overline{a_k}\langle \mathbf{v}_L, \mathbf{e}_k\rangle\right) + |a_k|^2\right) \\
&= \langle \mathbf{v}, \mathbf{v}\rangle - \langle \mathbf{v}_L, \mathbf{v}_L\rangle + \sum_{k=1}^n |\langle f, \mathbf{e}_k\rangle - a_k|^2
\end{aligned}$$

Varying $\mathbf{w}$, it is clear from this equation that $\|\mathbf{v}-\mathbf{w}\|^2$ is minimized precisely when $\sum_{k=1}^n |\langle f, \mathbf{e}_k\rangle - a_k|^2 = 0$. This occurs if and only if $a_k = \langle f, \mathbf{e}_k\rangle$ for each k, i.e. $\mathbf{w} = \mathbf{v}_L$. □

The vector $\mathbf{v}_L$ is called the *orthogonal projection* of $\mathbf{v}$ onto L, and Lemma 22.2 characterizes the orthogonal projection as the unique best approximation of $\mathbf{v}$ by vectors in L. This characterization allows us to prove that the formula (22.1) holds for any $\mathbf{v} \in H$.

Corollary 22.3. *Let H be a separable Hilbert space and let $\{\mathbf{e}_n\}_{n\in\mathbb{N}}$ be an orthonormal basis for H. Then for any $\mathbf{v} \in H$ the infinite series*

$$\sum_{n=1}^\infty \langle \mathbf{v}, \mathbf{e}_n\rangle \mathbf{e}_n$$

converges to $\mathbf{v}$.

Proof. Let $H_N = \mathrm{Span}\{\mathbf{e}_1, \dots, \mathbf{e}_N\}$, a finite dimensional subspace of H. For $\mathbf{v} \in H$ define

$$d(\mathbf{v}, H_N) = \inf_{\mathbf{w} \in H_N} \|\mathbf{v} - \mathbf{w}\|$$

The fact that $\{\mathbf{e}_n\}_{n\in\mathbb{N}}$ has dense linear span implies that $d(\mathbf{v}, H_N) \to 0$ as N tends to infinity. Setting $\mathbf{v}_N = \sum_{n=1}^{N} \langle \mathbf{v}, \mathbf{e}_n\rangle \mathbf{e}_n$, we have that $d(\mathbf{v}, H_N) = \|\mathbf{v} - \mathbf{v}_N\|$ by Lemma 22.2. Thus $\|\mathbf{v} - \mathbf{v}_N\| \to 0$ as N tends to infinity which means:

$$\mathbf{v} = \lim_{N\to\infty} \sum_{n=1}^{N} \langle \mathbf{v}, \mathbf{e}_n\rangle \mathbf{e}_n$$

as desired. □

Remark 22.4. We formulated Lemma 22.1 and Lemma 22.2 only for finite dimensional subspaces, but in fact it follows from Corollary 22.3 that the same statements are true for any closed subspace of a separable Hilbert space. This is because Corollary 22.3 guarantees that the computations with finite series in the proofs of the two lemmas can be replaced with analogous computations with infinite series.

Corollary 22.3 finally verifies that we can treat an orthonormal basis for a Hilbert space as a sort of infinite dimensional coordinate system. This often allows us to import results from finite dimensional linear algebra that might have been proven using bases or dimension arguments. One important example of this phenomenon is the statement that any closed subspace L of a Hilbert space H is *complemented*, meaning there exists a subspace L' of H such that $H = L \oplus L'$. We will prove this in the case where H is separable, but first we introduce some more terminology.

Definition 22.5. Let H be a separable Hilbert space and let $S \subseteq H$ be any subset. The *orthogonal complement* of S, denoted by $S^\perp$, is the set of all vectors in H which are orthogonal to every vector in S.

Lemma 22.6. *$S^\perp$ is a closed linear subspace of H for any set S. If L is a closed linear subspace of H then $(L^\perp)^\perp = L$.*

Proof. Given $\mathbf{v}_1, \mathbf{v}_2 \in S^\perp$ and any $\mathbf{x} \in S$ we have $\langle a_1\mathbf{v}_1 + a_2\mathbf{v}_2, \mathbf{x}\rangle = a_1 \langle \mathbf{v}_1, \mathbf{x}\rangle + a_2 \langle \mathbf{v}_2, \mathbf{x}\rangle = 0$ for any $a_1, a_2 \in \mathbb{C}$. Thus $S^\perp$ is a subspace, and it is closed since $\langle \cdot, \cdot\rangle$ is continuous: if $\{\mathbf{v}_n\}$ is a sequence in $S^\perp$ which converges to $\mathbf{v} \in H$ and $\mathbf{x} \in S$ then we have $0 = \langle \mathbf{v}_n, \mathbf{x}\rangle \to \langle \mathbf{v}, \mathbf{x}\rangle$ and hence $\mathbf{v} \in S^\perp$.

Now, given $\mathbf{x} \in S$ and any $\mathbf{y} \in S^\perp$ we have $\langle \mathbf{x}, \mathbf{y}\rangle = 0$ since $\mathbf{y}$ is orthogonal to every vector in S. Thus $\mathbf{x}$ is orthogonal to every vector in $S^\perp$, which shows that $S \subseteq (S^\perp)^\perp$ for any set S. Let us show that if L is a closed subspace of H then $(L^\perp)^\perp \subseteq L$. If $\mathbf{x}$ is not in L then by Lemma 22.1 (together with Remark 22.4) we can write $\mathbf{x} = \mathbf{x}_L + \mathbf{x}^\perp$ where $\mathbf{x}_L \in L$ and $\mathbf{x}^\perp$ is a nonzero vector in $L^\perp$. But then we have $\left\langle \mathbf{x}, \mathbf{x}^\perp\right\rangle = \left\langle \mathbf{x}^\perp, \mathbf{x}^\perp\right\rangle = \left\|\mathbf{x}^\perp\right\| \neq 0$ since $\left\langle \mathbf{x}_L, \mathbf{x}^\perp\right\rangle = 0$, so $\mathbf{x}$ is not orthogonal to the vector $\mathbf{x}^\perp \in L^\perp$. Thus $\mathbf{x}$ is not in $(L^\perp)^\perp$, as desired. □

As a brief side note, one can sometimes use this lemma to construct an explicit model for the completion of a normed vector space V: if V isometrically embeds into a known Hilbert space H then $(V^\perp)^\perp$ is the closure of V in H and hence is isometrically isomoprhic to the completion of V.

Proposition 22.7. *Let L be a closed subspace of a separable Hilbert space H. Then $H = L \oplus L^\perp$*

Proof. According to the definition of direct sum we must check that $L \cap L^\perp = 0$ and that $H \subseteq L \oplus L^\perp$.

For any $\mathbf{x} \in L$ we have $\langle \mathbf{x}, \mathbf{x} \rangle = \|\mathbf{x}\|$, so the only vector in L which is orthogonal to every vector in L is the zero vector. This shows that $L \cap L^\perp = 0$.

Suppose there is a vector in H which is not in $L \oplus L^\perp$. By Lemma 22.1 (together with Remark 22.4) there is a vector in H which is orthogonal to $L \oplus L^\perp$. This means that $\mathbf{v}$ is orthogonal to both L and $L^\perp$, so that $\mathbf{v} \in L^\perp \cap (L^\perp)^\perp = L^\perp \cap L$. But $L^\perp \cap L = 0$ and so $\mathbf{v} = 0$, a contradiction. □

One often says that H is the *orthogonal direct sum* of L and $L^\perp$. The existence of this decomposition is a special feature of Hilbert space theory: a theorem of Lindenstrauss and Tzafririy in 1967 asserts that if a Banach space V has the property that every closed subspace of V is complemented then V is isometrically isomorphic to a Hilbert space. But the real importance of Proposition 22.7 stems from the fact that it provides an explicit description for the direct summand $L^\perp$. We will see a number of applications of this result.

Another interesting consequence of Corollary 22.3 is that it allows us to completely classify separable Hilbert spaces up to isometric isomorphism:

Proposition 22.8. *Every separable Hilbert space H is isometrically isomorphic to ℓ^2*

Proof. By Proposition 21.4, H has a countable orthonormal basis $\{\mathbf{e}_n\}_{n\in\mathbb{N}}$. By the previous corollary every $\mathbf{v} \in H$ satisfies

$$\mathbf{v} = \sum_{n=1}^{\infty} \langle \mathbf{v}, \mathbf{e}_n \rangle \, \mathbf{e}_n$$

so it follows that

$$\|\mathbf{v}\|^2 = \sum_{n=1}^{\infty} |\langle \mathbf{v}, \mathbf{e}_n \rangle|^2$$

In particular the sequence of complex numbers $s(\mathbf{v}) = \{\langle \mathbf{v}, \mathbf{e}_n \rangle\}_{n\in\mathbb{N}}$ is an element of ℓ^2 with the same norm as $\mathbf{v}$. Clearly $s(\mathbf{v})$ depends linearly on $\mathbf{v}$, so we have a linear isometry $s : H \to \ell^2$. Any linear isometry is automatically injective; s is surjective because for any $\{a_n\} \in \ell^2$ the series $\sum_{n=1}^{\infty} a_n \mathbf{e}_n$ converges by the triangle inequality and by completeness of H. □

This result may make the theory of separable Hilbert spaces seem trivial since it says that there is only one example up to isomorphism. In fact this rigidity is one of the reasons why Hilbert space techniques are so powerful: separable Hilbert spaces arise naturally in a variety of different contexts, and it is useful to know that their overall structure is independent of the details of their presentation. For example, the abstract results about orthonormal bases that we proved in this section immediately imply nontrivial statements about Fourier approximations of L^2 functions.

We conclude our current exploration of Hilbert space theory with a discussion of duality. It is not a priori obvious that the dual of a Hilbert space is itself a Hilbert space, but we can say something much stronger: Proposition 22.8 together with a complex analogue of Proposition 12.5 imply that every Hilbert space is isometrically isomorphic to its dual (up to a small caveat which will be explained shortly). It is worthwhile to examine this result more closely and prove it using geometric methods; as a byproduct we will obtain an explicit way to represent the dual of a separable Hilbert space.

Given any vector $\mathbf{v}$ in a separable Hilbert space H, define a linear functional $\ell_{\mathbf{v}} : H \to \mathbb{C}$ by $\ell_{\mathbf{v}}(\mathbf{x}) = \langle \mathbf{x}, \mathbf{v} \rangle$. The assignment $\mathbf{v} \mapsto \ell_{\mathbf{v}}$ is conjugate linear in $\mathbf{v}$ (meaning $\ell_{a\mathbf{v}+a'\mathbf{v}'} = \overline{a}\ell_{\mathbf{v}} + \overline{a'}\ell_{\mathbf{v}'}$) and we shall prove that it gives an explicit identification between H and H^*.

Proposition 22.9. *The linear functional $\ell_{\mathbf{v}}$ is continuous for every $\mathbf{v} \in H$, and the map $H \to H^*$ given by $\mathbf{v} \mapsto \ell_{\mathbf{v}}$ is a conjugate linear isometric isomorphism of Banach spaces.*

Proof. For any unit vector $\mathbf{x}$ we have that $|\ell_{\mathbf{v}}(\mathbf{x})| \leq \|\mathbf{v}\|$ by the Cauchy-Schwarz inequality, so $\|\ell_{\mathbf{v}}\|^* \leq \|\mathbf{v}\|$ and hence $\ell_{\mathbf{v}}$ is a bounded linear functional. In fact we can see that $\|\ell_{\mathbf{v}}\|^* = \|\mathbf{v}\|$ by setting $\mathbf{x} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$ so we have shown that the map under consideration is an isometry. This automatically implies that it is injective, so it remains only to prove surjectivity.

Let $\ell \in H^*$ be any nonzero bounded linear functional and let L denote its kernel, a closed subspace of H. Since $\ell \neq 0$, L has codimension 1 in H and thus $L^{\perp}$ is a 1 dimensional subspace of H by Proposition 22.7. Let $\mathbf{w}$ be any unit vector in $L^{\perp}$ and define $\mathbf{v} = \overline{\ell(\mathbf{w})}\mathbf{w}$; we shall prove that $\ell = \ell_{\mathbf{v}}$. For any $\mathbf{x} \in H$, $\mathbf{x}$ has a unique decomposition $\mathbf{x} = \mathbf{x}_L + \mathbf{x}^{\perp}$ where $\mathbf{x}_L \in L$ and $\mathbf{x}^{\perp} \in L^{\perp}$ by Proposition 22.7. Note that $\mathbf{x} = a\mathbf{w}$ for some $a \in \mathbb{C}$ since $L^{\perp}$ is one dimensional.

Let us calculate $\ell_{\mathbf{v}}(\mathbf{x})$:

$$\begin{aligned}
\ell_{\mathbf{v}}(\mathbf{x}) &= \langle \mathbf{x}, \mathbf{v} \rangle \\
&= \left\langle \mathbf{x}_L + a\mathbf{w}, \overline{\ell(\mathbf{w})}\mathbf{w} \right\rangle \\
&= \ell(\mathbf{w}) \langle \mathbf{x}_L, \mathbf{w} \rangle + a\ell(\mathbf{w}) \langle \mathbf{w}, \mathbf{w} \rangle \\
&= a\ell(\mathbf{w})
\end{aligned}$$

But $\ell(\mathbf{x}) = \ell(\mathbf{x}_L) + \ell(a\mathbf{w}) = a\ell(\mathbf{w})$ since $\mathbf{x}_L \in L = \ker(\ell)$, so $\ell_{\mathbf{v}} = \ell$ as desired. $\square$

The proof of this proposition tells us how to define a natural inner product on H^*: given $\ell, \ell' \in H^*$ take unit vectors $\mathbf{w}$ and $\mathbf{w}'$ in the orthogonal complements of $\ker(\ell)$ and $\ker(\ell')$, respectively, and define

$$\langle \ell, \ell' \rangle^* = \left\langle \overline{\ell(\mathbf{w})}\mathbf{w}, \overline{\ell(\mathbf{w}')}\mathbf{w}' \right\rangle$$

Note that the vectors $\overline{\ell(\mathbf{w})}\mathbf{w}$ and $\overline{\ell(\mathbf{w}')}\mathbf{w}'$ depend only on ℓ and ℓ', respectively, so this is well-defined.

Proposition 22.9 also implies that every separable Hilbert space is reflexive. Reflexivity is invariant under isometric isomorphism, so it provides a way to prove that certain Banach spaces - such as $C[0,1]$ (which we will soon prove is not reflexive) - cannot be given Hilbert space structures that generate the topology of the space.

23. Lecture 23 (10/21/11): The dual of $C[0,1]$. Preliminaries

23.1. **Examples of linear functionals in $C[0,1]$.** So far we have introduced three main examples of infinite dimensional Banach spaces: $C[0,1]$, ℓ^p, and $L^p[0,1]$. We have explicitly identified the dual of ℓ^p with ℓ^q, where $\frac{1}{p} + \frac{1}{q} = 1$, and with some additional techniques from integration theory we would have been able to identify the dual of $L^p[0,1]$ with $L^q[0,1]$. We also constructed a conjugate linear isometric isomorphism between a separable Hilbert space and its dual, though this does not add to the list of examples of Banach spaces whose duals we can explicitly identify since we also proved that every separable Hilbert space is isometrically isomoprhic to ℓ^2. The goal of the next two lectures is to build an explicit model for the dual of our very first example: $C[0,1]$.

We will describe $C[0,1]$ as a fairly complicated space of functions on $[0,1]$. Before introducing this space, let us consider a few examples to illustrate the difficulties:

Example 23.1. Given $c \in [0,1]$, define a linear functional ℓ_c on $C[0,1]$ by $\ell_c(f) = f(c)$. This linear functional is continuous since uniform convergence of functions implies pointwise convergence.

Example 23.2. Define a linear functional ℓ on $C[0,1]$ by $\ell(f) = \int_0^1 f(x)dx$. This linear functional is continuous since $\int_0^1 f_n(x)dx = \int_0^1 f(x)dx$ if f_n converges to f uniformly.

These two examples of bounded linear functionals on $C[0,1]$ appear to have quite different behavior: the first example depends only on the value of f at a single point while the second example depends on the values of f at every point! The second example can be easily generalized: the linear functional $f \mapsto \int_0^1 f(x)g(x)dx$ on $C[0,1]$ is continuous for any Riemann integrable function g. There are even more complicated examples, and to capture all of them we need to dive more deeply into one variable analysis.

23.2. **Monotone functions and functions of bounded variation.** It turns out that the key tool in this discussion is the concept of the *variation* of a function. Informally, the variation of a function ϕ on $[0,1]$ measures the total vertical distance travelled by a point moving along the graph of ϕ. For a monotone function the variation is simply the absolute value of the difference of its values at the endpoints. So we begin with a discussion of the structure of monotone functions.

Lemma 23.3. *Let $\phi : [a,b] \to \mathbb{R}$ be a nondecreasing function. Then ϕ has a left and right limit at every point and at most countably many discontinuities.*

Proof. For any increasing sequence $\{x_n\}$ in $[a,b]$ converging to x, we have that $\{\phi(x_n)\}$ is a nondecreasing sequence of real numbers bounded by $\phi(b)$ since ϕ is nondecreasing. This sequence converges to its least upper bound, so ϕ has a left limit at x. A similar argument shows that ϕ has a right limit.

Define a function $\phi_+(x) = \lim_{y \to x+} \phi(y)$, and define $\phi_-(x)$ similarly. Since ϕ is nondecreasing, $\{(\phi_-(x), \phi_+(x))\}_{x \in [a,b]}$ is a disjoint family of open intervals whose union is $[a,b]$. Thus only countably many of them can be nonempty, meaning $\phi_-(x) = \phi_+(x)$ for all but at most countably many x. Since ϕ is continuous at x if and only if $\phi_-(x) = \phi_+(x)$, the proof is complete. $\square$

Thus a monotone function is not too far off from being continuous; how far off is measured by the following definition:

Definition 23.4. Suppose that $\phi : [a,b] \to \mathbb{R}$ is a function which has left and right limits at every point. Then ϕ is a *jump function* if $\sum_{x \in D} \phi_+(x) - \phi_-(x) = \phi(b) - \phi(a)$ where $\phi_+(x)$ (respectively, $\phi_-(x)$) denotes the right (respectively, left) limit of ϕ at x and D is the discontinuity locus of ϕ.

Intuitively a jump function is a function which does not continuously increase or decrease at any point; piecewise constant functions are always jump functions, while $\phi(x) = x$ is not. Jump functions can be quite strange; in fact, for any countable set S there is a monotonic jump function with discontinuity locus S. Just choose any sequence $\{c_n\}$ such that $c_n > 0$ and $\sum_n c_n < \infty$, and define

$$\phi = \sum_{x_n \in S} c_n \chi_{[x_n,b]}$$

where $\chi_{[x_n,b]}$ denotes the characteristic function of $[x_n, b]$.

Proposition 23.5. *Any nondecreasing function on* $[a, b]$ *is the sum of a continuous function and a jump function.*

Proof. Let ϕ be a monotone function with (countable) discontinuity locus D and set $\phi'(x) = \sum_{x\in D}(\phi_+(x) - \phi(x))\chi_{[x,b]}$ where $\phi_+(x)$ is the right limit of ϕ at x. We have that $\phi_+(x) \geq \phi(x)$ for all x and $\sum_{x\in D}(\phi_+(x) - \phi(x)) \leq \phi(b) - \phi(a)$ since ϕ is monotonic, so ϕ' is a jump function. It follows from basic properties of left and right limits that $\phi - \phi'$ is continuous. $\square$

Intuitively, this proposition makes the believable claim that the only ways a nondecreasing function can increase are continuously or by jumps. The reader is warned, however, not to extrapolate too much from this intuition; the next example shows that even continuous nondecreasing functions can exhibit strange behavior.

Example 23.6 (The Devil's Staircase)**.** Iteratively define a sequence of continuous functions $\{f_n\}$ on $[0, 1]$ as follows. Let $f_0(x) = x$, and given f_n define f_{n+1} by:

$$f_{n+1}(x) = \begin{cases} \frac{1}{2}f_n(3x) & x \in [0, \frac{1}{3}] \\ \frac{1}{2} & x \in [\frac{1}{3}, \frac{2}{3}] \\ \frac{1}{2} + \frac{1}{2}f_n(3x-2) & x \in [\frac{2}{3}, 1] \end{cases} \tag{23.1}$$

The function f_n is continuous and is constant on any complementary interval to the Cantor middle-thirds set of rank up to n, i.e of length at least 2^{-n}. Furthermore, an immediate inductive calculation shows that $|f_{n+1} - f_n| \leq \frac{1}{32^n}$. The sequence f_n stabilizes on complimentary interval to Cantor middle-thirds set C . Thus the sequence $\{f_n\}$ converges uniformly to a function f on $[0, 1]$. f is a continuous, nondecreasing function which satisfies $f(0) = 0$, $f(1) = 1$, and moreover f is constant on the complementary intervals and hence differentiable at every point x in the complement of the set C with $f'(x) = 0$. The set S is well-known to be a null set, so f is an example of a *singular* function which increases continuously inspite of having derivative 0 almost everywhere (i.e. outside of a null set). f can be regarded as a sort of "inverse" to a jump function.

Remark 23.7. The above construction represents the Devil's Staircase as the limit of a sequence of continuous functions. A slight modification of this construction represents it as the limit of ordinary "staircases", i.e jump functions with finitely many jumps. We start with the function f_0 that is equal to 0 on $(0, 1/3]$, $1/2$ on $(1/3, 2/3]$ and 1 on $[2/3, 1]$. and then construct the sequence f_n by the same iterative formula (23.1).

We will show that the dual of $C[0, 1]$ can be identified with the space of differences of nondecreasing functions on $[0, 1]$. A jump function ϕ has this form if and only if it has the form

$$\phi = \sum_{x_n \in S} c_n \chi_{[x_n,1]}$$

where S is a countable subset of $[0,1]$ and $\{c_n\}$ is a sequence of real numbers such that $\sum_n |c_n| < \infty$. To extend this characterization to more general functions, we approximate them by jump functions using the following definition:

Definition 23.8. Let ϕ be a function on $[a,b]$. Consider partitions P of $[a,b]$ of the form $P = \{a = x_0 < x_1 < \ldots < x_{n_P} = b\}$, and define the *variation* of ϕ to be the quantity:

$$\mathrm{Var}_{[a,b]}(\phi) = \sup_P \sum_{i=0}^{n_P-1} |\phi(x_{i+1}) - \phi(x_i)| \tag{23.2}$$

When the interval $[a,b]$ is clear from the context, we often write $\mathrm{Var}(f)$ instead of $\mathrm{Var}_{[a,b]}(f)$.

We say that ϕ has *bounded variation* on $[a,b]$ if $\mathrm{Var}_{[a,b]}(\phi) < \infty$.

It is immediately obvious form the definition that for a nondecreasing function ϕ on $[a,b]$ the sum in the right-hand part of (23.2) is independent of P and hence $\mathrm{Var}_{[a,b]}(\phi) = \phi(b) - \phi(a)$.

Exercise 23.9. Given a continuously differentiable function ϕ on $[a,b]$, show that $\mathrm{Var}_{[a,b]}(\phi) = \int_a^b |\phi'(x)|\, dx$.

Here is a convenient and useful characterization of functions of bounded variation.

Proposition 23.10. *A function f has bounded variation if and only if f is the difference of two nondecreasing functions.*

Proof. Obviously $\mathrm{Var}(f+g) \le \mathrm{Var}f + \mathrm{Var}g$ since this is true for any partition that figures in (23.2). This proves the "if" part. We deduce also that The set of all functions of bounded variation on $[a,b]$ form a vector space.

Now suppose that $f \in BV[a,b]$ and define $f_+(x) = \mathrm{Var}_{[a,x]}(f)$. If $x < y$ then we can calculate $f_+(y)$ using partitions which include x as a partition point and deduce that $f_+(y) = f_+(x) + \mathrm{Var}_{[x,y]}(f)$. Since the variation of a function on any interval is nonnegative, it follows that $f_+(x) \le f_+(y)$. Additionally we have $f_+(y) - f(y) = f_+(x) + \mathrm{Var}_{[x,y]}(f) - f(y)$, and since $\mathrm{Var}_{[x,y]}(f) \ge f(y) - f(x)$ (using the partition of $[x,y]$ consisting only of the two endpoints) it follows that $f_+(y) - f(y) \ge f_+(x) - f(x)$. So f_+ and $f_+ - f$ are two nondecreasing functions whose difference is f. □

23.3. **Riemann-Stieltjes integral.** Now we will associate a bounded linear functional on $C[0,1]$ to a function of bounded variation. This is achieved via the *Riemann-Stieltjes integral*, which we define next.

To motivate the definition of the Riemann-Stieltjes integral, let us briefly review how the ordinary Riemann integral is defined. Given a function f on $[a,b]$ and a partition $P = \{a = x_0 < x_1 < \ldots < x_{n_P} = b\}$, define the upper and lower sums of

f relative to P as follows:

$$U(f,P) = \sum_{i=0}^{n_P-1} \sup_{x\in[x_i,x_{i+1}]} f(x)(x_{i+1}-x_i)$$
$$L(f,P) = \sum_{i=0}^{n_P-1} \inf_{x\in[x_i,x_{i+1}]} f(x)(x_{i+1}-x_i)$$

If P_1 and P_2 are any two partitions of $[a,b]$ then $L(f,P_1) \leq U(f,P_2)$, so it follows that

$$\sup_P L(f,P) \leq \inf_P U(f,P)$$

f is Riemann integrable if and only if $\sup_P L(f,P) = \inf_P U(f,P)$, and this value is defined to be the Riemann integral of f over $[a,b]$ when f is Riemann integrable. The definition of the Riemann-Stieltjes integral is similar, only the lengths of the partition intervals are weighted by an auxillary function of bounded variation.

Definition 23.11. Let ϕ be a function of bounded variation on $[a,b]$ and let f be any function on $[a,b]$. For any partition $P = \{a = x_0 < x_1 < \ldots < x_{n_P} = b\}$ of $[a,b]$, define

$$U(f,\phi,P) = \sum_{i=0}^{n_P-1} \sup_{x\in[x_i,x_{i+1}]} f(x)(\phi(x_{i+1})-\phi(x_i))$$
$$L(f,\phi,P) = \sum_{i=0}^{n_P-1} \inf_{x\in[x_i,x_{i+1}]} f(x)(\phi(x_{i+1})-\phi(x_i))$$

We say that f is *ϕ-Riemann-Stieltjes integrable* if

$$\sup_P L(f,\phi,P) = \inf_P U(f,\phi,P)$$

If f is ϕ-Riemann-Stieltjes integrable then we define its *ϕ-Riemann-Stieltjes integral* to be

$$\int_a^b f d\phi = \sup_P L(f,\phi,P) = \inf_P U(f,\phi,P)$$

Thus the ordinary Riemann integral is the Riemann-Stieltjes integral relative to the identity function on $[a,b]$. As with ordinary Riemann integration, the Riemann-Stieltjes integral can be calculated using a sequence of increasingly refined partitions. Recall that the *mesh* of a partition $P = \{a = x_0 < x_1 < \ldots < x_{n_P} = b\}$ is defined to be $\max_i\{|x_{i+1}-x_i|\}$.

Lemma 23.12. *If f is ϕ-Riemann-Stieltjes integrable on $[a,b]$ and $\{P_k\}$ is a sequence of partitions such that mesh$(P_k) \to 0$ as $k \to \infty$, then*

$$\int_a^b f d\phi = \lim_{k\to\infty} L(f,\phi,P_k) = \lim_{n\to\infty} U(f,\phi,P_k)$$

The proof of this result is similar to the proof of the corresponding result for ordinary Riemann integrals and is left to the reader.

The next result explains why Riemann-Stieltjes integrals are defined relative to functions of bounded variation.

Proposition 23.13. *If $\phi \in BV[a,b]$ and $f \in C[a,b]$ then f is ϕ-Riemann-Stieltjes integrable.*

Proof. Choose a sequence of partitions $P_k = \{a = x_0^k < x_1^k < \ldots < x_{n_k}^k = b\}$ such that $\text{mesh}(P_k) \to 0$ as $k \to \infty$. It suffices to show that for every $\varepsilon > 0$ there exists K such that $k \geq K$ implies $U(f, \phi, P_k) - L(f, \phi, P_k) < \varepsilon$. We use the fact that f is uniformly continuous on $[a, b]$: given $\varepsilon > 0$ there exists δ such that $|f(x) - f(y)| < \frac{\varepsilon}{\text{Var}(\phi)}$ whenever $|x - y| < \delta$. Choose K large enough so that $k \geq K$ implies $\text{mesh}(P_k) < \delta$; for such k we have

$$\left(\sup_{x \in [x_{i+1}^k - x_i^k]} f(x)\right) - \left(\inf_{x \in [x_{i+1}^k - x_i^k]} f(x)\right) < \frac{\varepsilon}{\text{Var}(\phi)}$$

Thus for $k \geq K$ we have:

$$\begin{aligned}
&U(f, \phi, P_k) - L(f, \phi, P_k) \\
&= \sum_{i=0}^{n_k - 1} \left(\sup_{[x_{i+1}^k - x_i^k]} f(x)\right) - \left(\inf_{[x_{i+1}^k - x_i^k]} f(x)\right) (\phi(x_{i+1}) - \phi(x_i)) \\
&< \sum_{i=0}^{n_k - 1} \frac{\varepsilon}{\text{Var}(\phi)} (\phi(x_{i+1}) - \phi(x_i)) \leq \varepsilon
\end{aligned}$$

□

The Riemann-Stieltjes integral is clearly linear in both f and ϕ; we claim that the linear functional $\ell_\phi(f) = \int_a^b f d\phi$ on $C[a, b]$ is bounded. In fact the following lemma shows that $\|\ell_\phi\|^* \leq \text{Var}(\phi)$.

Lemma 23.14. *For any $f \in C[a, b]$, $\phi \in BV[a, b]$ we have*

$$\left| \int_a^b f d\phi \right| \leq \|f\| \; Var_{[a,b]}(\phi)$$

where $\|\cdot\|$ is the uniform norm on $C[a, b]$.

Proof. Given any partition $P = \{a = x_0 < \ldots < x_{n_P} = b\}$ of $[a, b]$ we have:

$$\begin{aligned}
\left| \int_a^b f d\phi \right| &\leq |U(f, \phi, P)| \\
&\leq \sum_{i=0}^{n_P - 1} \sup_{x \in [x_i, x_{i+1}]} |f(x)| \, |\phi(x_{i+1}) - \phi(x_i)| \\
&\leq \sup_{x \in [a,b]} |f(x)| \sum_{i=0}^{n_P - 1} |\phi(x_{i+1}) - \phi(x_i)| \\
&\leq \|f\| \, \text{Var}_{[a,b]}(\phi)
\end{aligned}$$

□

Thus we have associated to every function of bounded variation on $[a, b]$ a bounded linear functional in $C[a, b]$. Notice however that this identification is not injective. First if two functions of bounded variation are differ by a constant they define the same linear functional. For continuous functions this is the only source of non-uniqueness. For discontinuous functions the values at discontinuity points can be changed arbitrarily without changing the values of the Riemann-Stieltjes integral of *continuous functions*. So one cam assume that the functions of bounded variation are left-continuous. Then Var defines a norm on the space of equivalence classes of left-continuous functions of bounded variation: $f \sim g$ if $f - g = const$. We denote this space by $(BV[a,b]$ In the next lecture we will prove that Riemann-Stiltjes integration identifies $BV[a,b]$ with $C[a,b]^*$.

24. Lecture 24 (10/25/11)

24.1. The Riesz Representation Theorem.

Theorem 24.1 (Riesz Representation Theorem)**.** *For any $\phi \in BV[a,b]$, define $\ell_\phi \in C[a,b]^*$ by $\ell_\phi(f) = \int_a^b f d\phi$. The map $BV[a,b] \to C[a,b]^*$ given by $\phi \mapsto \ell_\phi$ is a linear isomorphism which satisfies $\|\ell_\phi\|^* = Var(\phi)$.*

It will follow from this theorem that $(BV[a,b], \mathrm{Var})$ is a Banach space which is isometrically isomorphic to $C[a,b]^*$ (that $BV[a,b]$ is complete with respect to Var).

Before proving the theorem, let us revisit the two examples of linear functionals identified at the beginning of the previous lecture and verify that they arise as Riemann-Stieltjes integration against functions of bounded variation.

Example 24.2. Recall the linear functional $\ell_c \in C[0,1]^*$ from Example 23.1 given by $\ell_c(f) = f(c)$ for $c \in [0,1]$. Define ϕ_c to be the characteristic function of the interval $(c,1]$; this function is in $BV[0,1]$ since its variation is 1. We can calculate the Riemann-Stieltjes integral $\int_0^1 f d\phi_c$ using partitions which include c as a partition point, and for such partitions the expression $\phi_c(x_{i+1}) - \phi_c(x_i)$ is 0 except when $x_{i+1} = c$, and for that value of i it is 1. So simple estimates show that $\int_0^1 f d\phi_c = f(c)$, and hence $\ell_c = \ell_{\phi_c}$.

Remark 24.3. This example shows that the Riemann-Stieltjes can be used to give some meaning to the "Dirac delta function" $\delta(x)$ which is meant to satisfy the equation $\int_{\mathbb{R}} f(x)\delta(x)dx = f(0)$. Of course there is no actual function which has this property, but there is a theory of *generalized functions* (also called *distributions*) which includes δ as an example. This theory is very closely related to duality theory.

Example 24.4. Recall the linear functional $\ell \in C[0,1]^*$ from Example 23.2 given by $\ell(f) = \int_0^1 f(x)dx$. We commented in the previous lecture that the Riemann integral is the Riemann-Stieltjes integral for the function $\phi(x) = x$, and thus $\ell = \ell_\phi$.

Proof of Theorem 24.1. Let B denote the vector space spanned by $C[a,b]$ and the piecewise constant functions on $[a,b]$, equipped with the uniform norm. Any functional $\ell \in C[a,b]^*$ extends to a functional ℓ' on B without increasing its norm by the Hahn-Banach theorem; by first extending to the span of $C[a,b]$ and the set of all characteristic functions of one point sets, we can assume that $\ell'(\chi_{\{c\}}) = 0$ where $\chi_{\{c\}}$ is the characteristic function of $\{c\}$. So we can define a function ϕ on $[a,b]$ by $\phi(x) = \ell'(\chi_{[a,x]})$ where $\chi_{[a,x]}$ is the characteristic function of $[a,x]$.

First, we show that $\phi \in BV[a,b]$. Given a partition $P = \{a = x_0 < \ldots < x_{n_P} = b\}$ of $[a,b]$, define λ_i to be the number

$$\lambda_i = \frac{\phi(x_{i+1}) - \phi(x_i)}{|\phi(x_{i+1}) - \phi(x_i)|}$$

if the denominator is nonzero and 0 otherwise. Consider the function $f \in B$ given by

$$f = \sum_{i=0}^{n_P-1} \lambda_i \chi_{[x_i,x_{i+1}]}$$

Note that $\|f\| \leq 1$. We calculate:

$$\begin{aligned}\sum_{i=0}^{n_P-1} |\phi(x_{i+1}) - \phi(x_i)| &= \sum_{i=0}^{n_P-1} \lambda_k(\phi(x_{i+1}) - \phi(x_i)) \\ &= \sum_{i=0}^{n_P-1} \lambda_i \ell'(\chi_{[x_i,x_{i+1}]}) \\ &= \ell'(f) \leq \|\ell\|^*\end{aligned}$$

Passing to the supremum over all partitions of $[a,b]$, we deduce that $\mathrm{Var}(\phi) \leq \|\ell\|^*$ and hence that $\phi \in BV[a,b]$.

Next we show that $\ell'(g) = \int_a^b g d\phi$ for every $g \in B$. Since the space of piecewise constant functions is dense in B and Riemann-Stieltjes integration commutes with (uniform) limits, it suffices to prove this in the case where $g = \chi_{(c,d)}$. Using the partition $P = \{a,c,d,b\}$ of $[a,b]$ we calculate:

$$\begin{aligned}\int_a^b g d\phi &= 0(\phi(c) - \phi(a)) + 1(\phi(d) - \phi(c)) + 0(\phi(b) - \phi(d)) \\ &= \phi(d) - \phi(c) \\ &= \ell'(\chi_{[c,d]}) - \ell'(\chi_{[c,c]}) \\ &= \ell'(\chi_{[c,d]})\end{aligned}$$

For any other partition of $[a,b]$ we may refine the partition so that c and d are partition points, and all of the terms in the sum defining the Riemann-Stieltjes integral other than the terms appearing in the calculation above will vanish. Thus $\int_a^b g d\phi = \ell'(g)$, as desired.

We have now shown that the map $BV[a,b] \to C[a,b]^*$ given by $\phi \mapsto \ell_\phi$ is onto. We showed in the previous lecture that $\|\ell_\phi\|^* \leq \mathrm{Var}(\phi)$, and we proved the reverse inequality above. We conclude that our map $BV[a,b] \to C[a,b]^*$ is an isometric isomorphism, as desired. □

One nontrivial consequence of the Riesz representation theorem is that $C[a,b]$ is not reflexive. To see this, note that for any pair of distinct points $x, y \in [a,b]$ we have that the characteristic functions $\chi_{\{x\}}$ and $\chi_{\{y\}}$ have bounded variation and satisfy $\mathrm{Var}(\chi_{\{x\}} - \chi_{\{y\}}) = 2$. This yields an uncountable discrete set in $BV[a,b]$, proving that $BV[a,b]$ is not separable.

Exercise 24.5. If V is a normed space whose dual is separable then V is separable.

Since $C[a,b]$ is separable but $BV[a,b]$ is not, $C[a,b]$ can't be the dual of $BV[a,b]$.

25. Lecture 25 (10/26/11): Weak Topologies

With our explicit construction of the dual of $C[0,1]$ (and of $C[a,b]$ more generally), we have built a substantial library of examples of infinite dimensional Banach spaces and their duals. In what we have done so far we have really only needed one tool belonging to the general theory of Banach spaces: the Hahn-Banach theorem. There are a variety of other geometric tools which aid in the study of Banach spaces, and we now turn our attention to developing some of those tools. When applied to some of the examples that we have constructed, they will help us prove some deep results in analysis.

One of the most important tricks of the trade in functional analysis is to generalize the notion of convergence in a normed space. The standard notion of convergence is norm convergence: we declare that a sequence $\{\mathbf{x}_n\}$ in a normed space V converges to $\mathbf{x}$ if $\|\mathbf{x} - \mathbf{x}_n\| \to 0$ as $n \to \infty$. However, for many purposes this notion is too restrictive. Here are two other natural notions of convergence of sequences in a normed space:

- A sequence $\{\mathbf{x}_n\}$ in V converges in the weak sense to $\mathbf{x}$ if for every $\ell \in V^*$ we have $\ell(\mathbf{x}_n) \to \ell(\mathbf{x})$.
- A sequence $\{\ell_n\}$ in V^* converges in the weak* sense to ℓ if for every $\mathbf{x} \in V$ we have $\ell_n(\mathbf{x}) \to \ell(\mathbf{x})$.

Exercise 25.1. Show that in finite dimensions weak convergence, weak* convergence, and norm convergence are all the same.

In infinite dimensions, however, the three notions of convergence can all be different. Norm convergence is the strongest of the three notions in the sense that fewer sequences converge in norm than in the weak or weak* sense.

Example 25.2. Let H be a separable Hilbert space and let $\{\mathbf{e}_n\}$ be an orthonormal basis for H. We proved that any $\ell \in H^*$ has the form $\ell(\mathbf{x}) = \langle \mathbf{x}, \mathbf{y} \rangle$ for some fixed $\mathbf{y} \in H$, so $\ell(\mathbf{e}_n) = \langle \mathbf{e}_n, \mathbf{y} \rangle$. Recall that

$$\|\mathbf{y}\|^2 = \sum_{n=1}^{\infty} |\langle \mathbf{e}_n, \mathbf{y} \rangle|^2$$

so $\langle \mathbf{e}_n, \mathbf{y} \rangle \to 0$ as $n \to \infty$. Thus $\{\mathbf{e}_n\}$ converges weakly to 0. But $\|\mathbf{e}_n - \mathbf{e}_m\| = 1$ whenever $n \neq m$, so $\{\mathbf{e}_n\}$ does not converge in norm.

We shall see shortly that weak convergence and weak* convergence do not agree in general, but first we need to examine these notions a bit more closely. What allows us to simply define a new notion of convergence? A priori we have no guarantee that such notions will behave in any reasonable way; it is not immediately given to us for free that weak and weak* limits are unique, for example. This would be obvious if there were a metric such that metric convergence was the same thing as weak (or weak*) convergence. Later we will find such a metric for a weak* convergence in $V*$ where V is a separable space, but it turns out that in general such a metric may not exist.

The proper framework for generalizing convergence is not metric space theory but rather the abstract theory of *topological spaces*. We already saw in Lecture 2 that many natural concepts in metric space theory such as convergence, continuity, and compactness depend only on the topology of a metric space - that is, on which sets are open. So there is good reason to hope that an appropriate abstraction of openness will allow us to discuss weak convergence and weak* convergence on solid footing, and this leads directly to the abstract notion of a topology. We will very rapidly review the basics of topological spaces, though the reader who has not encountered the definitions below may wish to consult another reference as well.

Definition 25.3. A *topological space* is a set X equipped with a collection $\mathcal{T}$ of subsets of X - called *open sets* - which satisfies the following axioms:

- The empty set $\emptyset$ and X are both in $\mathcal{T}$.
- If $\{U_\alpha\}$ is an arbitrary collection of sets in $\mathcal{T}$ then $\bigcup_\alpha U_\alpha \in \mathcal{T}$.
- If $U_1, \ldots, U_n$ is a finite collection of sets in $\mathcal{T}$ then $\bigcap_{i=1}^n U_i \in \mathcal{T}$.

Exercise 25.4. Show that if $\mathcal{T}_\alpha$ is any collection of topologies on a fixed set X then $\bigcap_\alpha \mathcal{T}_\alpha$ is a topology on X.

Every set X has two topologies: the *discrete topology* for which every subset of X is open, and the *trivial topology* for which the only two open subsets of X are $\emptyset$ and X. These topologies are not generally very useful (except perhaps as counterexamples). The key example, as discussed in Lecture 2, is the metric topology:

Example 25.5. Recall that a subset U of a metric space (X, d) is said to be open if for every $x \in U$ there exists $\varepsilon > 0$ such that the open ball $B_\varepsilon(x) \subseteq U$. The reader was asked to verify in Exercise 2.5 that the collection of all subsets of X which are open in this sense satisfy the axioms of a topology given above.

However, there are a wealth of examples of topologies which are not *metrizable*, i.e. that do not correspond to the metric topology for some metric. Here is one simple and important necessary condition for a topology to be metrizable:

Definition 25.6. A topological space X is said to be *Hausdorff* if for every pair of distinct points $x, y \in X$ there exist disjoint open sets U_x and U_y such that $x \in U_x$ and $y \in U_y$.

Every metrizable space X is Hausdorff. To see this, equip X with a metric d such that the topology of X coincides with the metric topology of (X, d) and observe that for any pair of distinct points x and y the open balls $U_x = B_\varepsilon(x)$ and $U_y = B_\varepsilon(y)$ have the required property if $\varepsilon < \frac{1}{2}d(x, y)$. An example of a non-Hausdorff topological space is given by the indiscrete topology on any set X with at least two points.

There are many useful examples of non-Hausdorff topologies, but in analysis the Hausdorff property is usually desirable. One reason is that the Hausdorff property is a natural requirement when one wishes to take limits of sequences, which we discuss next.

Definition 25.7. Let X be a topological space and let $\{x_n\}$ be a sequence in X. We say that $\{x_n\}$ converges to $x \in X$ if for every neighborhood U of x there exists N such that $n \geq N$ implies $x_n \in U$.

It is not difficult to check that the topological notion of convergence agrees with the metric notion of convergence when X is a metric space. The reader is warned that sequences are not always powerful enough to probe the structure of a topological space (even for metric spaces!); for instance, one can construct a topological space X and a set $C \subseteq X$ which is not closed but which contains the limits of all of its convergent sequences. Such examples can be avoided by sticking to topological spaces which are separable (i.e. which have countable dense subsets).

There is also a notion of continuity in topology:

Definition 25.8. Let $f : X \to Y$ be a map between topological spaces. We say that f is continuous at a point $x \in X$ if for every open set V in Y containing $f(x)$

there exists an open set U in X containing x such that $f(U) \subseteq V$. We say that f is continuous if it is continuous at every point, or equivalently if $f^{-1}(V)$ is an open set in X for every open set V in Y.

Exercise 25.9. Show that a function between metric spaces is continuous in the δ-ε sense if and only if it is continuous in the topological sense.

Exercise 25.10. Show that if $f : X \to Y$ is a continuous function between topological spaces and $\{x_n\}$ is any sequence in X which converges to x then $\{f(x_n)\}$ converges to $f(x)$ in Y. Prove the converse to this statement under the assumption that X is separable.

Finally we give a topological definition of compactness. Since we have already discussed convergence of sequences we can import our earlier definition from metric spaces and declare that a topological space is compact if and only if every sequence has a convergent subsequence. However, this is really only the right definition for separable spaces; a more general definition is as follows:

Definition 25.11. A topological space X is *compact* if for every collection of open sets whose union is X there is a finite subcollection whose union is still X.

Thus three of the most important notions in analysis - continuity, convergence, and compactness - generalize easily to topological spaces. Many, but not all, properties of these notions which may be familiar to the reader in the context of metric spaces also hold for more general topological spaces.

We now introduce a useful tool for constructing and describing topologies.

Definition 25.12. Let X be a set and let $\mathcal{B}$ be any collection of subsets of X. The topology $\mathcal{T}$ *generated* by $\mathcal{B}$ is defined to be the intersection of all topologies for which every set in $\mathcal{B}$ is open.

For example, the metric topology is the topology generated by the open balls in a metric space. The topology generated by $\mathcal{B}$ can be characterized as follows: the open sets are precisely $\emptyset$, X, and any $U \subseteq X$ with the property that for every $x \in U$ there exist $B_1, \ldots, B_n \in \mathcal{B}$ such that $x \in B_i$ for each i and $\bigcap_{i=1}^n B_i \subseteq U$.

A particularly useful application of this construction arises when the set X comes naturally equipped with a family of functions $\{f_\alpha : X \to \mathbb{R}\}$. It is natural to try to construct a topology on X such that every f_α is continuous; it is clear from the definition of continuity that any function on X is continuous relative to the discrete topology, so it is more interesting to look for the "smallest" topology (the one with the fewest open sets) which makes each f_α continuous.

Definition 25.13. Let X be a set and let $\{f_\alpha : X \to \mathbb{R}\}$ be any collection of functions on X. The *weak topology* on X relative to this collection of functions is the intersection of all topologies on X for which every function f_α is continuous.

More concretely, the weak topology for $\{f_\alpha\}$ is the topology generated by the sets $\{f_\alpha^{-1}(a, b)\}$ where (a, b) is an open interval in $\mathbb{R}$. The weak topology is often called the "topology of pointwise convergence" due to the following lemma:

Lemma 25.14. *A sequence $\{x_n\}$ in X converges to x in the weak topology on X relative to the functions $\{f_\alpha\}$ if and only if $f_\alpha(x_n)$ converges to x for every α*

Proof. If x_n converges to x in the weak topology then $f_\alpha(x_n)$ converges to $f_\alpha(x)$ since f_α is by definition continuous. Conversely, suppose $f_\alpha(x_n)$ converges to $f_\alpha(x)$ for every α. This means that for every $\varepsilon > 0$ there exists N_ε such that $n \geq N_\varepsilon$ implies that $|f_\alpha(x) - f_\alpha(x_n)| < \varepsilon$, i.e. $f_\alpha(x_n) \in (f_\alpha(x) - \varepsilon, f_\alpha(x) + \varepsilon)$. By the definition of the weak topology we have that for every open set U containing x there is a set of the form $U_x = \bigcap_{i=1}^k f_{\alpha_i}^{-1}(a_i, b_i)$ such that $x \in U_x$ and $U_x \subseteq U$. Note that each open interval (a_i, b_i) contains $f_{\alpha_i}(x)$, so we can assume that U_x has the form $\bigcap_{i=1}^k f_{\alpha_i}^{-1}(f_{\alpha_i}(x) - \varepsilon, f_{\alpha_i(x)} + \varepsilon)$. We have shown that x_n is in this set whenever $n \geq \max_{1 \leq i \leq k} N_i$, so it follows that x_n converges to x in the weak topology. □

Exercise 25.15. Show that the weak topology on a set X relative to the collection of functions $\{f_\alpha\}$ on X is Hausdorff if and only if $\{f_\alpha\}$ separates points (meaning for every pair of distinct points $x, y \in X$ there exists f_α such that $f_\alpha(x) \neq f_\alpha(y)$).

We are now ready to return to the problem of giving weak convergence and weak* convergence a proper foundation. While neither of these notions of convergence correspond in general to norm convergence for any norm, they each correspond to topological convergence for an appropriately chosen topology.

Definition 25.16. Let V be a normed space.

- The *weak topology* on V is the weak topology relative to the set of all bounded linear functionals on V.
- The *weak* topology* on V^* is the weak topology relative to the set of all bounded linear functionals on V^* which lie in the image of the canonical embedding $i : V \hookrightarrow V^{**}$.

By Lemma 25.14 we have that a sequence $\{\mathbf{x}_n\}$ converges to $\mathbf{x}$ in the weak topology on V if and only if $\ell(\mathbf{x}_n)$ converges to $\ell(\mathbf{x})$ for every $\ell \in V^*$, so convergence in the weak topology agrees with weak convergence as we defined it at the beginning of the lecture. Similarly a sequence $\{\ell_n\}$ converges to ℓ in the weak* topology on V^* if and only if $i_{\mathbf{x}}(\ell_n)$ converges to $i_{\mathbf{x}}(\ell)$ for every $\mathbf{x} \in V$; by definition of $i_{\mathbf{x}}$ this means that $\ell_n(\mathbf{x})$ converges to $\ell(\mathbf{x})$ and thus convergence in the weak* topology agrees with weak convergence.

Lemma 25.17. *The weak topology on V and the weak* topology on V^* are both Hausdorff.*

Proof. First, let $\mathbf{x}$ and $\mathbf{y}$ be two distinct vectors in V. By basic linear algebra there is a linear functional ℓ on the subspace of V spanned by $\mathbf{x}$ and $\mathbf{y}$ such that $\ell(\mathbf{x}) \neq \ell(\mathbf{y})$, and ℓ can be extended to all of V by the Hahn-Banach theorem. Thus the set of all bounded linear functionals on V separates points, and the claim that the weak topology on V is Hausdorff follows from Exercise 25.15.

Second, let ℓ and ℓ' be two distinct elements of V^*. To say that they are different as elements of V^* means they are different as functions on V, so there exists $\mathbf{x} \in V$ such that $\ell(\mathbf{x}) \neq \ell'(\mathbf{x})$. This means that $i_{\mathbf{x}}(\ell) \neq i_{\mathbf{x}}(\ell')$, so the linear functionals on V^* in the image of $V \hookrightarrow V^{**}$ separate points and hence the weak* topology on V^* is Hausdorff. □

There are a variety of reasons for inventing the weak and weak* topologies. A simple reason is that pointwise convergence is a natural notion of convergence, and it is interesting that it is strictly weaker than norm convergence in infinite dimensions. A much deeper reason is provided by the Banach-Alaoglu theorem which asserts that the unit ball in the dual of a Banach space is compact in the weak* topology. We will prove this theorem later in the case of a dual to a separable space. Compactness theorems in infinite dimensional spaces are rare and precious (recall the that unit ball in an infinite dimensional Banach space is never compact in the norm topology), so this result is reason enough to take the weak* topology seriously.

We shall investigate these applications and others in greater detail in future lectures. For now we settle for a simple application to the problem of deciding which Banach spaces are reflexive. If V is reflexive then the canonical embedding $V \hookrightarrow V^{**}$ is an isomorphism and thus the weak and weak* topologies on $V*$ necessarily coincide. One can show that $BV[a,b] = C[a,b]^*$ is separable with respect to the weak * topology but non-separable with respect to the weak topology, so this yields another proof that $C[a,b]^*$ is not reflexive.

26. Lecture 26 (10/28/11): The Baire Category Theorem

The next step in our program of developing tools for understanding the structure of Banach spaces in the abstract is to formulate and prove the Baire category theorem. This result is about the topology of complete metric spaces (as well as certain more general topological spaces), and it has a startlingly wide variety of nontrivial applications to analysis and geometry. Informally it asserts that complete metric spaces are either discrete or very big. The precise definition of the word "big" for these purposes allows one to prove the existence of rather exotic mathematical objects (such as functions on $[0,1]$ which are continuous but nowhere differentiable) by showing that the set of all non-exotic objects is too small. In this and the next few lectures we will prove the Baire category theorem and explain some of its many applications.

We begin by defining precisely what is meant by "big" and "small".

Definition 26.1. Let X be a topological space.

- A subset of X is *nowhere dense* if its closure contains no open subset of X.
- A subset of X is *meager* (or alternatively *of first category*) if it is the countable union of nowhere dense sets.

Example 26.2. The Cantor set is a nowhere dense subset of $[0,1]$ because it is closed and it contains no open intervals. The set of all rational numbers in $[0,1]$ is not nowhere dense because its closure is all of $[0,1]$. However the set of all rational numbers in $[0,1]$ is meager since it is the countable union of one-point sets.

The reader should think of meager sets as "small". We will see some interesting examples of meager sets when we discuss applications of the Baire category theorem. The "big" sets are simply sets whose complements are small; for instance, the "big" analogue of a nowhere dense set is simply a set which contains a dense open set. Here is the "big" analogue of a meager set:

Definition 26.3. Let X be a topological space. A subset of X is *residual* if it contains a countable intersection of dense open sets.

The Baire category theorem is simply the assertion that every residual set in a complete metric space is dense. Before proving this theorem, we give an alternative characterization of completeness which is of independent interest in metric space theory.

Lemma 26.4. *Let X be a metric space. Then X is complete if and only if for every nested sequence $B_1 \supseteq B_2 \supseteq \dots$ of open balls such that the diameter of B_n tends to 0 as n tends to infinity we have that $\bigcap_n \overline{B_n}$ contains exactly one point.*

Proof. First, suppose X is complete and let $\{B_n\}$ be a nested sequence of open balls. Assume B_n has center x_n and radius r_n, so that $r_n \to 0$; we will show that the sequence $\{x_n\}$ is Cauchy in X. Given $\varepsilon > 0$ choose N large enough so that $n \geq N$ implies $r_n < \varepsilon$. Then if $n \geq m \geq N$ we have $d(x_n, x_m) < \varepsilon$ since $x_n \in B_n \subseteq B_m$, so $\{x_n\}$ is indeed Cauchy. By completeness this sequence has a limit $x \in X$, and in fact $x \in \overline{B_n}$ for every n since x_m is in B_n for all m sufficiently large. Thus $\bigcap_n \overline{B_n}$ is nonempty. Suppose x and y are both in $\overline{B_n}$ for every n; this means that $d(x, y) \leq r_n$ for every n and hence $d(x, y) = 0$ since $r_n \to 0$. This forces $x = y$ and proves that $\bigcap_n \overline{B_n}$ consists of exactly one point.

Conversely, suppose that the nested open balls condition holds for X and let $\{x_n\}$ be a Cauchy sequence in X. The statement that $\{x_n\}$ is Cauchy means that for every $\varepsilon > 0$ there exists N_ε such that $d(x_n, x_m) < \varepsilon$ whenever $n, m \geq N_\varepsilon$. For each k let $M_k = N_{\frac{1}{2^k}}$; we can assume without loss of generality that M_k is a nondecreasing function of K. Let B_k denote the open ball of radius $\frac{1}{2^k}$ centered at $y_k := x_{M_k}$; we will show that B_k is a nested sequence of balls. Given $x \in B_{k+1}$ we have that $d(x, y_{k+1}) < \frac{1}{2^{k+1}}$, and thus $d(x, y_k) \leq d(x, y_{k+1}) + d(y_k, y_{k+1}) < \frac{1}{2^{k+1}} + \frac{1}{2^{k+1}} = \frac{1}{2^k}$. Thus $x \in B_k$, as claimed. By assumption $\bigcap_k \overline{B_k}$ consists of a single point x_∞, and as argued above the sequence of centers of the B_k's converges to x_∞. But the sequence of centers is a subseqence of $\{x_n\}$, and a Cauchy sequence converges to a point if some subsequence converges to that point. This completes the proof. □

We are now ready to prove the Baire category theorem.

Theorem 26.5 (Baire Category Theorem)**.** *Let X be a complete metric space and let $\{U_n\}$ be a countable family of dense open subsets of X. Then $\bigcap_n U_n$ is dense in X.*

Proof. Fix any open ball B_0 in X, and note that $U_n \cap B_0$ is an open sense subset of B_0. Choose any point $x_1 \in U_1 \cap B_0$ and choose r_1 such that the ball $B(x_1, r_1) \subseteq U_1 \cap B_0$. Set $B_1 = B(x_1, \frac{r_1}{2})$ and note that $\overline{B_1} \subseteq U_1$. Working in $\overline{B_1} \cap U_2$ we can choose a ball B_2 such that $B_2 \subseteq B_1$, $\overline{B_2} \subseteq U_2$, and the radius of B_2 is no more than half that of B_1. By induction there is a nested sequence of balls $\{B_n\}$ whose radii tend to 0 such that $\overline{B_n} \subseteq U_n$. By Lemma 26.4 the intersection $\bigcap_n \overline{B_n}$ consists of a single point $x \in B_0$, and $x \in \overline{B_n} \subseteq U_n$ for every n. Thus $\bigcap_n U_n$ intersects B_0, and since B_0 was arbitrary we have shown that $\bigcap_n U_n$ intersects every open ball in X. This means $\bigcap_n U_n$ is dense in X. □

Remark 26.6. The conclusion of the Baire category theorem is purely topological, so it holds for any topological space which is homeomorphic to a complete metric space. For example, the open interval $(0, 1)$ with its standard metric is not complete but the Baire category theorem still holds for $(0, 1)$ since it is homeomorphic to the complete metric space $\mathbb{R}$ (with its standard metric). Of course the proof assumes that a metric has been chosen on X which makes X a complete metric space.

The Baire category theorem has many important consequences in mathematics. We will formulate and prove some of the most interesting applications in the next few lectures, but first we leave as exercises two elementary statements which would be rather difficult to prove without the Baire Category theorem.

Exercise 26.7. Prove that if X is a complete metric space without isolated points then X is uncountable.

Exercise 26.8. Recall that a *Hammel basis* for a (possibly infinite dimensional) vector space V is a linearly independent set $\mathcal{B}$ of vectors in V such that every vector in V is a finite linear combination of vectors in $\mathcal{B}$. Prove that no Banach space has a countable Hammel basis. Deduce that there is no Banach space norm on the vector space of all polynomials with real or complex coefficients.

Any set which is not of first category is often called a set of *second category*. It immediately follows from the Baire Category Theorem that for a set A of second category one can find an open set U such that $A \cap U$ is the residual set in the complete metric space $\bar{U}$.

27. Lecture 27 (10/31/11): Applications of the Baire Category Theorem

The first application that we consider is to the theory of *Diophantine approximation* in number theory. One of the most basic questions in number theory is: how well can a given real number be approximated by rational numbers?

One of the most important results in this direction is a theorem of Liouville which characterizes *transcendental* numbers - that is, numbers like e and π which are not the root of any polynomial equation with rational coefficients - in terms of rational approximations.

Theorem 27.1. *(Liouville theorem) Suppose α is a real number with the property that for every n there exist integers p and $q > 1$ with the property that*

$$0 < \left|\alpha - \frac{p}{q}\right| < \frac{1}{q^n}$$

Then α is transcendental.

This theorem was used to produce the first examples of transcendental numbers (of course mathematicians were well aware of e and π, but it was not until later that anyone could prove that e and π are transcendental). For example, *Liouville's constant* $\sum_{n=1}^{\infty} 10^{-n!}$ is transcendental by Liouville's theorem.

Liouville's theorem is not within the scope of this course although its proof is not difficult. We mentioned it to motivate the following definition:

Definition 27.2. Let $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ be a continuous decreasing function. Say that a real number α is *ϕ-approximable* if there exists integers p, $q > 1$ such that

$$0 < \left|\alpha - \frac{p}{q}\right| < \phi(q)$$

For example, if we set $\phi_n(x) = \frac{1}{x^n}$ then Liouville's theorem becomes the statement that if a real number is ϕ_n-approximable for every n then it is transcendental. Our goal is to prove that for any fixed ϕ the set of all ϕ-approximable numbers is dense in $\mathbb{R}$.

Proposition 27.3. *Let $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ be a continuous decreasing function. Then the set of all ϕ-approximable real numbers is residual.*

Proof. For positive integer N let S_N denote the set of all real numbers x with the property that there exists an integer p and an integer $q \geq n$ such that $0 < \left|x - \frac{p}{q}\right| < \phi(q)$. Note that the set of all ϕ-approximable real numbers is precisely $\bigcap_{N=1}^{\infty} S_N$. If $x \in S_N$ and $0 < \varepsilon < \left|x - \frac{p}{q}\right| < \phi(q)$ where $q \geq N$ then every real number in the open interval $(x-\varepsilon, x+\varepsilon)$ is also in S_N, so S_N is open. Moreover S_N is dense since ϕ is decreasing and the set of all rational numbers $\frac{p}{q}$ with $q \geq N$ is dense, so the set of all ϕ-approximable numbers is the countable intersection of nowhere dense sets, as desired. □

Corollary 27.4. *For any ϕ as above, the set of all ϕ-approximable real numbers is dense.*

Proof. This follows immediately from the previous proposition and the Baire category theorem. □

One of the most striking and historically significant applications of the Baire category theorem is to the existence of continuous functions on the real line which are not differentiable at any point. [1]

The Baire category theorem implies an even more striking statement: not only do continuous nowhere differentiable functions exist, but they are everywhere! More precisely, we will show that the set of all such functions is residual in $C[0,1]$ and hence dense by the Baire category theorem.

Proposition 27.5. *The set of all functions in $C[0,1]$ which are differentiable at some point is of first category.*

Proof. Let C_n denote the set of all functions $f \in C[0,1]$ with the following property: there exists $x_0 \in [0,1]$ and $\delta > 0$ such that $|x - x_0| < \delta$ implies $|f(x) - f(x_0)| \leq n\,|x - x_0|$. Observe that if f is differentiable at some point x_0 and $|f'(x_0)| < n$ then $f \in C_n$, so $\bigcup_n C_n$ contains the set of all continuous functions on $[0,1]$ which are differentiable at some point. C_n is a closed subset of $C[0,1]$, so it suffices to show that C_n contains no open set.

Given a positive real number m, say that a piecewise linear function is *m-oscillating* if the slope of each linear piece is at least m. We claim that the set of all m-oscillating piecewise linear functions is dense in $C[0,1]$. By a uniform continuity argument it suffices to prove that any constant function on an interval $[a,b]$ can be uniformly approximated by m-oscillating functions. Given any $\varepsilon > 0$, we may partition $[a,b]$ into intervals of length $\frac{\varepsilon}{m}$ and construct a m-oscillating function f which is linear on each of these subintervals and which satisfies $c - \varepsilon \leq f(x) \leq c + \varepsilon$ for any prescribed constant c; the uniform distance between f and the constant function with value c is no larger than ε, as desired.

[1]Well into the 19th century many eminent mathematicians believed that no such function could exist - Cauchy even published a (incorrect) proof that they could not! In 1872 Weierstrass surprised the mathematical community by showing that a certain function of the form:

$$f(x) = \sum_{n=0}^{\infty} a^n \cos(b^n \pi x)$$

is continuous but nowhere differentiable. *Comment added by Paul Siegel*

To conclude the proof, observe that for every $f \in C_n$ and any $\varepsilon > 0$ we can choose a m-oscillating function g with $m > n$ and $\|f - g\| < \varepsilon$. Such a function g cannot be in C_n, so C_n does not contain any open ε-ball, as desired. □

Corollary 27.6. *The set of all continuous nowhere differentiable functions on* $[0, 1]$ *is residual in* $C[0, 1]$ *and hence dense.*

Note that in the proof of Proposition 27.5 we actually proved something stronger. We did not need the full strength of the notion of differentiability, just the fact that if a function f is differentiable at a point x_0 then there is a constant C such that $|f(x) - f(x_0)| \leq C\,|x - x_0|$ for all x sufficiently close to x_0. This property of a function is in fact weaker than differentiability, and it generalizes to any metric space.

Definition 27.7. Let X be a metric space and let $f : X \to \mathbb{R}$ be a function on X.

- f is *Lipschitz at* $x_0 \in X$ if there is a constant C such that $|f(x) - f(x_0)| \leq Cd(x, x_0)$ for every $x \in X$.
- f is said to be *Lipschitz* if it is Lipschitz at every point of X.

Thus if f is a function on $[0, 1]$ which is differentiable at $x_0 \in [0, 1]$ then f is Lipschitz near x_0. Our argument in the proof of Proposition 27.5 actually showed that the set of all functions on $[0, 1]$ which are Lipschitz near some point is of first category in $C[0, 1]$. In the next lecture we will generalize this even further.

28. Lecture 28 (11/2/11): Moduli of Continuity

In the last lecture we saw that the set of all continuous nowhere differentiable functions is residual in $C[0, 1]$, and we observed that the proof used only a geometric property of differentiable function called the Lipschitz property. The Lipschitz property can be regarded as a quantitative strengthening of the notion of continuity, and in this lecture we will explore more general properties of this form. We will prove that the set of all functions on a compact metric space X (without isolated points) which satisfy a given quantitative continuity condition is a nowhere dense subset of $C(X)$.

Definition 28.1. Let $\rho : \mathbb{R}^+ \to \mathbb{R}^+$ be a nondecreasing function which satisfies $\rho(t) = 0$ if and only if $t = 0$. Let X be a metric space and let $f : X \to \mathbb{R}$ be a function on X.

- f is said to have *modulus of continuity* ρ *at* $x_0 \in X$ if $|f(x) - f(x_0)| \leq \rho(d(x, x_0))$ for every $x \in X$.
- f is said to have *modulus of continuity* ρ if it has modulus of continuity ρ at every point of X.

Exercise 28.2. Prove that a function on X is uniformly continuous if and only if it has a modulus of continuity.

Thus a function is Lipschitz if it has modulus of continuity $\rho(t) = Ct$ for some constant C. There are other common moduli of continuity; for instance a function is said to be α*-Holder continuous* if it is has modulus of continuity $\rho(t) = Ct^\alpha$ for some constant C, and it is simply *Holder continuous* if it is α-Holder continuous for some α.

The condition that a function have a given modulus of continuity globally can be a very strong condition - for example, an important theorem in analysis asserts that any Lipschitz function on $\mathbb{R}^n$ is differentiable away from a null set. Also, while it is easy to see that any constant function has modulus of continuity ρ for some ρ there need not be any non-constant functions with this property.

Example 28.3. Suppose ρ is a modulus such that $\rho(t) = o(t)$ as $t \to 0$, meaning for every $\varepsilon > 0$ there exists $t_\varepsilon > 0$ such that $\rho(\varepsilon t) < \varepsilon t$ for $t \in (0, t_\varepsilon)$. We shall prove that any continuous function f on $[0,1]$ with modulus of continuity ρ is constant. Given any a in $(0,1]$, choose a partition $\{0 = x_0 < x_1 < \ldots < x_n = a\}$ of $[0,a]$ such that $|x_{i+1} - x_i| < t_\varepsilon$ for each i, and observe that

$$\begin{aligned}
|f(a) - f(0)| &= \left| \sum_{i=0}^{n-1} f(x_{i+1}) - f(x_i) \right| \\
&\leq \sum_{i=0}^{n-1} |f(x_{i+1}) - f(x_i)| \\
&\leq \sum_{i=0}^{n-1} \rho(x_{i+1} - x_i) \\
&< \sum_{i=0}^{n-1} \varepsilon(x_{i+1} - x_i) = \varepsilon a
\end{aligned}$$

Thus $|f(a) - f(0)| < \varepsilon a$ for every a and every ε, which forces $f(a) = f(0)$ for all $a \in [0,1]$.

An immediate consequence of this example is that there are no non-constant α-Holder continuous functions on $[0,1]$ for $\alpha > 1$. However, such functions can exist on other metric spaces (such as the Cantor set).

Having a modulus of continuity ρ at a single point is a much less stringent condition, but our next result shows that functions with this property still make up only a small subset of $C(X)$ when X is a compact metric space without isolated points.

Proposition 28.4. *Fix a modulus $\rho(t)$, and let X be a compact metric space without isoleted points. Then the set of all functions which have modulus of continuity ρ at some point in X is a nowhere dense subspace of $C(X)$.*

Proof. Let f be a function with modulus of continuity ρ at $x_0 \in X$ and let $\varepsilon > 0$ be given. Let x_1 be any point such that $\rho(d(x_0, x_1)) < \frac{\varepsilon}{2}$ and $|f(x_0) - f(x_1)| < \frac{\varepsilon}{4}$. Choose a continuous function g on X which is supported in a small ball centered at x_1, which takes values in $[0,1]$, and which satisfies $g(x_0) = 0$ and $g(x_1) = 1$. We have:

$$(f + \varepsilon g)(x_0) - (f + \varepsilon g)(x_1) = f(x_0) - f(x_1) + \varepsilon g(x_1)$$

Thus $|(f + \varepsilon g)(x_0) - (f + \varepsilon g)(x_1)| \geq \frac{3\varepsilon}{4} > \rho(d(x_0, x_1))$ which implies that $f + \varepsilon g$ does not have modulus of continuity ρ at x_0 even though $f + \varepsilon g \in B_\varepsilon(f) \subseteq C(X)$.

Similar estimates show that $f + \varepsilon g$ fails to have modulus of continuity ρ on an open set containing x_0 and that any function h such that $\|f + \varepsilon g - h\|$ is sufficiently small will also fail to have modulus of continuity ρ on that open set. Thus there is an open set $U_{x_0} \subseteq X$ containing x_0 and an open set $V_{x_0} \subseteq B_\varepsilon(f) \subseteq C(X)$ such that h does not have modulus of continuity ρ on U_{x_0}. Repeating the argument at every point in X, we can cover X by open sets U each of which comes equipped with a (maximal) open set $V \subseteq C(X)$ such that no function in V has modulus of continuity ρ on U. Since X is compact there is a finite subcover $U_1, \dots, U_n$ and corresponding open sets of functions $V_1, \dots, V_n$ such that every $f \in V_i$ fails to have modulus of continuity ρ on U_i. The set $\bigcap_i V_i$ is an open and dense subset of $C(X)$ consisting of functions which do not have modulus of continuity ρ at any point in X, proving that the set of all functions which do have modulus of continuity ρ is nowhere dense. □

In a future lecture we will prove a statement that in a sense is complementary : if A is a set of functions on X which have the same modulus of continuity ρ *everywhere* and which are uniformly bounded - meaning there is a constant C such that $\|f\| \leq C$ for every $f \in A$ - then A is a compact subset of $C(X)$. In fact, every compact subset of $C(X)$ has this form. This is a version of the *Arzela-Ascoli theorem*, and it is extremely useful in functional analysis and its applications.

29. Lecture 29 (11/2/11): Proof of Stone-Weierstrass Theorem

In this lecture we temporarily set aside our discussion of moduli of continuity to prove the Stone-Weierstrass theorem (Theorem 21.2). In the case of real-valued functions this theorem states that any subalgebra $\mathcal{A}$ of $C(X, \mathbb{R})$ which contains a nonzero constant function and separates points is dense. Recall that a subalgebra of $C(X, \mathbb{R})$ is a linear subspace which is closed under pointwise multiplication of functions, and the condition that a collection of functions separates points means that for every pair of distinct points x, y in X there is a function f in the collection such that $f(x) \neq f(y)$. There is a corresponding statement about subalgebras of $C(X, \mathbb{C})$ so long as those subalgebras are closed under complex conjugation, and the complex version follows from the real version by approximating the real and imaginary parts of a function.

Proof of Theorem 21.2.

Step 1: If $f \in \mathcal{A}$ then $|f| \in \bar{\mathcal{A}}$.

Notice that $|x| = \sqrt{x^2}$ and recall that the function $\sqrt{x}$ in the uniform limit on any closed interval $[0, N]$ of functions $\sqrt{x+t}$ as $t \to 0$. Consider the Taylor series of the function $\sqrt{x+t}$ at a point $T > 0$. Its radius of convergence is $T + t$ and hence it converges uniformly on any interval $[-t + \epsilon, 2T + t - \epsilon]$ for any $\epsilon > 0$. Thus an appropriate Taylor polynomial $P(x)$ of this series uniformly approximates $\sqrt{x}$ on the interval $[0, 2T]$. Now take $T > \max |f|$. Then the function $P(f^2) \in \mathcal{A}$ uniformly approximates $|f|$.

Step 2: If $f, g \in \mathcal{A}$, then $\max(f, g)$ and $\min(f, g)$ are in $\overline{\mathcal{A}}$.

This follows immediately from Step 1, using the formulae

$$\max(f,g) = \frac{f+g}{2} + \frac{|f-g|}{2}$$
$$\min(f,g) = \frac{f+g}{2} - \frac{|f-g|}{2},$$

Step 3: Given any continuous function $f : X \to \mathbb{R}$, $x \in X$ and $\varepsilon > 0$ there exists a function $g_x \in \overline{\mathcal{A}}$ such that $g_x(x) = f(x)$ and $g_x(t) > f(t) - \varepsilon$ for all $t \in X$.

By the separation property and inclusion of constants in $\mathcal{A}$ for any $y \in X$ there exists $h_y \in \mathcal{A}$ such that $h_y(x) = f(x)$ and $h_y(y) = f(y)$. Since h_y is continuous, there exists an open neighborhood $I_y \ni y$ such that $h_y(t) > f(t) - \varepsilon$ for all $t \in I_y$.

Since X is compact the cover of X by open sets I_y contains a finite subcover $I_{y_1}, I_{y_2}, \ldots, I_{y_n}$. Any $x \in X$ lies in some I_{y_k}. Take $g_x = \max(h_{y_1}, \ldots, h_{y_n})$. Then $g_x \in \overline{\mathcal{A}}$ with the required property.

Step 4: Given any continuous function $f : X \to \mathbb{R}$, $x \in X$ and $\varepsilon > 0$ there exists a function $h \in \overline{\mathcal{A}}$ such that $f(t) - \varepsilon < h(t) < f(t) + \varepsilon$ for all $t \in X$.

Now look at the collection $\{g_x\}$ constructed in Step 3. Since each function is continuous, by for each $x \in X$ there exists an open neighborhood $J_x \ni x$ on which $g_x(t) < f(t) + \varepsilon$. Use compactness and choose a finite subcover $J_{x_1}, J_{x_2}, \ldots, J_{x_m}$. Take $h = \min(g_{x_1}, \ldots, g_{x_m})$. Then $h \in \overline{\mathcal{A}}$ with the required property, and since also $h(t) > f(t) - \varepsilon$ for all $t \in X$, we have $\|f - h\| < \varepsilon$, proving the result. □

30. Lecture 30 (11/9/11): Compactness in $C(X)$

Our goal in this lecture is to characterize the compact subsets of $C(X)$ where X is a compact metric space. Our characterization is a generalization of the classical Arzela-Ascoli theorem, an extremely powerful result in analysis with numerous applications in functional analysis, geometry, differential equations, and other areas. Informally the theorem asserts that the precompact subsets of $C(X)$ are precisely the sets of uniformly bounded functions with a uniform constraint on their oscillation; the constraint on the oscillation can take the form of a modulus of continuity common to all functions in the set.

Here is some relevant terminology:

Definition 30.1. Let X be a metric space and let A be a subset of $C(X)$.

- A is *uniformly bounded* if there is a constant C such that $|f(x)| \leq C$ for every $f \in A$ and every $x \in X$.
- A is *equicontinuous* if for every $\varepsilon > 0$ there exists δ such that for any $f \in A$ we have $|f(x) - f(y)| < \varepsilon$ whenever $d(x, y) < \delta$

Exercise 30.2. Show that the set of functions $\{\sin(nx)\}_{n\in\mathbb{N}}$ is a uniformly bounded subset of $C[0, 1]$ which is not equicontinuous.

Exercise 30.3. Show that $A \subseteq C(X)$ is equicontinuous if and only if there is a modulus ρ such that every function in A has modulus of continuity ρ.

Recall that a subset of a topological space is precompact if its closure is compact. We will need the following characterization of precompact subsets of an arbigtrary metric space:

Lemma 30.4. *Let Y be a complete metric space. A subset $K \subseteq Y$ is precompact if and only if for every $\varepsilon > 0$ there is a set $p_1, \ldots, p_n$ in K such that for every $y \in K$ there exists i such that $d(y, y_i) < \varepsilon$.*

Proof. Suppose K is precompact. Consider the collection of open balls $B(y, \frac{\varepsilon}{2})$ where y ranges over all points in the closure $\overline{K}$; this is an open cover of $\overline{K}$ so by compactness there is a finite subcover $B_1, \ldots, B_n$. Since K is dense in $\overline{K}$ there are points $p_1, \ldots, p_n$ such that $p_i \in B_i \cap K$. Given any $y \in K$ we have $y \in B_i$ for some i since $B_1, \ldots, B_n$ covers K and thus $d(y, p_i) < \varepsilon$ since y and p_i lie in the same ball of radius $\frac{\varepsilon}{2}$.

Conversely, suppose that K can be approximated arbitrarily well by finite sets in the sense described in the lemma; we will show that every sequence $\{y_k\}$ in K has a Cauchy subsequence. Cover K by finitely many balls of radius 1 and let B_1

be a ball which contains infinitely many of the y_k's. Given a ball B_N of radius $\frac{1}{N}$ which contains infinitely many of the y_k's we can cover B_N by balls of radius $\frac{1}{N+1}$ and choose a ball B_{N+1} which also contains infinitely many of the y_k's. Thus by induction we have shown that there is a nested sequence of balls $B_1 \supseteq B_2 \supseteq \dots$ such that each B_N contains infinitely many of the y_k's and the radius of B_N is $\frac{1}{N}$. Define a subsequence y_{k_N} by letting k_N denote the smallest value of k such that $y_k \in B_N$; this subsequence is clearly Cauchy. □

Theorem 30.5 (Arzela-Ascoli)**.** *Let X be a compact metric space and let A be a subset of $C(X)$. Then A is precompact if and only if it is uniformly bounded and equicontinuous.*

Proof. First let us assume that A is precompact. A precompact subset of a metric space is necessarily bounded, and a subset of $C(X)$ is bounded in norm if and only if it is uniformly bounded as a set of functions. So it suffices to show that A is equicontinuous. Given any $\varepsilon > 0$ let $f_1, \dots, f_n$ be a set of functions such that every $f \in A$ satisfies $\|f - f_i\| < \frac{\varepsilon}{3}$ for some i. Since each f_i is continuous (and therefore uniformly continuous since X is compact) there exists $\delta > 0$ such that $|f_i(x) - f_i(y)| < \frac{\varepsilon}{3}$ whenever $d(x, y) < \delta$. Hence for any $f \in A$ and any x, y such that $d(x, y) < \delta$ we have:

$$|f(x) - f(y)| \le |f(x) - f_i(x)| + |f_i(x) - f_i(y)| + |f_i(Y) - f(y)| < \varepsilon$$

Thus A is equicontinuous.

Conversely, assume that A is uniformly bounded and equicontinuous. Let $\varepsilon > 0$ be given, and let $\delta > 0$ be a number with the property that $|f(x) - f(y)| < \frac{\varepsilon}{3}$ whenever $d(x, y) < \delta$ for any $f \in A$. By compactness there is a set $x_1, \dots, x_n$ such that the balls $B(x_i, \delta)$ cover X. Call an n-tuple $(t_1, \dots, t_n) \in \mathbb{R}^n$ *admissible* if there is a function $f \in A$ such that $f(x_i) = t_i$ for each i; the set T of all admissible n-tuples is a bounded (and therefore precompact) subset of $\mathbb{R}^n$ since A is uniformly bounded. Cover T by a finite set of balls of radius $\frac{\varepsilon}{3}$ in the ℓ^∞ norm centered at $\mathbf{t}_1, \dots, \mathbf{t}_m \in \mathbb{R}^n$ and choose functions $f_1, \dots, f_m$ in A such that $\mathbf{t}_j = (f_j(x_1), \dots, f_j(x_n))$.

Given any $f \in A$ the n-tuple $(f(x_1), \dots, f(x_n))$ is in T, so it lies within $\frac{\varepsilon}{3}$ of some $\mathbf{t}_j$ and hence $|f(x_i) - f_j(x_i)| < \frac{\varepsilon}{3}$ for each i. We claim that $\|f - f_j\| < \varepsilon$. Indeed, for any $x \in X$ there exists x_i such that $d(x, x_i) < \delta$, and thus:

$$|f(x) - f_j(x)| \le |f(x) - f(x_i)| + |f(x_i) - f_j(x_i)| + |f_j(x_i) - f_j(x)| < \varepsilon$$

Thus we have shown that A can be approximated arbitrarily well by finite sets, and we conclude that A is precompact. □

We may encounter some applications of the Arzela-Ascoli theorem later on, but for now we give a simple example which illustrates how the theorem is used in practice.

Example 30.6. Let $M > 0$ be a constant and let A denote the set of all differentiable functions f on $[0, 1]$ such that $f(0) = 0$ and $|f'(x)| \le M$ for all x. The set A is uniformly bounded by the fundamental theorem of calculus: if $f \in A$ then $f(x) = \int_0^x f'(t)dt \le Mx \le M$ for $x \in [0, 1]$. Moreover it is equicontinuous by the mean value theorem: given any $\varepsilon > 0$ set $\delta = \frac{\varepsilon}{M}$ and observe that for $f \in A$ we have $|f(x) - f(y)| \le M\,|x - y| < \varepsilon$ whenever $|x - y| < \delta$. Thus A is precompact by the Arzela-Ascoli theorem.

Notice however that this set is not compact because it is not closed in the uniform topology. it is immediate that any function in $\bar{A}$ is M-Lipschitz. With a little effort one can show that any such function can be uniformly approximated by differentiable functions with $|f'(x)| \leq M$

This example is particularly useful in the theory of differential equations. For certain differential equations one can use an iterative procedure to construct a sequence of functions which approximates a solution in an appropriate sense, and if there is a uniform bound on the derivatives of the approximate solutions then one can use the Arzela-Ascoli theorem to extract a uniformly convergent subsequence and prove that the limit is an actual solution. The interested reader may wish to consult the *Peano existence theorem* to see this in action.

31. Lecture 31 (11/11/11): Introduction to Weak* Compactness

In the next few lectures we will discuss another important compactness result in infinite dimensional spaces called the *Banach-Alaoglu theorem*. This result was introduced in Lecture 25 following our discussion of the weak and weak* topologies in Banach space theory; the reader may wish to review this material now. Here is the statement of the theorem:

Theorem 31.1 (Banach-Alaoglu)**.** *Let V be a normed space. Then the closed unit ball in V^* equipped with the weak* topology is compact.*

We have seen examples where the unit ball in V^* is not compact in the norm topology: for example, if V is an infinite dimensional separable Hilbert space then any orthonormal basis $\{\ell_n\}$ satisfies $\|\ell_n - \ell_m\| = 1$ and hence $\{\ell_n\}$ cannot have a norm convergent subsequence. But $\ell_n \to 0$ in the weak* topology, so there is hope that Theorem 31.1 is true. We will prove it in the case where V is separable to avoid complications related to the axiom of choice.

The Banach-Alaoglu theorem has many important applications, but mainly we will be interested in its implications in the (convex) geometry of infinite dimensional Banach spaces. Recall that our main tool for studying finite dimensional normed spaces was the notion of a supporting hyperplane, and we proved that supporting hyperplanes exist using the extreme value theorem together. The same basic idea works on the unit ball in the dual of an infinite dimensional normed space.

Corollary 31.2. *Let V be a normed space and let $i : V \to V^{**}$ be the canonical embedding. Any $\ell \in i(V)$ attains its maximum on the unit ball in V^*.*

Proof. By definition ℓ is continuous relative to the weak* topology and the unit ball in V^* is compact in that topology by the Banach-Alaoglu theorem, so ℓ attains its maximum by the extreme value theorem. □

As a side note, this yields another proof that $C[0,1]$ is not reflexive. If it were then there would exist a Banach space V such that $C[0,1] = V^*$, and the canonical embedding $i : V \to C[0,1]^*$ would be an isomorphism. Thus any continuous linear functional on $C[0,1]^*$ would attain its maximum on the unit ball in $C[0,1]$ at a continuous function f. But it is not hard to show that the continuous linear functional $\ell(f) = \int_0^{\frac{1}{2}} f(x)dx - \int_{\frac{1}{2}}^1 f(x)dx$ does not attain its maximum at any continuous function.

More importantly, the corollary above will allow us to prove an infinite dimensional analogue of the statement that every convex body in $\mathbb{R}^n$ is the convex hull of its extreme points (Corollary 15.8).

Theorem 31.3 (Krein-Millman)**.** *If V is a normed space and $C \subseteq V^*$ is convex set which is compact in the weak* topology then the convex hull of the set of all extreme points of C is dense in C.*

This theorem may not seem very surprising at first, but it is extremely poweful: it is not obvious at the outset that C has *any* extreme points, but the Krein-Millman theorem says that C has enough extreme points to determine C completely.

Exercise 31.4. Show that the set of all $f \in C[0,1]$ such that $\|f\| = 1$ and

$$\int_0^1 f(x)dx = 0$$

is a closed convex set with no extreme points.

32. Lecture 32 (11/16/11): More on the Weak* Topology

Before proving our main results about weak* compact sets (Theorem 31.1 and Theorem 31.3) we make an important connection between weak and strong notions of boundedness. Recall that a subset of a normed space is bounded in norm if the distance between any two points in the set is bounded by some fixed constant.

Definition 32.1. Let V be a normed space and $S \subseteq V$ be a subset of V. Say that S is *weakly bounded* if there is a constant R such that $|\ell(\mathbf{x})| \leq R$ for every $\ell \in V^*$.

Proposition 32.2. *Any weakly bounded subset of a normed space V is bounded in norm.*

Proof. Suppose $S \subseteq V$ is a subset which is unbounded in norm; we will show that S is not weakly bounded. Let $i : V \to V^{**}$ be the canonical embedding, and consider the subset $i(S) \subseteq V^{**}$. We begin by proving that $i(S)$ is unbounded on any ball in V^*.

Since S is unbounded in norm there is a sequence $\{\mathbf{x}_n\}$ in S with the property that $\|\mathbf{x}_n\| \to \infty$. For each n there is a linear functional ℓ_n on V of norm $\frac{R}{2}$ such that $\ell_n(\mathbf{x}_n) = \frac{R}{2}\|\mathbf{x}_n\|$ by the Hahn-Banach theorem, so $i_{\mathbf{x}_n}(\ell_n) = \frac{R}{2}\|\mathbf{x}_n\| \to \infty$ and thus the sequence $i_{\mathbf{x}_n}$ is unbounded on the ball $B_R(0) \subseteq V^*$. Clearly the same sequence is unbounded on $B_R(0) + \ell$ for any fixed $\ell \in V^*$, so S is unbounded on any ball in V^* as claimed.

Now we show that there is a linear functional in V^* which is unbounded on S. Since S is unbounded in norm the same holds for $i(S)$ and thus there is a point $\mathbf{x}_0 \in S$ and a linear functinoal $\ell_0 \in V^*$ such that $\ell_0(\mathbf{x}_0) > 1$. Since the expression $\ell(\mathbf{x}_0)$ depends continuously on ℓ, there is a ball $B_0 \subseteq V^*$ centered at ℓ_0 such that $\ell(\mathbf{x}_0) > 1$ for every $\ell \in B_0$. We showed that $i(S)$ is unbounded on every ball in V^*, so there exists $\mathbf{x}_1 \in S$ and $\ell_1 \in B_0$ such that $\ell_1(\mathbf{x}_1) > 2$; as before there is an entire ball centered $B_1 \subseteq B_0$ centered at ℓ_1 such that $\ell(\mathbf{x}_1) > 2$ for every $\ell \in B_1$. By induction there is a sequence $\{\mathbf{x}_n\}$ in S and a collection of balls $\{B_n\}$ such that $\ell(\mathbf{x}_n) > n+1$ for every $\ell \in B_n$; furthermore we can ensure that the diameter of B_n tends to 0 as n tends to inifinity.

Recall that V^* is complete even if V is not, so by Lemma 26.4 we have that $\bigcap_n B_n = \{\widetilde{\ell}\}$. The linear functional $\widetilde{\ell}$ satisfies $\widetilde{\ell}(\mathbf{x}_n) > n+1$ for every n, so $\widetilde{\ell}$ is unbounded on S. Hence S cannot be weakly bounded. $\square$

There is an important variation on this proposition with a very similar proof called the *principle of uniform boundedness*. It asserts that if V is a Banach space and S is a pointwise bounded subset of V^*, meaning that for each $\mathbf{x} \in V$ there is a number $C_{\mathbf{x}}$ such that $\ell(\mathbf{x}) \leq C_{\mathbf{x}}$ for every $\ell \in S$, then S is uniformly bounded in the sense that there is a constant C such that $\|\ell\| \leq C$ for every $\ell \in S$. The previous proposition follows from the principle of uniform boundedness by observing that if $S \subseteq V$ is weakly bounded then $i(S)$ is a pointwise bounded set of linear functionals in V^{**}.

Exercise 32.3. Use the Baire category theorem to prove the principle of uniform boundedness. Hint: given a pointwise bounded set $S \subseteq V^*$, define $V_n = \{\mathbf{x} \in V : \sup_{\ell \in S} |\ell(\mathbf{x})| \leq n\}$. Each V_n is closed and by hypothesis $V = \bigcup_n V_n$, so some V_n contains an open ball by the Baire category theorem. Deduce from this that S is uniformly bounded.

The principle of uniform boundedness is quite useful in the theory of bounded linear operators between Banach spaces, but we will not need it any further in this course.

We now turn our attention to Theorem 31.1. The proof of this theorem requires the axiom of choice in its fullest generality, but we can give a more direct argument in the separable case. This is made possible by the fact that the weak* topology on the closed unit ball B^* in the dual V^* of a separable Banach space V is actually metrizable. Let $\{\mathbf{x}_n\}$ be a countable dense set in V and define:

$$d(\ell, \ell') = \sum_{n=1}^{\infty} \frac{1}{2^n} |\ell(x_n) - \ell'(x_n)|$$

Of course the constants $\frac{1}{2^n}$ are not crucial: any summable sequence of positive real numbers does the job.

Exercise 32.4. Show that d defines a translation invariant metric on B^*.

Proposition 32.5. *The metric topology on B^* determined by d coincides with the weak* topology on B^*.*

Proof. Recall that the metric topology on B^* is generated by open metric balls while the weak* topology is generated by the weak* open sets:

$$\{\ell \in B^* : |\ell(y_i) - \ell_0(y_i)| < \varepsilon, 1 \leq i \leq k\}$$

where ℓ_0, ε, and $y_1, \ldots, y_k$ are given. Since both the metric topology and the weak* topology are translation invariant, it suffices to show that every open metric ball centered at 0 contains a weak* open subset of B^* which contains 0 and that every weak* open subset of B^* which contains 0 contains an open metric ball centered at 0.

So let U be the open metric ball of radius ε centered at 0. Choose n large enough so that $\frac{1}{2^n} < \varepsilon$ and consider the weak* open set

$$U' = \{\ell \in B^* : |\ell(x_i)| < \frac{\varepsilon}{2}, 1 \leq i \leq k\}$$

where the x_i's are elements of the countable dense set used to define the metric d. For $\ell \in U'$ we have:

$$\begin{aligned} d(\ell, 0) &= \sum_{i=1}^{\infty} \frac{1}{2^i} |\ell(\mathbf{x}_i)| \\ &= \sum_{i=1}^{k} \frac{1}{2^i} |\ell(\mathbf{x}_i)| + \sum_{i=k+1}^{\infty} \frac{1}{2^i} |\ell(\mathbf{x}_i)| \\ &< \frac{\varepsilon}{2} \sum_{i=1}^{k} \frac{1}{2^i} + \sum_{i=k+1}^{\infty} \frac{1}{2^i} < \varepsilon \end{aligned}$$

Thus $0 \in U' \subseteq U$.

Now consider the weak* open set

$$V = \{\ell \in B^* : |\ell(y_i)| < \varepsilon, 1 \le i \le k\}$$

Since $\{x_n\}$ is dense we can choose $x_{n_1}, \ldots, x_{n_k}$ in such a way that $\|y_i - x_{n_i}\| < \frac{\varepsilon}{2}$ for $1 \le i \le k$. Choose δ so that $\delta < \frac{\varepsilon}{2^{n_i+1}}$ for each i. If $d(\ell, 0) < \delta$ then $|\ell(\mathbf{x}_i)| < 2^i \delta$ by the definition of d, so in particular $|\ell(\mathbf{x}_{n_i})| < \frac{\varepsilon}{2}$ by our choice of δ. Thus for $1 \le i \le k$ we have:

$$\begin{aligned} |\ell(\mathbf{y}_i)| &\le |\ell(\mathbf{y}_i - \mathbf{x}_{n_i})| + |\ell(\mathbf{x}_{n_i})| \\ &< \|\ell\|^* \|\mathbf{y}_i - \mathbf{x}_{n_i}\| + \frac{\varepsilon}{2} < \varepsilon \end{aligned}$$

It follows that the metric ball of radius δ centered at 0 is contained in V, as desired. ◻

In the next lecture we will use this metric to prove that B^* is weak* compact.

33. Lecture 33 (11/17/11) - Weak* Compactness

In this lecture we finish the proof of Theorem 31.1 - the assertion that the closed unit ball B^* in the dual V^* of a normed space V is weak* compact - under the assumption that V is separable. The advantage of separability is that the weak* topology on B^* is metrizable as we saw in the last lecture and compactness for metrizable spaces can be checked sequentially. We will also use the following straightforward exercise involving the weak* topology:

Exercise 33.1. Let V be a normed space $L \subseteq V$ a dense subspace, and $\{\ell_n\}$ a bounded sequence in V^*. If $\ell_n(\mathbf{x})$ converges to $\ell(\mathbf{x})$ for every $\mathbf{x} \in L$ where $\ell \in V^*$ then ℓ_n converges to ℓ in the weak* topology.

Proof of Theorem 31.1 when V is separable. Fix a dense subset $\{\mathbf{x}_n\}$ of the unit ball $B \subseteq V$. Since the unit ball $B^* \subseteq V^*$ is metrizable by Proposition 32.5, it suffices to show that every sequence $\{\ell_n\}$ in B^* has a convergent subsequence. Observe that the set $\{\ell_n(\mathbf{x}_1)\}$ is a bounded set of real numbers, so there is a convergent subsequence since closed and bounded subsets of $\mathbb{R}$ are compact. Let A_1 denote the index set for this subsequence, so that A_1 is an inifinte subset of $\mathbb{N}$ and $\{\ell_n(\mathbf{x}_1) : n \in A_1\}$ converges to a limit a_1. Choose $n_1 \in A_1$ so that $|\ell_{n_1}(\mathbf{x}_1) - a_1| < 1$.

Now consider the set $\{\ell_n(\mathbf{x}_2) : n \in A_1\}$. This again is a bounded set of real numbers, so there is an infinite subset $A_2 \subseteq A_1$ such that $\{\ell_n(\mathbf{x}_2) : n \in A_2\}$ is a convergent sequence with limit a_2. Thus we can choose n_2 in such a way that $n_2 > n_1$, $|\ell_{n_2}(\mathbf{x}_1) - a_1| < \frac{1}{2}$ and $|\ell_{n_2}(\mathbf{x}_2) - a_2| < \frac{1}{2}$. Proceeding inductively we can construct a chain of infinite sets $\mathbb{N} \supseteq A_1 \supseteq A_2 \supseteq \ldots$ with the property that $\{\ell_n(\mathbf{x}_k) : n \in A_k\}$ converges to a real number a_k, and we can construct a chain of numbers $n_1 < n_2 < \ldots$ such that $|\ell_{n_k}(\mathbf{x}_i) - a_i| < \frac{1}{k}$ for $1 \leq i \leq k$.

It follows from this construction that $\ell_{n_k}(\mathbf{x}_n)$ converges to a_n as k tends to infinity for each fixed n. Since $\{\mathbf{x}_n\}$ is dense in V a standard $\frac{\varepsilon}{3}$ argument shows that $\ell_{n_k}(\mathbf{y})$ is a Cauchy sequence for every $\mathbf{y} \in V$, so we can define a function $\widetilde{\ell} : V \to \mathbb{R}$ by $\widetilde{\ell}(\mathbf{y}) = \lim_k \ell_{n_k}(\mathbf{y})$. $\widetilde{\ell}$ is linear and bounded in norm by 1 by the corresponding properties of the ℓ_{n_k}'s, so ℓ_{n_k} converges to $\widetilde{\ell}$ in the weak* topology by 33.1. This completes the proof. $\square$

Here is a somewhat fancier way to conceptualize this argument. For each $\mathbf{x} \in B$, let $I_{\mathbf{x}}$ denote a copy of the interval $[-1, 1]$ in the real line marked by the point $\mathbf{x}$ and let $P = \prod_{\mathbf{x} \in B} I_{\mathbf{x}}$ equipped with the product topology. The map $\phi : B^* \to P$ given by $\phi(\ell) = \prod_{\mathbf{x} \in B} (\ell(\mathbf{x}))_{\mathbf{x}}$ continuously embeds B^* with the weak* topology as a closed subspace P, and thus the result follows from the *Tychonoff theorem* which asserts that the product of arbitrarily many compat spaces is compact. The general form of the Tychonoff theorem requires the axiom of choice, but there are more concrete proofs for countable products of metric spaces that involve an analogue of the metric introduced in the previous lecture and the "diagonal" argument explained above.

We have now shown that closed and bounded subspaces of V^* are weak* compact. The reader is warned, however, that his or her finite dimensional intuition still does not always apply to the infinite dimensional setting; for example, the unit ball in the dual of any normed space is weak* compact but the unit sphere in an infinite dimensional Hilbert space is not compact because any orthonormal basis converges weakly to 0 as we saw in an earlier lecture. It turns out that there is a deep relationship between compactness and convexity in the weak* topology and thus the "flaw" with the unit sphere is its lack of convexity.

The interaction between compactness and convexity is well-illustrated by Krein Millman theorem which asserts that a convex weak* compact set is determined by its extreme points. We begin with a modest first step in this direction:

Proposition 33.2. *If V is a separable normed space then any convex weak* compact set $S \subseteq V^*$ has an extreme point.*

Proof. In this argument it will be convenient to break from our usual notation by using ℓ to represent elements of $i(V) \subseteq V^{**}$ and x to reprsent elements of $S \subseteq V^*$.

Let $\{\ell_k\}$ be a countable dense subset of $i(V)$ where i is the canonical embedding. Sicne S is weak* compact we have that ℓ_1 attains its maximum value M_1 at a point in S; let $S_1 = \{x \in S : \ell_1(x) = M_1\}$. This is a convex weak* compact set as well, so ℓ_2 attains its maximum value M_2 at a point in S_1 and we may consider the convex weak* compact set $S_2 = \{x \in S_1 : \ell_2(x) = M_2\}$. Iterating this procedure we obtain a chain of convex weak* compact sets $S \supseteq S_1 \supseteq S_2 \supseteq \ldots$ with the property that ℓ_k is constant on S_k and its constant value is its maximum value on S_{k-1}. Note that this chain of closed sets has the finite intersection property (the intersection of any finite subcollection is nonempty), so the intersection $\bigcap_k S_k$ is nonempty since S is compact.

I claim that any $\widetilde{x} \in \bigcap_k S_k$ is an extreme point of S. Suppose $\widetilde{x} = \frac{y+z}{2}$ where $y, z \in S$. If either y or z were not in S_1 then it would mean that $\ell_1(y) < M_1$ or $\ell_1(z) < M_1$ since M_1 is the maximum value of ℓ_1 on S, but in either case we would have $\ell_1(\widetilde{x}) = \frac{1}{2}\ell_1(y) + \frac{1}{2}\ell_1(z) < M$, contradicting the assumption that $\widetilde{x} \in S_1$. The same argument shows that y and z lie in S_k whenever they lie in S_{k-1}, so by induction we have that y and z lie in $\bigcap_k S_k$. But each ℓ_k is constant on $\bigcap_k S_k$, implying that $\ell_k(\widetilde{x}) = \ell_k(y) = \ell_k(z)$ for every k. Since the ℓ_k's are dense in $i(V)$ we have that $\ell(\widetilde{x}) = \ell(y) = \ell(z)$ for every $\ell \in i(V)$ by continuity. But $i(V)$ separates points in V^*, so it follows that $\widetilde{x} = y = z$. We conclue that $\bigcap_k S_k$ consists of exactly one point, and that point is an extreme point of S. □

Exercise 33.3. Use transfinite induction to adapt this proof to the case where V is not separable.

34. Lecture 34 (11/18/11): Proof of Krein-Millman

We are now ready to complete the proof of Theorem 31.3, the Krein-Millman theorem. Given a normed space V and a subset $S \subseteq V^*$, use the notation $E(S)$ for the set of extreme points of S and $C(E(S))$ for the convex hull of $E(S)$. The Krein-Millman theorem is simply the assertion that $C(E(S))$ is dense in S. Our proof will imitate that of Proposition 33.2 with a subtle difference: instead of using supporting hyperplanes as above we will need to use hyperplanes which separate a convex weak* compact set from a point. It is tempting to brazenly assert that such hyperplanes exist by the Hahn-Banach theorem, but the Hahn-Banach theorem as we stated it will only produce a linear functional in V^{**} instead of $i(V)$ as we need. Fortunately our proof of the Hahn-Banach theorem can easily be adapted to the stronger conclusion, and we leave the details to the reader.

Proof of Theorem 31.3 when V is separable. Let S' be the closure of $C(E(S))$; it is clear that $S' \subseteq S$ and thus S' is a convex weak* compact set. Suppose there is a point $x \in S$ which is not in S'. By a weak* variation on the Hahn-Banach theorem aluded to above, there exists $\ell \in i(V)$ such that $\ell(y) < \ell(x)$ for every $y \in S'$. Let M denote the maximum value attained by ℓ on S and consider the set $T = \{z \in S : \ell(z) = M\}$; this is a convex weak* compact subset of S which does not intersect S'. But T has an extreme point by Proposition 33.2, and any such point is an extreme point of S by the proof of that proposition. This gives a contradiction. □

This concludes our discussion of weak* compactness.

35. The Riesz Representation Theorem Revisited

Recall that we constructed an isometric isomorphism $BV[0,1] \to C[0,1]^*$ which sends a function ϕ of bounded variation to the bounded linear functional $\ell_\phi(f) = \int_0^1 f d\phi$, the ϕ-Riemann-Stieltjes integral. In our last few lectures we aim to generalize this construction to compact Hausdorff spaces more complicated than $[0,1]$. As in this more basic case we will identify bounded linear functionals on $C(X)$ with generalized integrals, but it will take some care to set up a sensible theory of integration on an arbitrary compact metric space.

Let us begin with a space which, from a certain point of view, is actually a little simpler than $[0,1]$: the Cantor set. Identify the cantor set K with the countable product $\{0,1\}^{\mathbb{N}}$, and consider the *cylinder* associated to a prefix $\mathbf{a} = (a_1, \ldots, a_n)$. This is the set $C_{\mathbf{a}}$ of all sequences $(x_1, x_2, \ldots)$ with the property that $x_i = a_i$ for $1 \leq i \leq n$. The cylinders associated to prefixes of length n yield a natural partition of the Cantor set into n disjoint open sets, and it is natural to define the integral of a function f which is constant on each of these cylindars by $\int_K f = \sum_{\mathbf{a}\in\{0,1\}^n} f(C_{\mathbf{a}})$. We can even choose a function ϕ which assigns different "weights" to the various cylindars and thereby construct a "Riemann-Stieltjes integral" $f \mapsto \int_K f d\phi$. As with the Riesz representation theorem, all of $C(K)^*$ can be recovered this way.

This suggests how one might handle even more general spaces.

Definition 35.1. Let X be a compact metric space and let $\ell \in C(X)^*$. Say that a set $A \subseteq X$ is ℓ-integrable if for every $\varepsilon > 0$ there exist continuous functions ϕ^+ and ϕ^- such that $\phi^- < \chi_A < \phi^+$ where χ_A is the characteristic function of A and $\ell(\phi^+) - \ell(\phi^-) < \varepsilon$.

Example 35.2. Consider the linear functional $\ell(f) = \int_0^1 f(x)dx$ on $C[0,1]$. Then the interval $[\frac{1}{2}, 1]$ is ℓ-integrable thanks to the functions:

$$\phi^-(x) = \begin{cases} 0 & 0 \leq x \leq \frac{1}{2} \\ n(x - \frac{1}{2}) & \frac{1}{2} \leq x \leq \frac{1}{2} + \frac{1}{n} \\ 1 & \frac{1}{2} + \frac{1}{n} \leq x \leq 1 \end{cases}$$

$$\phi^+(x) = \begin{cases} 0 & 0 \leq x \leq \frac{1}{2} - \frac{1}{n} \\ n(x - \leq \frac{1}{2} + \frac{1}{n}) & \frac{1}{2} - \frac{1}{n} \leq x \leq \frac{1}{2} \\ 1 & \frac{1}{2} \leq x \leq 1 \end{cases}$$

Example 35.3. Let X denote the disk of radius 2 in the plane and let C denote the unit circle in X. Define $\ell \in C(X)^*$ by $\ell(f) = \int_C f$. Then C is not ℓ-integrable because the only continuous function on X smaller than χ_C is the 0 function while any continuous function f larger than χ_C must satisfy $\int_C f \geq 2\pi$.

Our hope is that in a sufficiently wide variety of circumstances we can partition our metric space X into small ℓ-integrable pieces and then imitate the Riemann sum construction to define a notion of integral well-adapted to ℓ. We will continue this program in the final few lectures.