MATHEMATICS 3103 (Functional Analysis)
YEAR 2012–2013, TERM 2

HANDOUT #3: INTRODUCTION TO NORMED LINEAR SPACES

Already in Handout #1 I gave the definition of a normed linear space:

Definition 3.1 *Let X be a vector space over the field $\mathbb{R}$ of real numbers (or the field $\mathbb{C}$ of complex numbers). Then a* **norm** *on X is a function that assigns to each vector $x \in X$ a real number $\|x\|$, satisfying the following four conditions:*

(i) $\|x\| \geq 0$ *for all* $x \in X$ *(***nonnegativity***);*

(ii) $\|x\| = 0$ *if and only if* $x = 0$ *(***nondegeneracy***);*

(iii) $\|\lambda x\| = |\lambda| \, \|x\|$ *for all* $x \in X$ *and all* $\lambda \in \mathbb{R}$ *(or* $\mathbb{C}$*) (***homogeneity***);*

(iv) $\|x + y\| \leq \|x\| + \|y\|$ *for all* $x, y \in X$ *(***triangle inequality***).*

The pair $(X, \|\cdot\|)$ consisting of a vector space X together with a norm $\|\cdot\|$ on it is called a **normed linear space***.*

I stress once again that the normed linear space is the *pair* $(X, \|\cdot\|)$. The same vector space X can be equipped with many different norms, and these give rise to *different* normed linear spaces. However, we shall often refer informally to "the normed linear space X" whenever it is understood from the context what the norm is.

We saw last week that *completeness* is a very nice property for a metric space to have, and that every metric space can be embedded as a dense subset of a complete metric space (called its *completion*); indeed, if one is given an incomplete metric space, then it is advisable to first form its completion and henceforth work there.

Now, the process of forming the completion — by taking equivalence classes of Cauchy sequences — can be applied equally well to a normed linear space (which, after all, is a special case of a metric space), and *this process respects the vector-space structure*. Therefore, the completion of a normed linear space is again a normed linear space (or more precisely, can be given in an obvious way the structure of a normed linear space). And if one is given an incomplete normed linear space, it is advisable to first form its completion and henceforth work there.

A complete normed linear space is called a **Banach space**.[1] Most of the important spaces in functional analysis are Banach spaces.[2] Indeed, much of this course concerns the properties of Banach spaces.

---

[1]Polish mathematician Stefan Banach (1892–1945) was one of the leading contributors to functional analysis in the 1920s and 1930s. His book *Théorie des Opérations Linéaires* (1932) was extremely influential in consolidating the main ideas of functional analysis.

[2]But not all: some important areas of analysis, such as the *theory of distributions* (which plays a central role in the modern theory of partial differential equations), require the study of *non-normable* topological vector spaces.

Before beginning in earnest the general study of Banach spaces, I would like to clean up two issues that we have left pending concerning our favorite example spaces: namely, to prove the completeness of the spaces $\mathcal{C}(X)$ of bounded continuous functions, and to define the sequence spaces ℓ^p for $1 < p < \infty$.

Completeness of spaces $\mathcal{C}(X)$ of bounded continuous functions

In Handout #1 I defined the space $\mathcal{C}(X)$ of bounded continuous real-valued functions for a subset X of the real line. But now that we have defined continuity of maps from one metric space to another, we can see that the same definition holds when X is an arbitrary metric space. That is, we denote by $\mathcal{C}(X)$ the linear space of bounded continuous real-valued functions on X, and we equip it with the sup norm

$$\|f\|_\infty \;=\; \sup_{x \in X} |f(x)| \,. \tag{3.1}$$

Recall, also, that if A is any set, then $\mathcal{B}(A)$ denotes the space of bounded real-valued functions on A, equipped with the sup norm. In Problem 1 of Problem Set #2, you proved that $\mathcal{B}(A)$ is complete, i.e. is a Banach space.

Now, if X is any metric space, then $\mathcal{C}(X)$ is clearly a linear subspace of $\mathcal{B}(X)$, and of course the norm is the same. So to prove that $\mathcal{C}(X)$ is complete, it suffices to prove that $\mathcal{C}(X)$ is closed in $\mathcal{B}(X)$ (why?). Since convergence in $\mathcal{B}(X)$ [i.e. in the sup norm] is another name for *uniform* convergence of functions, what we need to prove is that the limit of a *uniformly convergent* sequence of bounded continuous functions on a metric space X is continuous. In your real analysis course you saw a proof of this fact when X is an interval of the real line (or a subset of $\mathbb{R}^n$); the proof in the general case is identical:

Proposition 3.2 *Let X be any metric space. Then the limit of a uniformly convergent sequence of bounded real-valued continuous functions on X is continuous.*

Corollary 3.3 *Let X be any metric space. Then the space $\mathcal{C}(X)$ is complete, i.e. is a Banach space.*

Proof of Proposition 3.2. Let (f_n) be a uniformly convergent sequence of bounded real-valued continuous functions on X, and let f be the limit function; we want to prove that f is continuous. By hypothesis we have $\lim_{n\to\infty} \|f_n - f\|_\infty = 0$. In other words, for any $\epsilon > 0$ there exists an integer n_0 such that $\|f_n - f\|_\infty < \epsilon/3$ for all $n \geq n_0$. (You will see later why I chose $\epsilon/3$ here.) Now consider any point $x_0 \in X$. By continuity of f_{n_0} at x_0, we can find $\delta > 0$ such that $d(x, x_0) < \delta$ implies $|f_{n_0}(x) - f_{n_0}(x_0)| < \epsilon/3$. But we then have

$$\begin{aligned} |f(x) - f(x_0)| \;&\leq\; |f(x) - f_{n_0}(x)| \,+\, |f_{n_0}(x) - f_{n_0}(x_0)| \,+\, |f_{n_0}(x_0) - f(x_0)| && \text{(3.2a)} \\ &<\; \epsilon/3 \,+\, \epsilon/3 \,+\, \epsilon/3 && \text{(3.2b)} \\ &=\; \epsilon && \text{(3.2c)} \end{aligned}$$

where the first and third summands are bounded by $\|f_{n_0} - f\|_\infty < \epsilon/3$ (why?) and the second summand is bounded by $\epsilon/3$ as we just saw. $\square$

This proof is often called the "$\epsilon/3$ proof" (for obvious reasons) or the "down-over-and-up" proof (you will see why if you draw a picture).

You should give an example to show that a *pointwise-convergent* sequence of bounded real-valued continuous functions need not be continuous, even if the metric space X is compact (e.g. $X = [0,1]$) and even if the convergence is *monotone* (i.e. the sequence increases or decreases to its limit for each $x \in X$).

Definition and completeness of the sequence spaces ℓ^p for $1 < p < \infty$

Let us begin by defining the ℓ^p norm on $\mathbb{R}^n$ (or $\mathbb{C}^n$); we will then define the ℓ^p space of infinite sequences.

Example 1. $\mathbb{R}^n$ with the ℓ^p norm. Let p be any positive real number. Then, for any n-tuple $x = (x_1, x_2, \ldots, x_n)$ of real (or complex) numbers, we define

$$\|x\|_p \;=\; \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} . \tag{3.3}$$

It is easy to see that this is a norm on $\mathbb{R}^n$ (or $\mathbb{C}^n$) when $p = 1$; and in Handout #1 we proved it also for $p = 2$ using the Cauchy–Schwarz inequality. Now we will do it for all p in the interval $(1, \infty)$ by using **Hölder's inequality**, which generalizes the Cauchy–Schwarz inequality (and reduces to it when $p = 2$). The proof will use the following sequence of inequalities:

Lemma 3.4 (Arithmetic-geometric mean inequality) *Let $x, y > 0$ and $0 < \lambda < 1$. Then*

$$x^\lambda y^{1-\lambda} \;\le\; \lambda x \,+\, (1-\lambda) y \, . \tag{3.4}$$

Corollary 3.5 (Young's inequality) *Let $p \in (1, \infty)$, and define $q \in (1, \infty)$ by $\dfrac{1}{p} + \dfrac{1}{q} = 1$. Then, for any real numbers $\alpha, \beta > 0$, we have*

$$\alpha\beta \;\le\; \frac{\alpha^p}{p} \,+\, \frac{\beta^q}{q} \, . \tag{3.5}$$

The relation $\dfrac{1}{p} + \dfrac{1}{q} = 1$ will come up over and over again in discussing the ℓ^p spaces. We refer to q as the *conjugate index* to p.

Proposition 3.6 (Hölder's inequality) *Consider once again $p, q \in (1, \infty)$ satisfying $\dfrac{1}{p} + \dfrac{1}{q} = 1$. Then, for any $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ in $\mathbb{R}^n$ (or $\mathbb{C}^n$), we have*

$$\sum_{i=1}^{n} |x_i y_i| \;\le\; \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \left(\sum_{i=1}^{n} |y_i|^q \right)^{1/q} , \tag{3.6}$$

or in other words

$$\sum_{i=1}^{n} |x_i y_i| \;\le\; \|x\|_p \, \|y\|_q \, . \tag{3.7}$$

Note that when $p = q = 2$ this is the Cauchy–Schwarz inequality. Note also that (3.7) also holds when $p = 1$ and $q = \infty$ (or the reverse) when we interpret $\|\cdot\|_\infty$ as the sup norm.[3]

Proposition 3.7 (Minkowski's inequality) *Let $1 \le p < \infty$. Then, for any $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ in $\mathbb{R}^n$ (or $\mathbb{C}^n$), we have*

$$\left(\sum_{i=1}^{n} |x_i + y_i|^p \right)^{1/p} \;\le\; \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} + \left(\sum_{i=1}^{n} |y_i|^p \right)^{1/p}, \qquad (3.8)$$

or in other words

$$\|x + y\|_p \;\le\; \|x\|_p + \|y\|_p \,. \qquad (3.9)$$

This is the needed triangle inequality for $\mathbb{R}^n$ (or $\mathbb{C}^n$) with the ℓ^p norm.

Proof of Arithmetic-geometric mean inequality. Since the exponential function is convex (why?), we have

$$x^\lambda y^{1-\lambda} \;=\; \exp[\lambda \log x + (1-\lambda) \log y] \;\le\; \lambda \exp[\log x] + (1-\lambda) \exp[\log y] \;=\; \lambda x + (1-\lambda) y \,. \qquad (3.10)$$

□

Proof of Young's inequality. Just put $\lambda = 1/p$, $x = \alpha^p$ and $y = \beta^q$ in the arithmetic-geometric mean inequality. □

Proof of Hölder's inequality. The result is trivial if $x = 0$ or $y = 0$ or both, so let x, y be nonzero vectors in $\mathbb{R}^n$ or $\mathbb{C}^n$. For each index i ($1 \le i \le n$), put

$$\alpha \;=\; \frac{|x_i|}{\|x\|_p} \quad \text{and} \quad \beta \;=\; \frac{|y_i|}{\|y\|_q} \qquad (3.11)$$

into the inequality of Corollary 3.5 and then sum over i. We get

$$\frac{\sum\limits_{i=1}^{n} |x_i y_i|}{\|x\|_p \, \|y\|_q} \;\le\; \frac{1}{p} + \frac{1}{q} \;=\; 1 \,. \qquad (3.12)$$

[3]A quirk of history: Hölder's inequality was first found by the English mathematician Leonard James Rogers (1862–1933) in 1888, and then discovered independently by the German mathematician Otto Hölder (1859–1937) one year later. Alas, the term "Hölder's inequality" is by now so well established that it would be very hard to change the name.

Rogers also discovered, in 1894, the famous Rogers–Ramanujan identities, which were rediscovered independently by the Indian mathematical genius Srinivasa Ramanujan (1887–1920) somewhere around 1913. As Hardy writes in his book on Ramanujan:

> The formulae have a very curious history. They were first found in 1894 by Rogers, a mathematician of great talent but comparatively little reputation, now remembered mainly from Ramanujan's rediscovery of his work. Rogers was a fine analyst, whose gifts were, on a smaller scale, not unlike Ramanujan's; but no one paid much attention to anything he did ...

[G.H. Hardy, *Ramanujan: Twelve Lectures on Subjects Suggested by his Life and Work* (1940), p. 91]

(You should check the algebra here!) $\square$

Proof of Minkowski's inequality. It is trivial when $p = 1$, so consider $1 < p < \infty$. We have

$$\sum_{i=1}^{n} |x_i + y_i|^p \;\le\; \sum_{i=1}^{n} |x_i|\, |x_i + y_i|^{p-1} \;+\; \sum_{i=1}^{n} |y_i|\, |x_i + y_i|^{p-1} \,. \tag{3.13}$$

Now use Hölder's inequality to bound each of the two sums on the right: we have

$$\sum_{i=1}^{n} |x_i|\, |x_i + y_i|^{p-1} \;\le\; \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} \left(\sum_{i=1}^{n} |x_i + y_i|^{(p-1)q}\right)^{1/q} \tag{3.14a}$$

$$= \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} \left(\sum_{i=1}^{n} |x_i + y_i|^{p}\right)^{1/q} \tag{3.14b}$$

since $(p-1)q = p$, and similarly

$$\sum_{i=1}^{n} |y_i|\, |x_i + y_i|^{p-1} \;\le\; \left(\sum_{i=1}^{n} |y_i|^p\right)^{1/p} \left(\sum_{i=1}^{n} |x_i + y_i|^{p}\right)^{1/q} \,. \tag{3.15}$$

We therefore have

$$\sum_{i=1}^{n} |x_i + y_i|^p \;\le\; (\|x\|_p + \|y\|_p) \left(\sum_{i=1}^{n} |x_i + y_i|^{p}\right)^{1/q} \,. \tag{3.16}$$

If $x + y \neq 0$ we can divide both sides by $\left(\sum_{i=1}^{n} |x_i + y_i|^{p}\right)^{1/q}$ to obtain the desired inequality (since $1 - 1/q = 1/p$); and if $x + y = 0$ the desired inequality is trivial. $\square$

Remarks. 1. You should show, by example, that $\|\cdot\|_p$ is *not* a norm when $0 < p < 1$, i.e. the triangle inequality *fails*.

2. Note that

$$\lim_{p \to \infty} \|x\|_p \;=\; \max_{1 \le i \le n} |x_i| \tag{3.17}$$

(why?). This justifies our notation $\|x\|_\infty$ for the max norm. $\square$

Example 2. The sequence space ℓ^p. For any sequence $x = (x_1, x_2, \ldots)$ of real (or complex) numbers and any real number $p \in [1, \infty)$, we define

$$\|x\|_p \;=\; \left(\sum_{i=1}^{\infty} |x_i|^p\right)^{1/p} \,. \tag{3.18}$$

Of course, $\|x\|_p$ defined in this way might be $+\infty$! We define ℓ^p to be the set of sequences for which $\|x\|_p < \infty$. We must now verify that

(a) ℓ^p is a linear space, i.e. $x, y \in \ell^p$ imply $x + y \in \ell^p$; and

(b) $\| \cdot \|_p$ is a norm on ℓ^p.

As always, the only nontrivial thing to be verified in (b) is the triangle inequality $\|x+y\|_p \leq \|x\|_p + \|y\|_p$; and this will also demonstrate (a).

Proposition 3.8 (Minkowski's inequality for infinite sequences) *Let $1 \leq p < \infty$. Then, for any $x = (x_1, x_2, \ldots)$ and $y = (y_1, y_2, \ldots)$ in ℓ^p, we have $x + y \in \ell^p$ and*

$$\left(\sum_{i=1}^{\infty} |x_i + y_i|^p \right)^{1/p} \leq \left(\sum_{i=1}^{\infty} |x_i|^p \right)^{1/p} + \left(\sum_{i=1}^{\infty} |y_i|^p \right)^{1/p} , \tag{3.19}$$

or in other words

$$\|x + y\|_p \leq \|x\|_p + \|y\|_p . \tag{3.20}$$

PROOF. From the finite-dimensional Minkowski's inequality we have for every n

$$\left(\sum_{i=1}^{n} |x_i + y_i|^p \right)^{1/p} \leq \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} + \left(\sum_{i=1}^{n} |y_i|^p \right)^{1/p} \tag{3.21a}$$

$$\leq \left(\sum_{i=1}^{\infty} |x_i|^p \right)^{1/p} + \left(\sum_{i=1}^{\infty} |y_i|^p \right)^{1/p} , \tag{3.21b}$$

which is $< \infty$ since by hypothesis $x, y \in \ell^p$. We can now take $n \to \infty$ to conclude that $\sum_{i=1}^{\infty} |x_i + y_i|^p < \infty$ and that (3.19) holds. $\square$

So we have shown that ℓ^p is a normed linear space for each $p \in [1, \infty)$. The next step is to show that ℓ^p is *complete*, hence is a Banach space. The proof is given in Kreyszig, Section 1.5 (handout) and is essentially the same for general $p \in [1, \infty)$ as it is for $p = 1$.

Elementary properties of normed linear spaces

Let $(X, \| \cdot \|)$ be a normed linear space. The following properties are easy consequences of the triangle inequality, and you should supply the proofs:

Lemma 3.9 *Let $(X, \| \cdot \|)$ be any normed linear space. Then:*

(a) For any $x, y \in X$, we have

$$\Big| \|x\| - \|y\| \Big| \leq \|x - y\| . \tag{3.22}$$

In particular, the norm is a continuous *function on X (in fact, a Lipschitz-continuous function of Lipschitz constant 1).*

(b) *For any* $x, y \in X$ *and* $0 \le \lambda \le 1$*, we have*

$$\|\lambda x + (1-\lambda)y\| \;\le\; \lambda\|x\| \;+\; (1-\lambda)\|y\| \;, \tag{3.23}$$

i.e. the norm is a convex function *on* X*. In particular, open and closed balls are* convex sets *in* X*, i.e.* $x, y \in B(x_0, r)$ *imply* $\lambda x + (1-\lambda)y \in B(x_0, r)$ *for all* $0 \le \lambda \le 1$*, and likewise for* $\overline{B}(x_0, r)$*.*

(c) *The vector-space operations are* jointly continuous*: i.e. if* $x_n \to x$ *and* $y_n \to y$*, then* $x_n + y_n \to x + y$*; and if* $x_n \to x$ *and* $\lambda_n \to \lambda$*, then* $\lambda_n x_n \to \lambda x$*.*

Subspaces and quotient spaces

Whenever one introduces a new mathematical structure, it is natural to seek operations that construct new examples of that structure out of old ones. In linear algebra we are familiar with two elementary ways of getting new vector spaces out of old ones: namely, *linear subspaces* and *quotient spaces*. Here we want to investigate how these constructions work in the context of a normed linear space.

Subspaces are easy: Let $(X, \|\cdot\|)$ be any normed linear space, and let M be any linear subspace of X. Then the restriction of $\|\cdot\|$ to M is obviously a norm on M, let us call it $\|\cdot\|_M$. So $(M, \|\cdot\|_M)$ is a normed linear space; it is called the **normed linear subspace** of X corresponding to the linear subspace M.

Usually we will just refer to "the subspace M of X" rather than the more pedantic "the subspace $(M, \|\cdot\|_M)$ of $(X, \|\cdot\|)$" — the norm being understood from the context.

We then have the following fundamental result:

Proposition 3.10 *Let* X *be a Banach space (i.e. a* complete *normed linear space) and let* M *be a* closed *linear subspace of* X*. Then* M *is a Banach space (i.e. is complete).*

Proof. This is an immediate consequence of the elementary fact that a *closed* subspace of a *complete* metric space is complete. (The fact that M is a *linear* subspace of X is irrelevant to the proof that M is complete.) $\square$

We also have the following easy fact:

Proposition 3.11 *Let* X *be a normed linear space, and let* M *be a linear subspace of* X*. Then the closure* $\overline{M}$ *is also a linear subspace of* X*.*

This is an easy consequence of the joint continuity of algebraic operations [Lemma 3.9(c)], and you should supply the details of the proof yourself.

The second, and slightly more complicated, construction is that of a *quotient space*. Let us first recall how quotient spaces are constructed in linear algebra. Let X be any vector space and let M be any linear subspace of X. Then the **quotient space** X/M is defined to be the linear space of *cosets*

$$[x] \;=\; x + M \;=\; \{x + y \colon\, y \in M\} \tag{3.24}$$

under the algebraic operations

$$[x]+[y] \;\equiv\; (x+M)+(y+M) \;=\; (x+y)+M \;=\; [x+y] \qquad \text{for } [x],[y]\in X/M \tag{3.25}$$

and

$$\lambda[x] \;\equiv\; \lambda(x+M) \;=\; \lambda x+M \;=\; [\lambda x] \qquad \text{for } [x]\in X/M \text{ and } \lambda\in\mathbb{R} \text{ (or } \mathbb{C}) \;. \tag{3.26}$$

The coset $[0]$ is the zero vector of X/M and is simply denoted 0. It is routine to verify that all the axioms of a vector space are satisfied.

Now let us equip the quotient space with a norm. A little thought shows that the right definition is

$$\|[x]\|_{X/M} \;=\; d(0,x+M) \;=\; \inf_{y\in x+M} \|y\| \;=\; \inf_{m\in M} \|x+m\| \;. \tag{3.27}$$

We call $\|\cdot\|_{X/M}$ the **quotient norm** on the space X/M that is induced by the norm $\|\cdot\|$ on X. Of course, we still need to verify that this *is* a norm; and one slight subtlety arises in this verification:

Proposition 3.12 *$\|\cdot\|_{X/M}$ is a norm on X/M if and only if M is a* closed *linear subspace of X.*

Proof. It is routine to verify properties (i), (iii) and (iv) of the norm $\|\cdot\|_{X/M}$, and I leave it as an exercise for you (in particular, you should make sure you know how to prove the triangle inequality). The property that can go wrong is the nondegeneracy property (ii), in case M is not closed. Indeed, if M is not closed, we can take any $x\in\overline{M}\setminus M$, and then $\|[x]\|_{X/M}=0$ (why?) but $[x]\neq[0]$ (why?). On the other hand, if M is closed, then so is $x+M$ (why?), and $d(0,x+M)=0$ if and only if $0\in x+M$ by Proposition 1.14 of Handout #1, i.e. if and only if $[x]=[0]$. $\square$

The next step is to verify completeness:

Proposition 3.13 *Let X be a Banach space (i.e. a* complete *normed linear space) and let M be a* closed *linear subspace of X. Then X/M is a Banach space (i.e. is complete).*

Proof. Consider a Cauchy sequence $\{x_n+M\}$ in $(X/M,\|\cdot\|_{X/M})$. Then for each positive integer k there exists an integer n_k such that

$$\|(x_m+M)-(x_n+M)\|_{X/M} \;<\; 2^{-k} \qquad \text{for all } m,n\ge n_k \tag{3.28}$$

and we can also choose the n_k so that $n_1<n_2<\ldots$. Now consider the subsequence $\{x_{n_k}+M\}$. We have

$$\|(x_{n_k}+M)-(x_{n_{k+1}}+M)\|_{X/M} \;<\; 2^{-k} \;. \tag{3.29}$$

By definition of the quotient norm, it follows that we can successively choose $y_k\in x_{n_k}+M$ such that

$$\|y_k-y_{k+1}\| \;<\; 2^{-k} \;. \tag{3.30}$$

But it follows from this that $\{y_k\}$ is a Cauchy sequence in X (why? do you see why we used 2^{-k} instead of e.g. $1/k$?). Therefore, since X is complete, $\{y_k\}$ converges to some $y \in X$. But then

$$\|(x_{n_k}+M)-(y+M)\|_{X/M} \;=\; \|(y_k+M)-(y+M)\|_{X/M} \;\le\; \|y_k-y\| \;\to\; 0 \text{ as } k \to \infty\ , \tag{3.31}$$

so $\{x_{n_k} + M\}$ is convergent to $y + M$ in $(X/M, \|\cdot\|_{X/M})$. But $\{x_{n_k} + M\}$ is a convergent subsequence of the original *Cauchy* sequence $\{x_n + M\}$, so $\{x_n + M\}$ also converges to $y + M$ by Proposition 1.27 of Handout #1. $\square$

Very soon we will introduce a third operation for constructing new normed linear spaces out of old ones: namely, forming *spaces of linear mappings*.

Continuous linear mappings

In linear algebra you studied **linear mappings** T from one vector space X to another vector space Y. (Also called **linear operators** — the two terms are synonyms.) Here we would like to see how linear mappings fit together with the normed structure.

So let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed linear spaces, and let $T\colon X \to Y$ be a linear mapping. Let us define

$$\|T\|_{X\to Y} \;\equiv\; \sup_{\substack{x \in X \\ \|x\|_X = 1}} \|Tx\|_Y \;=\; \sup_{\substack{x \in X \\ \|x\|_X \le 1}} \|Tx\|_Y \;=\; \sup_{x \in X\setminus\{0\}} \frac{\|Tx\|_Y}{\|x\|_X}\ . \tag{3.32}$$

(You should make sure you understand why all these definitions are equivalent.) Of course, $\|T\|_{X\to Y}$ might be $+\infty$. We say that the map T is **bounded** if $\|T\|_{X\to Y} < \infty$.[4]

Remark. Another equivalent definition is

$$\|T\|_{X\to Y} \;=\; \inf\{M\colon \|Tx\|_Y \le M\|x\|_X \text{ for all } x \in X\} \tag{3.33}$$

provided that the infimum of an empty set is defined to be $+\infty$. You should supply the details of the proof that this is indeed equivalent to the other definitions.

We then have the following fundamental fact:

Proposition 3.14 *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed linear spaces, and let $T\colon X \to Y$ be a linear mapping. Then the following are equivalent:*

(a) T is continuous at 0.

(b) T is continuous.

[4]**Warning:** This terminology slightly conflicts with the usual meaning of "bounded" in discussing functions from one metric space to another, i.e. a function is called bounded (in the usual sense) if its range is a bounded subset of the range space. Now, the only *linear* map that is bounded in the usual sense is the zero map! (Why?) So it is silly, for a *linear* map T, to ask for boundedness in the usual sense; the more interesting question, for a linear map, is whether *the restriction of T to a bounded subset of the domain space* is bounded in the usual sense; and that is exactly what $\|T\|_{X\to Y} < \infty$ says (why?).

(c) T is uniformly continuous.

(d) T is bounded.

PROOF. The equivalence of (a), (b) and (c) is an easy consequence of the properties of the norm and the linearity of T (you should work out the details for yourself). Furthermore, if T is bounded, then it is clearly continuous at 0 (why?). On the other hand, suppose that T is not bounded; then for each positive integer n there exists $x_n \in X$ with $\|x_n\|_X = 1$ and $\|Tx_n\|_Y \geq n$. We then have $x_n/n \to 0$ in X but $T(x_n/n) \not\to 0$ in Y since $\|T(x_n/n)\|_Y \geq 1$. This shows that T is not continuous at 0. □

In the special case where T is a *bijection*, the inverse map $T^{-1}\colon Y \to X$ is well-defined, and we have the following criterion for its continuity:

Proposition 3.15 *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed linear spaces, and let $T\colon X \to Y$ be a bijective linear mapping. Then the following are equivalent:*

(a) T^{-1} is continuous at 0.

(b) T^{-1} is continuous.

(c) T^{-1} is uniformly continuous.

(d) T^{-1} is bounded.

(e) There exists $m > 0$ such that $\|Tx\|_Y \geq m\|x\|_X$ for all $x \in X$.

PROOF. The equivalence of (a)–(d) is just Proposition 3.14 applied to T^{-1}. On the other hand, (e) is equivalent to the statement that $\|T^{-1}y\|_X \leq m^{-1}\|y\|_Y$ for all $y \in Y$, which is precisely the statement that T^{-1} is bounded (by m^{-1}). □

A bijective linear mapping $T\colon X \to Y$ is called a **topological isomorphism** (or **linear homeomorphism**) if both T and T^{-1} are continuous. Putting together Propositions 3.14 and 3.15, we have the following immediate corollary:

Corollary 3.16 *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed linear spaces, and let $T\colon X \to Y$ be a bijective linear mapping. Then T is a topological isomorphism if and only if there exist numbers $m > 0$ and $M < \infty$ such that*

$$m\|x\|_X \;\leq\; \|Tx\|_Y \;\leq\; M\|x\|_X \tag{3.34}$$

for all $x \in X$.

Of course, if $T\colon X \to Y$ is a topological isomorphism, then so is $T^{-1}\colon Y \to X$; and if $S\colon X \to Y$ and $T\colon Y \to Z$ are topological isomorphisms, then so is $T \circ S\colon X \to Z$ (why?).

Two normed linear spaces X and Y are said to be **topologically isomorphic** if there exists a topological isomorphism $T\colon X \to Y$. From what has just been said, topological isomorphism is an equivalence relation among normed linear spaces (i.e. is reflexive, symmetric and transitive).

It is worth observing that completeness of normed linear spaces is invariant under topological isomorphism:

Proposition 3.17 *Suppose that the normed linear spaces X and Y are topologically isomorphic. If X is complete, then so is Y.*

PROOF. Let $T\colon X \to Y$ be a topological isomorphism. If (y_n) is a Cauchy sequence in Y, then $(T^{-1}y_n)$ is a Cauchy sequence in X (why?). Since X is complete, there exists $x \in X$ such that $T^{-1}y_n \to x$ in X. But then $y_n \to Tx$ in Y (why?). $\square$

Do you see how the boundedness of both T and T^{-1} was used in this proof?

Two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on the same space X are said to be **equivalent** if the identity mapping id: $(X, \|\cdot\|_1) \to (X\|\cdot\|_2)$ is a topological isomorphism. By Corollary 3.16 we see that $\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent if and only if there exist numbers $m > 0$ and $M < \infty$ such that

$$m\|x\|_1 \;\le\; \|x\|_2 \;\le\; M\|x\|_1 \,. \tag{3.35}$$

(This property is sometimes taken as the definition of equivalent norms, and then the topological-isomorphism property is proven to be an equivalent characterization.) Of course, if norms $\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent and norms $\|\cdot\|_2$ and $\|\cdot\|_3$ are equivalent, then norms $\|\cdot\|_1$ and $\|\cdot\|_3$ are equivalent; so equivalence of norms is an equivalence relation.

If normed linear spaces X and Y are topologically isomorphic, then they are essentially identical for all *topological* purposes (i.e. things relating to open sets, convergence, continuity, etc.) but not necessarily for *metric* purposes (i.e. things relating to distances, balls, etc.). A stronger notion is **isometric isomorphism**: a bijective linear mapping $T\colon X \to Y$ is called an isometric isomorphism if $\|Tx\|_Y = \|x\|_X$ for all $x \in X$, i.e. if we can take $m = M = 1$ in (3.34). Two normed linear spaces X and Y are said to be **isometrically isomorphic** if there exists an isometric isomorphism $T\colon X \to Y$. If normed linear spaces X and Y are isometrically isomorphic, then they are essentially identical for *all* purposes.

Example: $(\mathbb{R}^n, \|\cdot\|_2)$ and $(\mathbb{R}^n, \|\cdot\|_\infty)$ are topologically isomorphic but not isometrically isomorphic. To see that there cannot exist any isometric isomorphism from $(\mathbb{R}^n, \|\cdot\|_2)$ to $(\mathbb{R}^n, \|\cdot\|_\infty)$, it suffices to note that the closed unit ball of $(\mathbb{R}^n, \|\cdot\|_2)$ is round, while the closed unit ball of $(\mathbb{R}^n, \|\cdot\|_\infty)$ has flat sides (you should draw these two balls for the case $n = 2$). Stated more precisely, every boundary point of the closed unit ball of $(\mathbb{R}^n, \|\cdot\|_2)$ is an extreme point, while this is not so for the closed unit ball of $(\mathbb{R}^n, \|\cdot\|_\infty)$.[5]

Spaces of continuous linear mappings

At the purely algebraic level, we can define **spaces of linear mappings** as follows: If X and Y are vector spaces over the same field (either $\mathbb{R}$ or $\mathbb{C}$), then the space $\mathcal{L}(X, Y)$ of all linear mappings from X to Y is a vector space under the obvious definitions of pointwise addition and multiplication by a scalar.

Now let X and Y be *normed* linear spaces (again over the same field); we define $\mathcal{B}(X, Y)$ to be the space of all *continuous* (or equivalently, bounded, hence the choice of the letter $\mathcal{B}$)

[5]If C is a convex subset of a real vector space V, a point $x \in C$ is called an **extreme point** of C if there do not exist *distinct* points $y, z \in C$ and a number $0 < \lambda < 1$ such that $x = \lambda y + (1 - \lambda)z$.

linear mappings from X to Y.[6] It is obviously a linear subspace of $\mathcal{L}(X,Y)$ (why?). More importantly, it is a *normed* linear space under the **operator norm** $\|\cdot\|_{X\to Y}$:

Proposition 3.18 *$\|\cdot\|_{X\to Y}$ is a norm on the vector space $\mathcal{B}(X,Y)$.*

The easy verification of the properties of a norm is left to you. We call $\mathcal{B}(X,Y)$, equipped with the operator norm, the **space of continuous linear mappings** from X to Y.

The fundamental result is the following:

Proposition 3.19 *Let X and Y be normed linear spaces, with Y complete (i.e. a Banach space). Then $\mathcal{B}(X,Y)$ is complete (i.e. is a Banach space).*

Proof. Let (T_n) be a Cauchy sequence in $\mathcal{B}(X,Y)$. We aim to show that (T_n) converges to some $T\in\mathcal{B}(X,Y)$. The structure of the proof is the same as in our previous completeness proofs: First we identify the putative limit T, by using convergence in a *weaker* sense than convergence in the norm of $\mathcal{B}(X,Y)$; next we prove that T belongs to the space $\mathcal{B}(X,Y)$; and finally, we prove that (T_n) actually converges to T in the norm of $\mathcal{B}(X,Y)$.

Step 1. Since (T_n) is a Cauchy sequence in $\mathcal{B}(X,Y)$, it follows that $(T_n x)$ is a Cauchy sequence in Y for each $x\in X$ (why?). Since Y is assumed complete, this means that $(T_n x)$ converges to some (necessarily unique) element of Y, call it Tx. The mapping $T\colon X\to Y$ defined by $x\mapsto Tx$ is linear (why?).

Step 2. Since (T_n) is a Cauchy sequence in $\mathcal{B}(X,Y)$, it is necessarily bounded, i.e. there exists M such that $\|T_n\|_{X\to Y}\le M$ for all n. It easily follows from this that $\|T\|_{X\to Y}\le M$ (why?). So T is indeed a *bounded* linear operator from X to Y, i.e. it belongs to the space $\mathcal{B}(X,Y)$.

Step 3. Fix $\epsilon>0$. Since (T_n) is a Cauchy sequence in $\mathcal{B}(X,Y)$, there exists n_0 such that

$$\|T_m - T_n\|_{X\to Y} \;\le\; \epsilon \text{ whenever } m,n\ge n_0\ , \tag{3.36}$$

and hence that

$$\|T_m x - T_n x\|_Y \;\le\; \epsilon\|x\|_X \text{ for all } x\in X \text{ whenever } m,n\ge n_0\ . \tag{3.37}$$

Now let $n\to\infty$ with m fixed; we have $T_n x\to Tx$ by construction, and hence

$$\|T_m x - Tx\|_Y \;\le\; \epsilon\|x\|_X \text{ for all } x\in X \text{ whenever } m\ge n_0 \tag{3.38}$$

(why? do you see how Lemma 3.9(a) is being used here?). But this means that

$$\|T_m - T\|_{X\to Y} \;\le\; \epsilon \text{ whenever } m\ge n_0\ , \tag{3.39}$$

which is exactly what is needed to prove that $T_m\to T$ in the space $\mathcal{B}(X,Y)$. $\square$

Note that it makes no difference in the foregoing proposition whether X is complete or not. This is related to the following fact:

[6] **Warning 1:** Some authors (e.g. Dieudonné) write $\mathcal{L}(X,Y)$ to denote the space of all *continuous* linear mappings from X to Y.

Warning 2: I trust that there will be no confusion between the space $\mathcal{B}(X,Y)$ of bounded linear mappings from a normed linear space X to another normed linear space Y, and the space $\mathcal{B}(A)$ of all bounded real-valued functions (in the usual sense of "bounded") on a set A. I apologize for using the same letter $\mathcal{B}$ in both cases; but there are, alas, only 26 letters in the alphabet, and only one that starts the word "bounded".

Proposition 3.20 *Let X and Y be normed linear spaces, with Y complete (i.e. a Banach space), and let A be a dense linear subspace of X. Then every continuous linear mapping $T\colon A \to Y$ has a unique extension to a continuous linear mapping $\widetilde{T}\colon X \to Y$, and $\|\widetilde{T}\|_{X\to Y} = \|T\|_{A\to Y}$. Indeed, the mapping $\iota\colon T \mapsto \widetilde{T}$ is an isometric isomorphism of $\mathcal{B}(A,Y)$ onto $\mathcal{B}(X,Y)$.*

Proof. If a continuous extension $\widetilde{T}$ exists, it is certainly unique, since a continuous function is determined by its values on a dense subset of the domain (why?), and indeed $\widetilde{T}$ must satisfy $\widetilde{T}x = \lim\limits_{n\to\infty} Ta_n$ for any sequence (a_n) in A converging to x (why?).

To see that a continuous extension $\widetilde{T}$ exists, let us try to define it by this formula. That is, consider any $x \in X$, and let (a_n) be any sequence in A converging to x (why must such a sequence exist?). Then (a_n) is a Cauchy sequence in A, so that (Ta_n) is a Cauchy sequence in Y (why?). Since Y is complete, (Ta_n) converges to some $y \in Y$; we define $\widetilde{T}x = y$. To see that the definition makes sense, we must verify that for any other sequence (a'_n) in A converging to x, we also have $Ta'_n \to y$; but this follows from $\|Ta_n - Ta'_n\|_Y \le \|T\|_{A\to Y}\|a_n - a'_n\|_X$ (why?). In particular, this means that $\widetilde{T} \restriction A = T$ (why?). I leave it to you to verify that $\widetilde{T}$ is linear.

Let us now prove that $\|\widetilde{T}\|_{X\to Y} = \|T\|_{A\to Y}$. The $\ge$ is trivial, because $\widetilde{T}$ is an extension of T. On the other hand, if (a_n) is a sequence in A converging to x, we have

$$\|\widetilde{T}x\|_Y \;=\; \lim_{n\to\infty} \|Ta_n\|_Y \;\le\; \|T\|_{A\to Y} \lim_{n\to\infty} \|a_n\|_X \;=\; \|T\|_{A\to Y}\|x\|_X \,. \tag{3.40}$$

It is easy to easy to see that the map $T \mapsto \widetilde{T}$ is linear, and that it is a bijection of $\mathcal{B}(A,Y)$ onto $\mathcal{B}(X,Y)$ (why?). Since $\|\widetilde{T}\|_{X\to Y} = \|T\|_{A\to Y}$, it is an isometric isomorphism of $\mathcal{B}(A,Y)$ onto $\mathcal{B}(X,Y)$. $\square$

Remark. Proposition 3.20 is actually a special case of more general results concerning extension of continuous mappings from one *metric space* to another: see e.g. Dieudonné, (3.15.5) and (3.15.6).

The relation between the preceding two propositions arises from the fact that every incomplete normed linear space X can be isometrically embedded as a dense linear subspace of its completion $\widehat{X}$. Proposition 3.20 then tells us that $\mathcal{B}(X,Y)$ is isometrically isomorphic to $\mathcal{B}(\widehat{X},Y)$, so it makes no difference which of these two spaces we consider: they are essentially the *same* space.

Some examples and counterexamples

It is instructive to look at some examples of bounded and unbounded linear operators:

Example 1: Linear mappings from $\mathbb{R}^n$ to $\mathbb{R}^m$. As you undoubtedly recall from linear algebra, any $m \times n$ matrix $A = (a_{ij})$ of real numbers defines a linear mapping (by abuse of language let us also call it A) from $\mathbb{R}^n$ to $\mathbb{R}^m$ by

$$(Ax)_i \;=\; \sum_{j=1}^{n} a_{ij}x_j \qquad \text{for } i = 1,2,\ldots,m \,. \tag{3.41}$$

And conversely, every linear mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$ is induced by such a matrix (why?).

Let us now show that every such mapping is bounded. To give meaning to this statement, we first have to equip $\mathbb{R}^n$ and $\mathbb{R}^m$ with norms. We have already studied three such norms — the ℓ^1, ℓ^2 and ℓ^∞ norms — and it is easy to see that these three norms are all equivalent on any space $\mathbb{R}^n$ (why? note that this statement does *not* survive if we try to take $n \to \infty$: why?). So it doesn't matter which one we choose; let's choose for simplicity the ℓ^∞ norm for both the domain space $\mathbb{R}^n$ and the range space $\mathbb{R}^m$. I then claim that the operator norm $\|A\|_{(\mathbb{R}^n, \|\cdot\|_\infty) \to (\mathbb{R}^m, \|\cdot\|_\infty)}$ for such a mapping is given in terms of the matrix $A = (a_{ij})$ by

$$\|A\|_{(\mathbb{R}^n, \|\cdot\|_\infty) \to (\mathbb{R}^m, \|\cdot\|_\infty)} \;=\; \max_{1 \le i \le m} \sum_{j=1}^{n} |a_{ij}| \,. \tag{3.42}$$

I leave it to you to prove this (Problem 4(a) on Problem Set #3). It follows, in particular, that every linear mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$ is bounded.

The same considerations apply if A is an $m \times n$ matrix of *complex* numbers, and lead to a bounded linear mapping from $\mathbb{C}^n$ to $\mathbb{C}^m$. The formula for the norm is the same.

We will see later, in fact, that every linear mapping from a finite-dimensional normed linear space to *any* normed linear space (finite-dimensional or not) is bounded (Proposition 3.25).

Example 2: Linear mappings on sequence spaces induced by infinite matrices. In a similar way, we can start from an "infinite matrix" $A = (a_{ij})_{i,j=1}^\infty$ of real numbers, and try to define a linear operator on a sequence space (for example, on one of our spaces ℓ^p) by

$$(Ax)_i \;=\; \sum_{j=1}^{\infty} a_{ij} x_j \qquad \text{for } i = 1, 2, \ldots \,. \tag{3.43}$$

But now things become nontrivial: we have to verify, first of all, that the sum is indeed convergent for all x in the domain space, and secondly, that the resulting sequence Ax indeed belongs to the desired range space. Only if we can do this does it then make sense to inquire whether the resulting linear map is bounded or not.

I won't try to do this in general, but will simply give a few important examples:

Consider the infinite matrix

$$A \;=\; \begin{pmatrix} 0 & 1 & 0 & 0 \cdots \\ 0 & 0 & 1 & 0 \cdots \\ 0 & 0 & 0 & 1 \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} , \tag{3.44}$$

whose matrix elements are

$$a_{ij} \;=\; \begin{cases} 1 & \text{if } j - i = 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.45}$$

This matrix can be interpreted as acting on any of the ℓ^p spaces ($1 \le p \le \infty$) and gives

$$A\,(x_1, x_2, x_3, \ldots) \;=\; (x_2, x_3, x_4, \ldots) \tag{3.46}$$

(why?). You should convince yourself that $x \in \ell^p$ implies $Ax \in \ell^p$ and that $\|A\|_{\ell^p \to \ell^p} = 1$. This operator is called **left shift**.

Similarly, one can define the **right shift** on ℓ^p: you should figure out what is the corresponding infinite matrix B and show that again $\|B\|_{\ell^p \to \ell^p} = 1$.

There are also **weighted shifts**, in which the infinite sequence of 1's in the matrix of A (or B) is replaced by a suitable sequence $(a_i)_{i=1}^\infty$ of scalars, so that, for instance, the weighted left shift sends $(x_1, x_2, x_3, \ldots)$ to $(a_1 x_2, a_2 x_3, a_3 x_4, \ldots)$. Of course, we have to be careful in our choice of weights $(a_i)_{i=1}^\infty$ if we want to make sure that the domain space is mapped into the desired range space and that the mapping is bounded. Can you see that if the sequence $(a_i)_{i=1}^\infty$ is *bounded*, then the weighted left (or right) shift defines a bounded linear map from any ℓ^p space to itself? The class of weighted shift operators on ℓ^2 plays an important role in the theory of linear operators on Hilbert spaces.

Some more examples of linear maps defined by infinite matrices are given in Problem 4(c,d) of Problem Set #3.

Example 3: Integral operators. We consider the space $\mathcal{C}[a,b]$ of (automatically bounded) continuous functions on a bounded closed interval $[a,b] \subset \mathbb{R}$, equipped with the sup norm; we will define some linear mappings from $\mathcal{C}[a,b]$ to itself via integrals. To do this, let $K\colon [a,b] \times [a,b] \to \mathbb{R}$ be a continuous function (called the **kernel**), and define a mapping $T\colon \mathcal{C}[a,b] \to \mathcal{C}[a,b]$ by

$$(Tf)(s) \;=\; \int\limits_a^b K(s,t)\, f(t)\, dt\,. \tag{3.47}$$

It is not hard to show that Tf is continuous whenever f is, and that T defines a bounded linear map from $\mathcal{C}[a,b]$ to itself, of norm

$$\|T\|_{\mathcal{C}[a,b] \to \mathcal{C}[a,b]} \;=\; \sup_{s \in [a,b]} \int_a^b |K(s,t)|\, dt \tag{3.48}$$

(Problem 6 on Problem Set #3). Do you see that this is an infinite-dimensional version of Example 1, in which the matrix (a_{ij}) is replaced by the kernel $K(s,t)$ and the sum over j is replaced by integration over t? And do you see that (3.48) is the analogue of (3.42)? The operator T defined by (3.47) is called a *Fredholm integral operator of the first kind*.[7]

If T is a Fredholm integral operator of the first kind, then $U = I - T$ (where I denotes the identity operator) is called a *Fredholm integral operator of the second kind*: explicitly it is given by

$$(Uf)(s) \;=\; f(s) \;-\; \int\limits_a^b K(s,t)\, f(t)\, dt\,. \tag{3.49}$$

If in (3.47) or (3.49), we change the upper limit of integration from b (a fixed limit) to s (a variable limit), then T or U is called a *Volterra integral operator* (of the first or second

[7] Named after the Swedish mathematician Erik Ivar Fredholm (1866–1927), who founded the modern theory of integral equations.

kind, respectively).[8]

Now let us look at some examples of what can go wrong in infinite dimensions:

Example 4: An unbounded linear map. Let A be the infinite diagonal matrix $\mathrm{diag}(1, 2, 3, \ldots)$, and let us try to define the corresponding linear operator A on (for example) the sequence space ℓ^2, i.e.

$$(Ax)_k \;=\; k\, x_k \;. \tag{3.50}$$

The trouble is that this formula does not map ℓ^2 into itself — can you see why? (You should give an example of an $x \in \ell^2$ with the property that $Ax \notin \ell^2$.) Indeed, the largest space on which the operator A can be sensibly defined (as a mapping into ℓ^2) is the *subspace* $M \subsetneq \ell^2$ given by

$$M \;=\; \{x \in \ell^2\colon \sum_{k=1}^{\infty} k^2\, |x_k|^2 < \infty\} \;. \tag{3.51}$$

It is not hard to see that M is indeed a linear subspace of ℓ^2, but it is not *closed*; indeed, it is a *proper dense subspace* of ℓ^2. (Can you prove this? Indeed, you might as well go farther and show that the subspace $c_{00} \subsetneq M$ consisting of sequences with at most finitely many nonzero entries is already dense in ℓ^2.)

So now consider A as a linear map from M (equipped with the ℓ^2 norm that it inherits as a subspace of ℓ^2) into ℓ^2. Is it bounded? The answer is no: consider the vector e_n having a 1 in the nth coordinate and 0 elsewhere; it belongs to M (and indeed to c_{00}) and we have $\|e_n\|_{\ell^2} = 1$ but $\|Ae_n\|_{\ell^2} = n$ (why?); so the operator norm of A is infinite.

This is a typical situation: Unbounded operators cannot be defined in any sensible way on the whole Banach space, but only on a proper dense subspace.

Here is another example of the same phenomenon:

Example 5: A differential operator. In the space $\mathcal{C}[0, 1]$ of (bounded) continuous functions on the interval $[0, 1]$, equipped with the sup norm, let us try to define the operator D of differentiation. The trouble, once again, is that differentiation makes no sense for arbitrary functions $f \in \mathcal{C}[0, 1]$ (at least, not if we want the derivative f' to also belong to $\mathcal{C}[0, 1]$). Rather, we need to restrict ourselves to (for example) the space $\mathcal{C}^1[0, 1]$ consisting of functions that are once continuously differentiable on $[0, 1]$ (where we insist on having also one-sided derivatives at the two endpoints, and the derivative f' is supposed to be continuous on all of $[0, 1]$, including at the endpoints). This is a linear subspace of $\mathcal{C}[0, 1]$, but it is not closed; indeed, with some work it can be shown that it is a proper *dense* subspace of $\mathcal{C}[0, 1]$. (There is an easy proof using the Weierstrass approximation theorem, if you know this theorem.) Then the map $D\colon f \mapsto f'$ is a well-defined linear map from $\mathcal{C}^1[0, 1]$ into $\mathcal{C}[0, 1]$. But it is *unbounded*: to see this, it suffices to look at the functions $f_n \in \mathcal{C}^1[0, 1]$ defined by $f_n(x) = x^n$. We have $\|f_n\|_\infty = 1$ but $\|Df_n\|_\infty = n$ (why?), so the operator norm of D (as an operator from $\mathcal{C}^1[0, 1]$ into $\mathcal{C}[0, 1]$, where both spaces are equipped with the sup norm) is infinite.

[8]Named after the Italian mathematician Vito Volterra (1860–1940), who made important contributions to the theory of integral and integrodifferential equations as well as to mathematical biology.

Example 6: A bounded linear bijection with unbounded inverse. Consider the linear map $T\colon \ell^2 \to \ell^2$ defined by

$$(Tx)_j \;=\; \frac{1}{j}\, x_j \;. \tag{3.52}$$

Can you see that T is *bounded*? What is its operator norm?

Can you also see that T is an *injection*? It is not a surjection, i.e. its image is not all of ℓ^2; rather, its image is precisely the dense linear subspace $M \subsetneq \ell^2$ defined in Example 4 above — can you see why?

Since T is a linear bijection from ℓ^2 to M, the inverse map T^{-1} is well-defined as a map from M to ℓ^2; it is obviously given by

$$(T^{-1}x)_j \;=\; j\, x_j \;, \tag{3.53}$$

which is nothing other than the unbounded linear map $A\colon M \to \ell^2$ discussed in Example 4 above. So a bounded linear bijection $T\colon \ell^2 \to M$ can have an unbounded inverse.

This pathology arises from the fact that the space M is not complete. We will see later in this course that a bounded linear bijection from a *Banach* space onto another *Banach* space always has a bounded inverse: this is a corollary of one of the fundamental theorems of functional analysis, the Open Mapping Theorem.

This same map T could have been considered in any of the spaces ℓ^p $(1 \le p \le \infty)$, not just $p = 2$, with similar results; I leave it to you to work out the details.

Special properties of finite-dimensional spaces

In this section we would like to study some special properties possessed by finite-dimensional normed linear spaces — properties that do *not* hold in general for infinite-dimensional normed linear spaces. The culmination of this section is F. Riesz's theorem stating that the closed unit ball of a normed linear space is compact if and *only if* the space is finite-dimensional.

Proposition 3.21 *Let $(X, \|\cdot\|)$ be a finite-dimensional normed linear space over the real (resp. complex) field, and let $\{a_1, \ldots, a_n\}$ be any basis for X. Then the mapping*

$$T\colon\; (\xi_1, \ldots, \xi_n) \;\mapsto\; \sum_{i=1}^{n} \xi_i a_i \tag{3.54}$$

of $(\mathbb{R}^n, \|\cdot\|_\infty)$ [resp. $(\mathbb{C}^n, \|\cdot\|_\infty)$] onto $(X, \|\cdot\|)$ is a topological isomorphism.

Proof. I shall give the proof for $\mathbb{R}^n$; it is identical for $\mathbb{C}^n$.

Clearly T is a linear bijection from $\mathbb{R}^n$ to X (why?). Moreover, for $\xi = (\xi_1, \ldots, \xi_n) \in \mathbb{R}^n$ we have

$$\|T\xi\| \;=\; \left\| \sum_{i=1}^{n} \xi_i a_i \right\| \tag{3.55a}$$

$$\leq \sum_{i=1}^{n} |\xi_i| \, \|a_i\| \tag{3.55b}$$

$$\leq \left(\sum_{i=1}^{n} \|a_i\| \right) \max_{1 \leq i \leq n} |\xi_i| \tag{3.55c}$$

$$= \left(\sum_{i=1}^{n} \|a_i\| \right) \|\xi\|_\infty \,, \tag{3.55d}$$

so T is bounded, hence continuous.

On the other hand, the unit sphere in $(\mathbb{R}^n, \| \cdot \|_\infty)$, namely

$$S = \{x \in \mathbb{R}^n \colon \|x\|_\infty = 1\} \,, \tag{3.56}$$

is clearly a closed and bounded subset of $(\mathbb{R}^n, \| \cdot \|_\infty)$, hence compact. Moreover, the function $\xi \mapsto \|T\xi\|$ is continuous on $\mathbb{R}^n$ (why?) and hence on S; so it attains its minimum on S (why?) and this minimum value m cannot be zero (why?). So we have $\|T\xi\| \geq m > 0$ for all $\xi \in S$, hence $\|T\xi\| \geq m\|\xi\|_\infty$ for all $\xi \in \mathbb{R}^n$ (why?).

It then follows from Corollary 3.16 that T is a topological isomorphism. □

Corollary 3.22 *On a finite-dimensional vector space, all norms are equivalent.*

Proof. Let X be a vector space of finite dimension n, and let $\| \cdot \|_1$ and $\| \cdot \|_2$ be norms on X. Fix a basis $\{a_1, \ldots, a_n\}$ of X, and define the linear bijection $T \colon \mathbb{R}^n \to X$ by (3.54). By Proposition 3.21, T is a topological isomorphism $(\mathbb{R}^n, \| \cdot \|_\infty) \to (X, \| \cdot \|_1)$ and also $(\mathbb{R}^n, \| \cdot \|_\infty) \to (X, \| \cdot \|_2)$. But then $\mathrm{id} = T \circ T^{-1}$ is a topological isomorphism from $(X, \| \cdot \|_1)$ to $(X, \| \cdot \|_2)$. □

Corollary 3.23 *Every finite-dimensional normed linear space is complete.*

Proof. This follows from Propositions 3.17 and 3.21 and the completeness of $(\mathbb{R}^n, \| \cdot \|_\infty)$. □

Corollary 3.24 *Every finite-dimensional subspace of a normed linear space is closed.*

Proof. This follows from the preceding corollary and the fact that a complete set in any metric space is necessarily closed. □

In Problem 7 of Problem Set #3 you will prove a generalization of Corollary 3.24.

Proposition 3.25 *Let X and Y be normed linear spaces, with X finite-dimensional. (Here the space Y is arbitrary.) Then every linear mapping of X into Y is continuous.*

PROOF. If X is an n-dimensional real normed linear space, then X is topologically isomorphic to $(\mathbb{R}^n, \|\cdot\|_\infty)$ by Proposition 3.21. On the other hand, any linear map from $\mathbb{R}^n$ to Y is of the form $T(\xi_1, \ldots, \xi_n) = \sum_{i=1}^n \xi_i y_i$ for suitable vectors $y_1, \ldots, y_n$ in Y (why?); and such a map is necessarily bounded [from $(\mathbb{R}^n, \|\cdot\|_\infty)$ to $(Y, \|\cdot|_Y)$] by the same calculation that was done during the proof of Proposition 3.21. $\square$

We now turn to the deepest result characterizing finite-dimensional spaces among normed linear spaces, due to F. Riesz[9]:

Theorem 3.26 (F. Riesz's theorem) *For a normed linear space X, the following are equivalent:*

(a) The closed unit ball in X is compact.

(b) Every closed bounded set in X is compact.

(c) X is locally compact.

(d) X is finite-dimensional.

The proof of this theorem will rely on a simple but important lemma, also due to F. Riesz:

Lemma 3.27 (F. Riesz's lemma) *Let X be a normed linear space, and let M be a proper closed linear subspace of X. Then for each $\epsilon > 0$ there exists a point $x \in X$ such that $\|x\| = 1$ and $d(x, M) \geq 1 - \epsilon$.*

PROOF OF F. RIESZ'S LEMMA. Since M is proper, there exists a point $x_1 \in X \setminus M$; and since M is closed, we have $d(x_1, M) = d > 0$. But this means that there exists $x_0 \in M$ such that $\|x_1 - x_0\|$ is arbitrarily close to d, say $\leq d/(1-\epsilon)$. Now set

$$x \;=\; \frac{x_1 - x_0}{\|x_1 - x_0\|} \;. \tag{3.57}$$

By construction $\|x\| = 1$. Moreover, for any $y \in M$ we have

$$\|x - y\| \;=\; \left\| \frac{x_1 - x_0}{\|x_1 - x_0\|} - y \right\| \tag{3.58a}$$

$$=\; \frac{\|x_1 - (x_0 + \|x_1 - x_0\| y)\|}{\|x_1 - x_0\|} \tag{3.58b}$$

(why?). Since $x_0 + \|x_1 - x_0\| y \in M$, the numerator is $\geq d$ (why?); and the denominator is $\leq d/(1-\epsilon)$; so the ratio is $\geq 1 - \epsilon$. $\square$

[9]Hungarian mathematician Frigyes Riesz (1880–1956) was another leading contributor to functional analysis. We use his initial "F" to distinguish him from his younger brother Marcel Riesz (1886–1969), who also made important contributions to analysis, particularly partial differential equations.

Remark: It is natural to ask whether we can actually achieve $d(x, M) = 1$ in F. Riesz's lemma. This turns out to be delicate: the answer is yes in some normed linear spaces and no in others. For instance, in Euclidean spaces (which we will discuss later), we can take x to belong to the orthogonal complement $M^\perp$ of M, and thereby achieve $d(x, M) = 1$. For the ℓ^p spaces with $1 < p < \infty$, the answer also turns out to be yes. On the other hand, you will see in Problem 9 of Problem Set #3 a case in which the answer turns out to be no.

PROOF OF F. RIESZ'S THEOREM. It is easy to see that (a), (b) and (c) are all equivalent; you should supply the proof.

If X is finite-dimensional, then it is topologically isomorphic to $(\mathbb{R}^n, \|\cdot\|_\infty)$, where we know that every closed bounded set is compact; so the same is true for X.

Conversely, suppose that the closed unit ball $\overline{B}(0, 1) \subset X$ is compact. Then there exists a finite sequence of points $a_1, \ldots, a_n \in \overline{B}(0, 1)$ such that $\overline{B}(0, 1)$ is contained in the union of the closed balls of center a_i and radius $1/2$. Let M be the finite-dimensional subspace of X generated by $a_1, \ldots, a_n$; by Corollary 3.24 it is closed. Suppose now that M is proper (i.e. that $M \neq X$); then by F. Riesz's lemma there exists $x \in X$ such that $\|x\| = 1$ and $d(x, M) \geq 3/4$. But this contradicts the hypothesis that $\overline{B}(0, 1) \subseteq \bigcup_{i=1}^{n} \overline{B}(a_i, 1/2)$. We conclude that $X = M$, hence that X is finite-dimensional. □

Warning: For topological vector spaces, it is still true that a space is locally compact if and only if it is finite-dimensional. But there do exist infinite-dimensional non-normable topological vector spaces in which every closed bounded set is compact — the most important of these are called *Montel spaces*, and they arise in many applications in both real and complex analysis.[10]

Introduction to duality for normed linear spaces

Among the spaces $\mathcal{B}(X, Y)$ of continuous linear mappings, there are two choices of the target space Y that are of particular importance:

- $Y = \mathbb{R}$ (or $\mathbb{C}$ in the case of a complex normed linear space), considered as a one-dimensional vector space. This is obviously the simplest case.
- $Y = X$. In this case maps can be freely composed: therefore $\mathcal{B}(X, X)$, in addition to being a vector space, also has the structure of an **operator algebra**.

The case $Y = X$ leads to the subject of **spectral theory**, which is of especial importance in quantum mechanics (among other applications) and which you are invited to study in a fourth-year course. Here I would like to begin the analysis of the case $Y = \mathbb{R}$ (or $\mathbb{C}$), which plays a central role in the theory of Banach spaces; we will continue this study a few weeks from now, after proving the Hahn–Banach theorem. In what follows I will assume (simply to fix the notation) that the field of scalars is $\mathbb{R}$, but everything here works with minor

[10]In honor of the French mathematician Paul Montel (1876–1975). Montel's theorem in complex analysis can be interpreted as saying that certain spaces of analytic functions are Montel spaces.

modifications when the field of scalars is $\mathbb{C}$. (When we discuss the Hahn–Banach theorem we will have to treat the complex case separately.)

A linear mapping from a vector space X to the field of scalars (here $\mathbb{R}$) is called a **linear functional** on X. The space $\mathcal{L}(X, \mathbb{R})$ of all linear functionals (continuous or not) on X is called the **algebraic dual space** of X, and is denoted $X^\sharp$. I assume that in your course in linear algebra you studied (even if only briefly) the theory of algebraic duals of finite-dimensional vector spaces, and proved in particular that the algebraic dual of an n-dimensional vector space is n-dimensional.[11] The theory of algebraic duals of infinite-dimensional vector spaces belongs to a more advanced course in linear algebra.

In analysis, however, we are principally interested in *continuous* linear functionals. The space $\mathcal{B}(X, \mathbb{R})$ of all continuous linear functionals on a normed linear space X is called the **(topological) dual space** of X, and is denoted X^*. It is itself a normed linear space under the norm

$$\|\ell\|_{X^*} \;=\; \|\ell\|_{X \to \mathbb{R}} \;=\; \sup_{\|x\|_X \le 1} |\ell(x)| \;. \tag{3.59}$$

By Proposition 3.19, the normed linear space X^* is *complete*, i.e. is a Banach space; this holds whether or not X is complete. Moreover, Proposition 3.20 tells us that if A is a dense linear subspace of X, then A^* is isometrically isomorphic to X^* under the obvious mapping.

Let us begin with some purely algebraic considerations. If X is a vector space (not necessarily normed) and M is a linear subspace of X (not necessarily closed), we say that M has **codimension $\boldsymbol{n}$** in X (where n is a positive integer) if the quotient space X/M has dimension n, or in other words if there exist $e_1, \ldots, e_n \in X \setminus M$ such that every vector $x \in X$ can be *uniquely* represented in the form $x = \lambda_1 e_1 + \ldots + \lambda_n e_n + z$ where $\lambda_1, \ldots, \lambda_n$ are scalars and $z \in M$. A linear subspace of codimension 1 is called a **(linear) hyperplane**. Note that in this case one can choose *any* vector in $X \setminus M$ to play the role of e_1 (why?).

If M is a (linear) hyperplane in X, then each coset $x_0 + M$ (where $x_0 \in X$) is called an **affine hyperplane** in X.[12]

The relevance of hyperplanes to linear functionals is explained by the following proposition, which is purely algebraic:

Proposition 3.28 *Let X be a vector space.*

(a) The kernel of a nonzero linear functional $\ell \in X^\sharp$,

$$\ker \ell \;=\; \{x \in X \colon \, \ell(x) = 0\} \;, \tag{3.60}$$

[11]**Warning:** Though this implies that the vector spaces X and $X^\sharp$ are isomorphic, they should nevertheless be kept conceptually distinct, because they are not *naturally* isomorphic. That is, though there *exists* a linear bijection from X to $X^\sharp$, there in fact exist *infinitely many* such bijections, and no one of them is any more "natural" than any other. Rather, to pick out one of those isomorphisms as "canonical", one has to introduce *additional* structure beyond the vector-space structure of X: for example, by choosing a basis $e_1, \ldots, e_n$ of X. *Then* there is determined in a natural way a "dual basis" $f_1, \ldots, f_n$ of $X^\sharp$ (satisfying $f_i(e_j) = \delta_{ij}$ for all i, j) and an isomorphism $T \colon X \to X^\sharp$ defined by $T(e_i) = f_i$ for all i. For a rant in which I attempted (probably in vain) to explain this principle to my fellow physicists, see A.D. Sokal, Some comments on multigrid methods for computing propagators, *Phys. Lett. B* **317**, 399–408 (1993), `http://arxiv.org/abs/hep-lat/9307020`

[12]**Warning:** Some authors (e.g. Dieudonné) use the term "hyperplane" without adjectives to mean a *linear* hyperplane; others (e.g. Kolmogorov–Fomin) use it to mean an *affine* hyperplane; and still others (e.g. Giles) oscillate between the two meanings! I shall follow Dieudonné in adhering to the first meaning.

is a hyperplane in X. Conversely, every hyperplane $H \subset X$ is the kernel of some nonzero linear functional $\ell \in X^\sharp$.

(b) *The set where a nonzero linear functional $\ell \in X^\sharp$ takes the value 1,*

$$A_\ell \;=\; \ell^{-1}[1] \;=\; \{x \in X \colon \ell(x) = 1\}\;, \tag{3.61}$$

is an affine hyperplane in X, not passing through the origin. Conversely, every affine hyperplane $A \subset X$ not passing through the origin is of the form A_ℓ for some nonzero linear functional $\ell \in X^\sharp$.

Proof. (a) Let ℓ be a nonzero linear functional on X. Then $\ker \ell$ is a proper linear subspace of X (why?). Now fix any $x_0 \in X \setminus \ker \ell$, and note that $\ell(x_0) \neq 0$. We can then represent each $x \in X$ uniquely in the form $x = \lambda x_0 + y$ with $\lambda \in \mathbb{R}$ and $y \in \ker \ell$; indeed, such a representation can hold only if we take $\lambda = \ell(x)/\ell(x_0)$ [just apply ℓ to both sides to see this], and if we do choose this value of λ, then $x - \lambda x_0 \in \ker \ell$ (why?). This shows that $\ker \ell$ is a hyperplane in X.

Conversely, if H is a hyperplane in X, choose any $x_0 \in X \setminus H$; then by hypothesis each $x \in X$ can be represented uniquely in the form $x = \lambda x_0 + y$ with $\lambda \in \mathbb{R}$ and $y \in H$. Defining $\ell(x) = \lambda$ then does the trick.

(b) If ℓ is a nonzero linear functional on X, there clearly exists $x_0 \in X$ such that $\ell(x_0) = 1$ (why?). Then we have $A_\ell = x_0 + \ker \ell$ (why?), so using the result of part (a) we conclude that A_ℓ is an affine hyperplane in X; and clearly $0 \notin A_\ell$.

Conversely, if $A \subset X$ is an affine hyperplane not passing through the origin, then we can write $A = x_0 + H$ for some hyperplane H and some $x_0 \in X \setminus H$. By the result of part (a), there exists a nonzero linear functional $\ell_0 \in X^\sharp$ such that $H = \ker \ell_0$; in particular, $\ell_0(x_0) \neq 0$. Defining $\ell(x) = \ell_0(x)/\ell_0(x_0)$ then does the trick. □

Let us now pass from algebra to analysis, and assume that X is a normed linear space.

Proposition 3.29 *In a normed linear space X, a hyperplane $M \subset X$ is either closed or dense.*

Proof. By Proposition 3.11, $\overline{M}$ is a linear subspace of X containing M. Since M has codimension 1, the only possibilities are $\overline{M} = M$ and $\overline{M} = X$. □

Note that the affine hyperplanes $x_0 + M$ are also closed or dense according as M is.

Proposition 3.30 *Let X be a normed linear space, and let $\ell \in X^\sharp$ be a linear functional. Then ℓ is continuous if and only if $\ker \ell$ is closed. Moreover, we have the quantitative relation*

$$\|\ell\|_{X^*} \;=\; \frac{1}{d(0, A_\ell)} \tag{3.62}$$

where

$$A_\ell \;=\; \ell^{-1}[1] \;=\; \{x \in X \colon \ell(x) = 1\}\;. \tag{3.63}$$

[If $\ell = 0$ we have $A_\ell = \varnothing$ and we define $d(0, \varnothing) = +\infty$.]

PROOF. Everything is trivial (including the quantitative relation) if $\ell = 0$, so we henceforth assume that $\ell \neq 0$.

If ℓ is continuous, then $\ker \ell$ is obviously closed. To prove the converse, suppose that $H = \ker \ell$ is a closed, and choose any $x_0 \in X \setminus H$ such that $\ell(x_0) = 1$; we then have $d(x_0, H) > 0$ (why?). Then every $x \in X$ can be uniquely written in the form $x = \lambda x_0 + z$ with $z \in H$ (why?), and such an x satisfies

$$\|x\| \;=\; d(x,0) \;\geq\; d(x,H) \;=\; |\lambda|\, d(x_0,H) \tag{3.64}$$

(why?) and

$$\ell(x) \;=\; \lambda \tag{3.65}$$

(why?), hence

$$|\ell(x)| \;\leq\; \frac{1}{d(x_0,H)}\, \|x\| \;, \tag{3.66}$$

which proves that ℓ is continuous.

The quantitative relation clearly holds when ℓ is discontinuous, because we have $\|\ell\|_{X\to\mathbb{R}} = +\infty$ and $d_X(0, A_\ell) = 0$ (why?). When ℓ is continuous, we have

$$\|\ell\|_{X^*} \;=\; \sup_{x\neq 0} \frac{|\ell(x)|}{\|x\|} \;=\; \frac{1}{\displaystyle\inf_{x\colon \ell(x)\neq 0} \frac{\|x\|}{|\ell(x)|}} \;=\; \frac{1}{\displaystyle\inf_{x\colon \ell(x)=1} \|x\|} \;=\; \frac{1}{d(0,A_\ell)} \tag{3.67}$$

(you should make sure you understand each equality here). □

In order to make further progress in the theory of dual spaces, we need to ensure that the dual space X^* is "large enough", i.e. that there are "enough" continuous linear functionals. What is "enough"? Well, first of all, we would like the continuous linear functionals to *separate points of* X: that is, for each pair $x, y \in X$ with $x \neq y$, there should exist $\ell \in X^*$ such that $\ell(x) \neq \ell(y)$. By linearity this is equivalent to the assertion that for every $x \in X$ with $x \neq 0$, there exists $\ell \in X^*$ such that $\ell(x) \neq 0$. A few weeks from now we will deduce, as a corollary of the Hahn–Banach theorem, that this is indeed true, and indeed that ℓ can be chosen to make $\ell(x)$ as large as it is allowed to be:

Proposition 3.31 (Corollary of Hahn–Banach theorem, to be proven later) *Let X be a normed linear space. Then for each nonzero $x_0 \in X$, there exists a nonzero $\ell_0 \in X^*$ such that $\ell_0(x_0) = \|\ell_0\|_{X^*}\, \|x_0\|_X$.*

Equivalently, we can choose $\ell_0 \in X^*$ with $\|\ell_0\|_{X^*} = 1$ such that $\ell_0(x_0) = \|x_0\|_X$.

Let us go a little bit farther to explore one of the consequences of Proposition 3.31. Given a normed linear space, we can form the dual space X^* and then the **second dual space** $(X^*)^*$ [usually written X^{**} for short]. It is automatically complete, hence a Banach space.

Proposition 3.32 *Let X be a normed linear space, and let $x \in X$. Then the linear functional $\widehat{x}$ on X^* defined by*

$$\widehat{x}(\ell) \;=\; \ell(x) \tag{3.68}$$

is continuous and hence belongs to X^{**}. Moreover, the map $\iota\colon x \mapsto \widehat{x}$ is an isometric isomorphism of X onto its image $\iota[X]$, which is a linear subspace of X^{**}. If X is complete, then $\iota[X]$ is a closed *linear subspace of X^{**}.*

PROOF. $\widehat{x}$ is obviously a linear functional on X^*. Since

$$|\widehat{x}(\ell)| \;=\; |\ell(x)| \;\le\; \|x\|_X \, \|\ell\|_{X^*} \;, \tag{3.69}$$

$\widehat{x}$ is in fact a *continuous* linear functional on X^* (i.e. belongs to X^{**}) with $\|\widehat{x}\|_{X^{**}} \le \|x\|_X$. On the other hand, for any nonzero $x \in X$, Proposition 3.31 tells us that there exists $\ell \in X^*$ with $\|\ell\|_{X^*} = 1$ and $\widehat{x}(\ell) = \ell(x) = \|x\|_X$, so $\|\widehat{x}\|_{X^{**}} \ge \|x\|_X$. Therefore $\|\widehat{x}\|_{X^{**}} = \|x\|_X$. In particular, $\widehat{x} \neq 0$ whenever $x \neq 0$.

The map $\iota\colon x \mapsto \widehat{x}$ is obviously linear, and we have just seen that it is injective and an isometry from X into X^{**}. So it is an isometric isomorphism from X onto its image $\iota[X]$, which is a linear subspace of X^{**}. If X is complete, then by Proposition 3.17 so is $\iota[X]$, so it must be closed in X^{**}. $\square$

The map ι is called the **natural embedding** of X into X^{**}. It is then natural to ask whether the image of $\iota[X]$ is all of X^{**} or is only a proper subspace. We make the following definition:

Definition 3.33 *A normed linear space X is said to be* **reflexive** *if the natural embedding maps X* onto *X^{**}, i.e. if $\iota[X] = X^{**}$.*

For instance, all finite-dimensional normed linear spaces are reflexive (can you work out a proof of this?), and we will see next week that all Hilbert spaces are reflexive. Among the ℓ^p spaces, we will see that the spaces with $1 < p < \infty$ are reflexive, while ℓ^1 and ℓ^∞ are nonreflexive. One of the central themes in functional analysis is to study the property of reflexivity and to relate it to other properties that a Banach space may or may not possess.

Warning: Please note that reflexivity says *more* than just that X is isometrically isomorphic to X^{**}: it says that X is isometrically isomorphic to X^{**} *under the natural embedding*. Indeed, R.C. James has constructed an example of a nonreflexive separable Banach space X for which X is isometrically isomorphic to X^{**}! See R.C. James, A non-reflexive Banach space isometric with its second conjugate space, *Proc. Nat. Acad. Sci. USA* **37**, 174–177 (1951).

Here is one final important concept:

Definition 3.34 *Let X and Y be normed linear spaces, and let $T\colon X \to Y$ be a continuous linear mapping. Then the* **adjoint** *(or* **dual** *or* **transpose***) map $T^*\colon Y^* \to X^*$ [note the reversal of order here!] is defined by*

$$(T^*(\ell))(x) \;=\; \ell(Tx) \tag{3.70}$$

for $\ell \in Y^*$ and $x \in X$.

First of all, we should check that this definition makes sense! It is obvious that $(T^*(\ell))(x) = \ell(Tx)$ is linear in x, so $T^*(\ell)$ is indeed a linear functional on X. Moreover, we have

$$|(T^*(\ell))(x)| \;=\; |\ell(Tx)| \;\le\; \|\ell\|_{Y^*} \, \|Tx\|_Y \;\le\; \|\ell\|_{Y^*} \, \|T\|_{X\to Y} \, \|x\|_X \;, \tag{3.71}$$

so $T^*(\ell)$ is a *bounded* linear functional on X, of norm at most $\|\ell\|_{Y^*} \, \|T\|_{X \to Y}$. Finally, the map $T^*\colon Y^* \to X^*$ given by $T^*\colon \ell \mapsto T^*(\ell)$ is manifestly linear; and we have just shown that it is a *bounded* linear operator, whose operator norm satisfies

$$\|T^*\|_{Y^* \to X^*} \;\le\; \|T\|_{X \to Y} \tag{3.72}$$

(why?). So we indeed have $T^* \in \mathcal{B}(Y^*, X^*)$, as desired.

Lemma 3.35 *We have* $\|T^*\|_{Y^* \to X^*} = \|T\|_{X \to Y}$.

Proof. We have just shown that $\|T^*\|_{Y^* \to X^*} \le \|T\|_{X \to Y}$, so we need only show the reverse inequality. So let $x \in X$. Then by Proposition 3.31 there exists $\ell \in Y^*$ such that $\|\ell\|_{Y^*} = 1$ and $\ell(Tx) = \|Tx\|_Y$. Therefore

$$\begin{aligned} \|Tx\|_Y \;=\; \ell(Tx) \;=\; (T^*(\ell))(x) \;&\le\; \|T^*(\ell)\|_{X^*} \, \|x\|_X \\ &\le\; \|T^*\|_{Y^* \to X^*} \, \|\ell\|_{Y^*} \, \|x\|_X \;=\; \|T^*\|_{Y^* \to X^*} \, \|x\|_X \,. \end{aligned} \tag{3.73}$$

Since this holds for every $x \in X$, we have shown that $\|T\|_{X \to Y} \le \|T^*\|_{Y^* \to X^*}$. $\square$

Lemma 3.36

(a) If $T\colon X \to Y$ is a topological isomorphism, then $T^*\colon Y^* \to X^*$ is also a topological isomorphism, and in fact $(T^*)^{-1} = (T^{-1})^*$.

(b) If $T\colon X \to Y$ is an isometric isomorphism, then $T^*\colon Y^* \to X^*$ is also an isometric isomorphism.

Proof. (a) Let T be a topological isomorphism, and let us write $S = T^{-1}$. Since by hypothesis $S \in \mathcal{B}(Y, X)$, the adjoint $S^* \in \mathcal{B}(X^*, Y^*)$ is well-defined. Then, for any $x \in X$ and $\ell \in X^*$, we have by definition

$$(T^*(S^*(\ell)))(x) \;=\; (S^*(\ell))(Tx) \;=\; \ell(S(Tx)) \;=\; \ell(x) \tag{3.74}$$

(why?), so that $T^*(S^*(\ell)) = \ell$, i.e. $T^* \circ S^* = I_{X^*}$. A similar calculation shows that $S^* \circ T^* = I_{Y^*}$. Hence T^* and S^* are inverses, as was to be shown.

(b) Let T be an isometric isomorphism. Then we have just shown that T^* is a topological isomorphism; we need only show that it is an isometry. But this is easy: since T is an isometric isomorphism, we have $\|T\|_{X \to Y} = 1$ and $\|T^{-1}\|_{Y \to X} = 1$. But then by Lemma 3.35 we have $\|T^*\|_{Y^* \to X^*} = \|T\|_{X \to Y}$ and $\|(T^*)^{-1}\|_{X^* \to Y^*} = \|(T^{-1})^*\|_{X^* \to Y^*} = \|T^{-1}\|_{Y \to X} = 1$ (why does the first equality hold?). So T^* is an isometry. $\square$

Much more can be said about dual spaces and adjoint operators, but this is enough for now.

Duals of the sequence spaces c_0 and ℓ^p $(1 \le p < \infty)$

We are now ready to determine the duals of the sequence spaces c_0 and ℓ^p for $1 \le p < \infty$ (but *not* ℓ^∞, which is more difficult). We shall do this "with our bare hands", i.e. we shall not use the not-yet-proven Proposition 3.31 or any other "deep" result.

The intuition behind all of this is the following elementary fact: every linear functional on $\mathbb{R}^n$ is of the form

$$\iota(a)\colon\ x \ \mapsto\ \sum_{i=1}^{n} a_i x_i \tag{3.75}$$

for some unique $a \in \mathbb{R}^n$; and conversely, every $a \in \mathbb{R}^n$ defines a linear functional $\iota(a)$ on $\mathbb{R}^n$ by the above formula.

It is natural to try to adapt this idea to spaces of *infinite* sequences, defining linear functionals by the formula (3.75) but with the sum now taken to infinity. But now we have to be careful about the precise spaces in which x and a are taken to lie, and to make sure that the infinite sum is convergent and defines a *bounded* linear functional.

Let us start with the case $X = c_0$ (equipped as usual with the sup norm). We will show that the dual X^* can be naturally identified with the space ℓ^1. Here and in what follows, we will write e_i to denote the infinite sequence having a 1 in the ith coordinate and 0 elsewhere.

Proposition 3.37 *For any $a \in \ell^1$, the sum $\sum\limits_{i=1}^{\infty} a_i x_i$ is absolutely convergent for all $x \in c_0$, and defines a linear functional*

$$\iota(a)\colon\ x \ \mapsto\ \sum_{i=1}^{\infty} a_i x_i \tag{3.76}$$

on c_0 that is bounded (i.e. belongs to c_0^*) and has norm $\|\iota(a)\|_{c_0^*} = \|a\|_{\ell^1}$. Moreover, every bounded linear functional on c_0 arises in this way, and the map ι is an isometric isomorphism of ℓ^1 onto c_0^*.

PROOF. Fix $a \in \ell^1$. Then for all $x \in c_0$ we have

$$\sum_{i=1}^{\infty} |a_i x_i| \ \le\ \|a\|_{\ell^1}\, \|x\|_{c_0} \tag{3.77}$$

(why?), so the sum $\sum\limits_{i=1}^{\infty} a_i x_i$ is absolutely convergent and satisfies

$$\left| \sum_{i=1}^{\infty} a_i x_i \right| \ \le\ \sum_{i=1}^{\infty} |a_i x_i| \ \le\ \|a\|_{\ell^1}\, \|x\|_{c_0}\ . \tag{3.78}$$

It follows that $\iota(a) \in c_0^*$ and that $\|\iota(a)\|_{c_0^*} \le \|a\|_{\ell^1}$.

To see the reverse inequality, consider for each n the vector $x^{(n)} \in c_{00} \subset c_0$ defined by

$$x_i^{(n)} \ =\ \begin{cases} \operatorname{sgn}(a_i) & \text{for } 1 \le i \le n \\ 0 & \text{for } i > n \end{cases} \tag{3.79}$$

where

$$\operatorname{sgn}(x) \;=\; \begin{cases} 1 & \text{if } x > 0 \\ -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \end{cases} \tag{3.80}$$

We then have $\|x^{(n)}\|_{c_0} \le 1$ (why?) and

$$\iota(a)(x^{(n)}) \;=\; \sum_{i=1}^{n} a_i \, \operatorname{sgn}(a_i) \;=\; \sum_{i=1}^{n} |a_i| \;. \tag{3.81}$$

It follows that $\|\iota(a)\|_{c_0^*} \ge \sum\limits_{i=1}^{n} |a_i|$ for all n, and hence $\|\iota(a)\|_{c_0^*} \ge \|a\|_{\ell^1}$.

The map $\iota\colon \ell^1 \to c_0^*$ is clearly linear and injective (why is it injective?), so it remains only to show that it is surjective, i.e. that every $\ell \in c_0^*$ is of the form $\iota(a)$ for some (necessarily unique) $a \in \ell^1$. Indeed, we claim that $\ell = \iota(a)$ where a is given by the formula

$$a_i \;=\; \ell(e_i) \tag{3.82}$$

(recall that $e_i \in c_0$ is the vector having a 1 in the ith coordinate and 0 elsewhere). To see this, note first that if $x = (x_1, x_2, \ldots) \in c_0$, then

$$\left\| x - \sum_{i=1}^{n} x_i e_i \right\|_{c_0} \;=\; \sup_{i>n} |x_i| \;\to\; 0 \tag{3.83}$$

as $n \to \infty$ (why?), i.e. $\sum\limits_{i=1}^{n} x_i e_i \to x$ in c_0 as $n \to \infty$. Now consider any $\ell \in c_0^*$; since ℓ is continuous, we have

$$\ell\left(\sum_{i=1}^{n} x_i e_i\right) \;=\; \sum_{i=1}^{n} x_i \, \ell(e_i) \;\to\; \ell(x) \tag{3.84}$$

as $n \to \infty$, or in other words

$$\ell(x) \;=\; \sum_{i=1}^{\infty} a_i x_i \tag{3.85}$$

where the sum is convergent (but we have not yet proven that the convergence is absolute). The final step is to prove that $a \in \ell^1$ (this will imply in particular that the convergence is absolute). To see this, let vectors $x^{(n)} \in c_{00} \subset c_0$ be defined once again by (3.79), and recall that $\|x^{(n)}\|_{c_0} \le 1$. Then

$$\sum_{i=1}^{n} |a_i| \;=\; \sum_{i=1}^{n} a_i \, \operatorname{sgn}(a_i) \;=\; \ell(x^{(n)}) \;\le\; \|\ell\|_{c_0^*} \;, \tag{3.86}$$

so taking $n \to \infty$ we conclude that $a \in \ell^1$ (with $\|a\|_{\ell^1} \le \|\ell\|_{c_0^*}$). It now follows that $\ell = \iota(a)$ (and hence, from the first part of the proof, that $\|a\|_{\ell^1} = \|\ell\|_{c_0^*}$). $\square$

Remarks. 1. Since a dual space is automatically complete by Proposition 3.19, and Proposition 3.37 tells us that ℓ^1 is isometrically isomorphic to the dual space c_0^*, we obtain from Proposition 3.37 an alternate proof of the completeness of ℓ^1.

2. The first part of this proof works for $x \in \ell^\infty$ just as well as it does for $x \in c_0$. It follows that, for any $a \in \ell^1$, the formula (3.76) defines a bounded linear functional $\iota(a) \in (\ell^\infty)^*$ of norm $\|\iota(a)\|_{(\ell^\infty)^*} = \|a\|_{\ell^1}$, and that the map ι is an isometric isomorphism of ℓ^1 *into* $(\ell^\infty)^*$. But this map is *not* onto: we will see later that there exist linear functionals in $(\ell^\infty)^*$ that are *not* of the form $\iota(a)$ for some $a \in \ell^1$.

3. Essentially the same proof works for the *complex* sequence spaces; we need only replace $\operatorname{sgn}(a_i)$ in (3.79) by $\overline{a_i}/|a_i|$, where $\overline{}$ denotes complex conjugate.

Let us now treat, by a similar method, the case $X = \ell^p$ with $1 \le p < \infty$ (note that the case $p = 1$ is included but $p = \infty$ is not). We let $q \in (1, \infty]$ be the conjugate index defined by $1/p + 1/q = 1$.

Proposition 3.38 *Fix $p \in [1, \infty)$, and let $q \in (1, \infty]$ be the conjugate index defined by $1/p + 1/q = 1$. Then, for any $a \in \ell^q$, the sum $\sum\limits_{i=1}^{\infty} a_i x_i$ is absolutely convergent for all $x \in \ell^p$, and defines a linear functional*

$$\iota(a)\colon\ x \;\mapsto\; \sum_{i=1}^{\infty} a_i x_i \tag{3.87}$$

on ℓ^p that is bounded (i.e. belongs to $(\ell^p)^*$) and has norm $\|\iota(a)\|_{(\ell^p)^*} = \|a\|_{\ell^q}$. Moreover, every bounded linear functional on ℓ^p arises in this way, and the map ι is an isometric isomorphism of ℓ^q onto $(\ell^p)^*$.

We will imitate the proof of the preceding proposition, but using Hölder's inequality in place of the elementary bound $\sum\limits_{i=1}^{\infty} |a_i x_i| \le \|a\|_1 \, \|x\|_\infty$. This will handle the case $1 < p < \infty$. The case $p = 1$ is even easier (i.e. uses elementary inequalities rather than Hölder) and is left as an exercise for you.

Proof of Proposition 3.38 for the case $1 < p < \infty$. Fix $a \in \ell^q$. Then for all $x \in \ell^p$ we have

$$\sum_{i=1}^{\infty} |a_i x_i| \;\le\; \|a\|_q \, \|x\|_p \tag{3.88}$$

by Hölder's inequality, so the sum $\sum\limits_{i=1}^{\infty} a_i x_i$ is absolutely convergent and satisfies

$$\left| \sum_{i=1}^{\infty} a_i x_i \right| \;\le\; \sum_{i=1}^{\infty} |a_i x_i| \;\le\; \|a\|_q \, \|x\|_p \, . \tag{3.89}$$

It follows that $\iota(a) \in (\ell^p)^*$ and that $\|\iota(a)\|_{(\ell^p)^*} \le \|a\|_q$.

To see the reverse inequality, consider for each n the vector $x^{(n)} \in c_{00} \subset \ell^p$ defined by

$$x_i^{(n)} \;=\; \begin{cases} \operatorname{sgn}(a_i)\, |a_i|^{q-1} & \text{for } 1 \le i \le n \\ 0 & \text{for } i > n \end{cases} \tag{3.90}$$

We then have

$$\|x^{(n)}\|_p \;=\; \left(\sum_{i=1}^{n} |a_i|^{(q-1)p}\right)^{1/p} \;=\; \left(\sum_{i=1}^{n} |a_i|^q\right)^{1/p} \tag{3.91}$$

since $(q-1)p = q$, while

$$\iota(a)(x^{(n)}) \;=\; \sum_{i=1}^{n} |a_i|^q \tag{3.92}$$

(why?). It follows that

$$\|\iota(a)\|_{(\ell^p)^*} \;\ge\; \frac{\iota(a)(x^{(n)})}{\|x^{(n)}\|_p} \;=\; \left(\sum_{i=1}^{n} |a_i|^q\right)^{1-\frac{1}{p}} \;=\; \left(\sum_{i=1}^{n} |a_i|^q\right)^{1/q} \tag{3.93}$$

since $1 - 1/p = 1/q$. Taking $n \to \infty$, we conclude that $\|\iota(a)\|_{(\ell^p)^*} \ge \|a\|_q$.

The map $\iota\colon \ell^q \to (\ell^p)^*$ is clearly linear and injective (why is it injective?), so it remains only to show that it is surjective, i.e. that every $\ell \in (\ell^p)^*$ is of the form $\iota(a)$ for some (necessarily unique) $a \in \ell^q$. Indeed, we claim that $\ell = \iota(a)$ where a is given by the formula

$$a_i \;=\; \ell(e_i)\ . \tag{3.94}$$

To see this, note first that if $x = (x_1, x_2, \ldots) \in \ell^p$, then

$$\left\| x - \sum_{i=1}^{n} x_i e_i \right\|_p \;=\; \left(\sum_{i=n+1}^{\infty} |x_i|^p\right)^{1/p} \;\to\; 0 \tag{3.95}$$

as $n \to \infty$ (why?), i.e. $\sum\limits_{i=1}^{n} x_i e_i \to x$ in ℓ^p as $n \to \infty$. Now consider any $\ell \in (\ell^p)^*$; since ℓ is continuous, we have

$$\ell\left(\sum_{i=1}^{n} x_i e_i\right) \;=\; \sum_{i=1}^{n} x_i\, \ell(e_i) \;\to\; \ell(x) \tag{3.96}$$

as $n \to \infty$, or in other words

$$\ell(x) \;=\; \sum_{i=1}^{\infty} a_i x_i \tag{3.97}$$

where the sum is convergent (but we have not yet proven that the convergence is absolute). The final step is to prove that $a \in \ell^q$ (this will imply in particular that the convergence is absolute). To see this, let vectors $x^{(n)} \in c_{00} \subset \ell^p$ be defined once again by (3.90), and recall that the formula (3.91) for $\|x^{(n)}\|_p$. Then

$$\sum_{i=1}^{n} |a_i|^q \;=\; \sum_{i=1}^{n} a_i\ \mathrm{sgn}(a_i)\, |a_i|^{q-1} \;=\; \ell(x^{(n)}) \;\le\; \|\ell\|_{(\ell^p)^*}\, \|x^{(n)}\|_p \;=\; \|\ell\|_{(\ell^p)^*} \left(\sum_{i=1}^{n} |a_i|^q\right)^{1/p}\ , \tag{3.98}$$

hence

$$\left(\sum_{i=1}^{n} |a_i|^q\right)^{1/q} \;\le\; \|\ell\|_{(\ell^p)^*} \tag{3.99}$$

since $1 - 1/p = 1/q$. Taking $n \to \infty$ we conclude that $a \in \ell^q$ (with $\|a\|_q \le \|\ell\|_{(\ell^p)^*}$). It now follows that $\ell = \iota(a)$ (and hence, from the first part of the proof, that $\|a\|_q = \|\ell\|_{(\ell^p)^*}$). $\square$

Since a dual space is automatically complete by Proposition 3.19, and Proposition 3.38 tells us that ℓ^q for $1 < q \le \infty$ is isometrically isomorphic to the dual space $(\ell^p)^*$, we obtain from Proposition 3.38 an alternate proof of the completeness of the spaces ℓ^q for $1 < q \le \infty$.

Even more importantly, we can obtain from Proposition 3.38 the following corollary:

Corollary 3.39 *The spaces ℓ^p for $1 < p < \infty$ are reflexive.*

Remark. We will see later that ℓ^1 and ℓ^∞ are nonreflexive.

At first glance one might think that the reflexivity of ℓ^p for $1 < p < \infty$ is an *immediate* consequence of Proposition 3.38: after all, the dual of ℓ^p is ℓ^q (where $1/p + 1/q = 1$), and the dual of ℓ^q is ℓ^p, so the second dual of ℓ^p is ℓ^p, and we are done — right? *Well, no ... not quite!* As I have stressed, reflexivity says *more* than just that X is isometrically isomorphic to X^{**}: it says that X is isometrically isomorphic to X^{**} *under the natural embedding*. So we need to verify that the isometric isomorphisms constructed by Proposition 3.38 piece together to yield the natural embedding of ℓ^p into $(\ell^p)^{**}$. This is not difficult, but it requires a bit of attention to detail.

Let $\iota_q\colon \ell^q \to (\ell^p)^*$ be the isometric isomorphism constructed by Proposition 3.38: that is,

$$(\iota_q(a))(x) \;=\; \sum_{i=1}^{\infty} a_i x_i \tag{3.100}$$

for $a \in \ell^q$ and $x \in \ell^p$. Of course we also have $\iota_p\colon \ell^p \to (\ell^q)^*$ defined in the same way. Finally, let $J_p\colon \ell^p \to (\ell^p)^{**}$ be the natural embedding. [I now use the letter J rather than ι to avoid confusion.] Since ι_q maps ℓ^q to $(\ell^p)^*$, its adjoint ι_q^* maps $(\ell^p)^{**}$ to $(\ell^q)^*$, and I claim that:

Lemma 3.40 $\iota_q^* \circ J_p = \iota_p$.

To start with, you should check that the statement of this lemma makes sense, i.e. that the maps act between the correct spaces. We can then prove the lemma as follows:

Proof. We need to show that $(\iota_q^* \circ J_p)(x) = \iota_p(x)$ for all $x \in \ell^p$, or in other words that

$$((\iota_q^* \circ J_p)(x))(y) \;=\; (\iota_p(x))(y) \tag{3.101}$$

for all $x \in \ell^p$ and all $y \in \ell^q$. Now

$$(\iota_p(x))(y) \;=\; \sum_{i=1}^{\infty} x_i y_i \tag{3.102}$$

by definition of ι_p. On the other hand, we have

$$((\iota_q^* \circ J_p)(x))(y) \;=\; (\iota_q^*(J_p(x)))(y) \;=\; (J_p(x))(\iota_q(y)) \;=\; (\iota_q(y))(x) \;=\; \sum_{i=1}^{\infty} y_i x_i \tag{3.103}$$

by the definitions of adjoint, J_p and ι_q. $\square$

Proof of Corollary 3.39. By the lemma we have $\iota_q^* \circ J_p = \iota_p$. But by Lemma 3.36(b), the map ι_q^* is an isometric isomorphism. Then $J_p = (\iota_q^*)^{-1} \circ \iota_p$ is an isometric isomorphism, as was to be proved. $\square$