MATH41001/MATH61001: LINEAR ANALYSIS

EXTRA READING

A. APPLICATION OF THE STONE-WEIERSTRASS THEOREM: SEPARABILITY

Definition. We say that a metric space $\mathcal{Z}$ is *separable* if it contains a countable subset $\{z_n\}_{n=0}^{\infty}$ which is dense in $\mathcal{Z}$ (i.e., for all $z \in \mathcal{Z}$ and all $\epsilon > 0$, there exists z_n such that $d_{\mathcal{Z}}(z, z_n) < \epsilon$.

We shall apply this to $C(X, \mathbb{R})$, where X is a compact metric space, with the metric given by the uniform norm $\|\cdot\|_\infty$. Then the definition requires that there exists a countable set $\{f_n\}_{n=0}^{\infty} \subset C(X, \mathbb{R})$ such that for all $f \in C(X, \mathbb{R})$ and all $\epsilon > 0$, there exists f_n such that $\|f - f_n\|_\infty < \epsilon$.

Theorem A.1. *Suppose that X is a compact metric space. Then $C(X, \mathbb{R})$ is separable (with respect to the uniform norm).*

To prove this theorem, we will use the following exercise for the reader.

Exercise. If X is a compact metric space then X is separable.

Proof of Theorem A.1. Since X is separable, there exists $\{x_n\}_{n=0}^{\infty} \subset X$ such that for all $x \in X$ and all $\epsilon > 0$ there exists x_n with $d_X(x, x_n) < \epsilon$. We define a sequence of functions $g_n \in C(X, \mathbb{R})$ by $g_n(x) = d_X(x, x_n)$.

Given an arbitrary finite string of indices $(n_1, \dots, n_k)$, $n_i \in \mathbb{N}$, we define

$$g_{n_1,\dots,n_k}(x) = g_{n_1}(x) \cdots g_{n_k}(x).$$

This gives us a countable collection of functions. Now choose $\mathcal{A}$ to be the set of all finite linear combinations of these functions and the constant function 1:

$$\mathcal{A} = \left\{ a + \sum_{(n_1,\dots,n_k)} a_{n_1,\dots,n_k} g_{n_1,\dots,n_k}(x) \,:\, a, a_{n_1,\dots,n_k} \in \mathbb{R} \right\},$$

where the above sums are all finite.

Clearly, $\mathcal{A}$ is an algebra. Also, $\mathcal{A}$ contains the non-zero constant function 1, so we only need to check that $\mathcal{A}$ separates points.

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

Suppose that $x, y \in X$ with $x \neq y$. Let $d_X(x, y) = 2\epsilon > 0$ and choose x_n with $d_X(x, x_n) < \epsilon$. Now consider $g_n \in \mathcal{A}$; by definition, $g_n(x) < \epsilon$. However,

$$g_n(y) = d_X(y, x_n) \geq d_X(x, y) - d_X(x, x_n) \geq 2\epsilon - \epsilon = \epsilon,$$

so that $g_n(x) \neq g_n(y)$.

Therefore $\mathcal{A}$ satisfies the hypotheses of the Stone-Weierstrass Theorem and is thus a dense subset of $C(X, \mathbb{R})$.

Now define

$$\mathcal{A}_{\mathbb{Q}} = \left\{ a + \sum_{(n_1,\dots,n_k)} a_{n_1,\dots,n_k} g_{n_1,\dots,n)}(x) \,:\, a, a_{n_1,\dots,n_k} \in \mathbb{Q} \right\} \subset \mathcal{A}.$$

Restricting the coefficients to lie in $\mathbb{Q}$ means that $\mathcal{A}_{\mathbb{Q}}$ is countable. (Check that this is true.) Furthermore, since $\mathbb{Q}$ is dense in $\mathbb{R}$, given $\epsilon > 0$ and $f \in \mathcal{A}$, one can find $\widetilde{f} \in \mathcal{A}_{\mathbb{Q}}$ such that $\|f - \widetilde{f}\|_\infty < \epsilon$. (Check that this is true.) Since $\mathcal{A}$ is dense in $C(X, \mathbb{R})$, this shows that $\mathcal{A}_{\mathbb{Q}}$ is also dense in $C(X, \mathbb{R})$. Thus $C(X, \mathbb{R})$ is separable. $\square$

B. MINKOWSKI'S INEQUALITY

Recall that in Chapter 2 we use the following.

Theorem B.1 = Lemma 2.3 (Minkowski's Inequality). *For* $p \geq 1$,

$$\left(\sum_{i=1}^{n} |a_i + b_i|^p\right)^{1/p} \leq \left(\sum_{i=1}^{n} |a_i|^p\right)^{1/p} + \left(\sum_{i=1}^{n} |b_i|^p\right)^{1/p}.$$

Furthermore, if $a_i, b_i \in \mathbb{C}$, $i \geq 1$, *are such that* $\sum_{i=1}^{\infty} |a_i|^p$ *and* $\sum_{i=1}^{\infty} |b_i|^p$ *both converge then* $\sum_{i=1}^{\infty} |a_i + b_i|^p$ *converges and*

$$\left(\sum_{i=1}^{\infty} |a_i + b_i|^p\right)^{1/p} \leq \left(\sum_{i=1}^{\infty} |a_i|^p\right)^{1/p} + \left(\sum_{i=1}^{\infty} |b_i|^p\right)^{1/p}.$$

To prove Minkowski's inequality, we need the following generalization of the Cauchy-Schwartz inequality.

Theorem B.2 (Hölder's Inequality). *Suppose that* $p, q > 1$ *satisfy* $1/p + 1/q = 1$. *Then, for* $a_i, b_i \in \mathbb{C}$, $i = 1, \ldots, n$,

$$\sum_{i=1}^{n} |a_i b_i| \leq \left(\sum_{i=1}^{n} |a_i|^p\right)^{1/p} \left(\sum_{i=1}^{n} |b_i|^q\right)^{1/q},$$

with equality if and only if $|a_i|^p/|b_i|^q$ *is constant independent of* i.

Furthermore, if $a_i, b_i \in \mathbb{C}$, $i \geq 1$, *are such that* $\sum_{i=1}^{\infty} |a_i|^p$ *and* $\sum_{i=1}^{\infty} |b_i|^q$ *both converge then* $\sum_{i=1}^{\infty} |a_i b_i|$ *converges and*

$$\sum_{i=1}^{\infty} |a_i b_i| \leq \left(\sum_{i=1}^{\infty} |a_i|^p\right)^{1/p} \left(\sum_{i=1}^{\infty} |b_i|^q\right)^{1/q},$$

with equality if and only if $|a_i|^p/|b_i|^q$ *is constant independent of* i.

We just give the proofs in the finite case. The infinite case is identical – one just needs to check that everything converges at each stage.

Proof of Theorem B.2.
Step 1. We shall show that if $a, b > 0$ and $0 < \lambda < 1$ then

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

with equality if and only if $a = b$. Set $t = a/b$; then (dividing by b) the result is equivalent to showing that $t^{\lambda} \leq \lambda t + (1-\lambda)$.

Set $\phi(t) = \lambda t + (1-\lambda) - t^{\lambda}$; then we need to show that $\phi(t) \geq 0$. However, $\phi'(t) = \lambda - \lambda t^{\lambda-1} = \lambda(1-t^{\lambda-1})$, so

$$\phi'(t) \quad \begin{cases} < 0 \text{ if } t < 1 \\ = 0 \text{ if } t = 1 \\ > 0 \text{ if } t > 1 \end{cases}.$$

Since $\phi(1) = 0$, this gives the result.

Step 2. Define

$$A_i = \frac{|a_i|^p}{\sum_{i=1}^n |a_i|^p}, \quad B_i = \frac{|b_i|^q}{\sum_{i=1}^n |b_i|^q}.$$

Let $\lambda = 1/p$; then, by Step 1,

$$A_i^{1/p} B_i^{1/q} \leq \frac{A_i}{p} + \frac{B_i}{q}$$

(with equality if and only if $|a_i|^p/|b_i|^q = \left(\sum_{i=1}^n |a_i|^p\right) / \left(\sum_{i=1}^n |b_i|^q\right)$, a constant). Writing this out fully, we get, for any $i = 1, \ldots, n$,

$$\frac{|a_i|}{\left(\sum_{i=1}^n |a_i|^p\right)^{1/p}} \frac{|b_i|}{\left(\sum_{i=1}^n |b_i|^q\right)^{1/q}} \leq \frac{1}{p} \frac{|a_i|^p}{\sum_{i=1}^n |a_i|^p} + \frac{1}{q} \frac{|b_i|^q}{\sum_{i=1}^n |b_i|^q}.$$

Summing over $i = 1, \ldots, n$, we obtain

$$\frac{\sum_{i=1}^n |a_i||b_i|}{\left(\sum_{i=1}^n |a_i|^p\right)^{1/p} \left(\sum_{i=1}^n |b_i|^q\right)^{1/q}} \leq \frac{1}{p} + \frac{1}{q} = 1$$

(with equality if and only if $|a_i|^p/|b_i|^q$ is constant). This completes the proof of Hölder's inequality. □

Proof of Theorem B.1. The case $p = 1$ is clear, so suppose $p > 1$ and define $q > 1$ by $1/p + 1/q = 1$. We have that

$$\begin{aligned} &\sum_{i=1}^n |a_i + b_i|^p = \sum_{i=1}^n |a_i + b_i||a_i + b_i|^{p-1} \\ &\leq \sum_{i=1}^n |a_i||a_i + b_i|^{p-1} + \sum_{i=1}^n |b_i||a_i + b_i|^{p-1} \\ &\leq \left(\sum_{i=1}^n |a_i|^p\right)^{1/p} \left(\sum_{i=1}^n |a_i + b_i|^{(p-1)q}\right)^{1/q} + \left(\sum_{i=1}^n |b_i|^p\right)^{1/p} \left(\sum_{i=1}^n |a_i + b_i|^{(p-1)q}\right)^{1/q} \end{aligned}$$

(using Hölder's inequality). However,

$$(p-1)q = (p-1)\left(1 - \frac{1}{p}\right)^{-1} = p,$$

so we may rewrite the above inequality as

$$\sum_{i=1}^{n} |a_i + b_i|^p \leq \left(\left(\sum_{i=1}^{n} |a_i|^p \right)^{1/p} + \left(\sum_{i=1}^{n} |b_i|^p \right)^{1/p} \right) \left(\sum_{i=1}^{n} |a_i + b_i|^p \right)^{1/q} .$$

Dividing by $\left(\sum_{i=1}^{n} |a_i + b_i|^p\right)^{1/q}$, we get

$$\left(\sum_{i=1}^{n} |a_i + b_i|^p \right)^{1-1/q} \leq \left(\sum_{i=1}^{n} |a_i|^p \right)^{1/p} + \left(\sum_{i=1}^{n} |b_i|^p \right)^{1/p} .$$

Since $1 - 1/q = 1/p$, this is Minkowski's inequality. $\square$

C. l^p ($p > 1$) AND l^∞ ARE BANACH SPACES

The proofs of these results are similar to the proof that l^1 is a Banach space. We start with l^∞.

Theorem C.1. *l^∞ is a Banach space.*

Proof. We need to show that l^∞ is complete. This means that any Cauchy sequence $\{(x_i^{(n)})_{i=1}^\infty\}_{n=1}^\infty$ in l^∞ converges to a point $(x_i)_{i=1}^\infty \in l^\infty$ with respect to the norm $\|\cdot\|_\infty$. This means that one has to find a suitable candidate limit point $(x_i)_{i=1}^\infty$, show that it is in l^∞ and then show that

$$\lim_{n\to+\infty} \|(x_i^{(n)})_{i=1}^\infty - (x_i)_{i=1}^\infty\|_\infty = 0.$$

Now let's start the proof. Suppose that $\{(x_i^{(n)})_{i=1}^\infty\}_{n=1}^\infty$ is a Cauchy sequence for $\|\cdot\|_\infty$. Then, given $\epsilon > 0$, there exists $N \geq 1$ such that, for $n, m \geq N$ and any $i \geq 0$,

$$|x_i^{(n)} - x_i^{(m)}| \leq \|(x_i^{(n)})_{i=1}^\infty - (x_i^{(m)})_{i=1}^\infty\|_\infty < \epsilon. \qquad (*)$$

In other words, for each fixed i, $\{x_i^{(n)}\}_{n=1}^\infty$ is a Cauchy sequence in $\mathbb{C}$ and so, since $\mathbb{C}$ is complete, it has a limit $x_i \in \mathbb{C}$, say. We may now let $m \to +\infty$ in $(*)$ to obtain, for $n \geq N$,

$$|x_i^{(n)} - x_i| = \lim_{m\to+\infty} |x_i^{(n)} - x_i^{(m)}| \leq \limsup_{m\to+\infty} \|(x_i^{(n)})_{i=1}^\infty - (x_i^{(m)})_{i=1}^\infty\|_\infty \leq \epsilon. \qquad (**)$$

Then, in particular, for each i,

$$|x_i| = |x_i^{(N)} - (x_i^{(N)} - x_i)| \leq |x_i^{(N)}| + |x_i^{(N)} - x_i| \leq \|(x_i^{(N)})_{i=1}^\infty\|_\infty + \epsilon,$$

so, taking the supremum over i, $\|(x_i)_{i=1}^\infty\|_\infty < +\infty$, i.e., $(x_i)_{i=1}^\infty \in l^\infty$. Furthermore, taking the supremum over i in $(**)$, gives that, for $n \geq N$,

$$\|(x_i^{(n)})_{i=1}^\infty - (x_i)_{i=1}^\infty\|_\infty = \sup_{i\geq 1} |x_i^{(n)} - x_i| \leq \epsilon,$$

so $\lim_{n\to+\infty} \|(x_i^{(n)})_{i=1}^\infty - (x_i)_{i=1}^\infty\|_\infty = 0$, as required. $\square$

And now for l^p.

Theorem C.2. *l^p is a Banach space.*

Proof. We need to show that l^p is complete. This means that any Cauchy sequence $\{(x_i^{(n)})_{i=1}^\infty\}_{n=1}^\infty$ in l^p converges to a point $(x_i)_{i=1}^\infty \in l^p$ with respect to the norm $\|\cdot\|_p$. This means that one has to find a suitable candidate limit point $(x_i)_{i=1}^\infty$, show that it is in l^p and then show that

$$\lim_{n\to+\infty} \|(x_i^{(n)})_{i=1}^\infty - (x_i)_{i=1}^\infty\|_p = 0.$$

Now let's start the proof. Suppose that $\{(x_i^{(n)})_{i=1}^\infty\}_{n=1}^\infty$ is a Cauchy sequence for $\|\cdot\|_p$. Then, given $\epsilon > 0$, there exists $N \geq 1$ such that, for $n, m \geq N$ and any $i \in \mathbb{N}$,

$$|x_i^{(n)} - x_i^{(m)}|^p \leq \sum_{i=1}^\infty |x_i^{(n)} - x_i^{(m)}|^p = \|(x_i^{(n)})_{i=1}^\infty - (x_i^{(m)})_{i=1}^\infty\|_p^p < \epsilon^p. \qquad (*)$$

So, for each fixed i, $\{x_i^{(n)}\}_{n=1}^{\infty}$ is a Cauchy sequence in $\mathbb{C}$ and thus has a limit $x_i \in \mathbb{C}$, say.

Next, note that, for any $M \geq 1$ and $n, m \geq N$, we have

$$\sum_{i=1}^{M} |x_i^{(n)} - x_i^{(m)}|^p \leq \sum_{i=1}^{\infty} |x_i^{(n)} - x_i^{(m)}|^p < \epsilon^p.$$

If we let $m \to +\infty$, this gives

$$\sum_{i=1}^{M} |x_i^{(n)} - x_i|^p \leq \epsilon^p, \tag{$**$}$$

for any $M \geq 1$ and $n \geq N$. By applying Minkowski's inequality (Theorem B.1)) to $a_i = x_i^{(N)} - x_i$ and $b_i = x_i^{(N)}$,

$$\begin{aligned}\left(\sum_{i=1}^{M} |x_i|^p\right)^{1/p} &\leq \left(\sum_{i=1}^{M} |x_i^{(N)} - x_i|^p\right)^{1/p} + \left(\sum_{i=1}^{M} |x_i^{(N)}|^p\right)^{1/p} \\ &\leq \left(\sum_{i=1}^{M} |x_i^{(N)} - x_i|^p\right)^{1/p} + \left(\sum_{i=1}^{\infty} |x_i^{(N)}|^p\right)^{1/p} \\ &\leq \epsilon + \|(x_i^{(N)})_{i=1}^{\infty}\|_p.\end{aligned}$$

Letting $M \to +\infty$, we see that $\left(\sum_{i=1}^{\infty} |x_i|^p\right)^{1/p}$ is finite, so $(x_i)_{i=1}^{\infty} \in l^p$.

Finally, letting $M \to +\infty$ in $(**)$ gives

$$\sum_{i=1}^{\infty} |x_i^{(n)} - x_i|^p \leq \epsilon^p,$$

for all $n \geq N$, so that $\lim_{n \to +\infty} \|(x_i^{(n)})_{i=1}^{\infty} - (x_i)_{i=1}^{\infty}\|_p = 0$, as required. □

D. THE DUAL SPACE OF l^p

In this section we prove Theorem 3.6. We split it into two statements, about l^p and l^∞, and prove them separately. We write $a = (a_i)_{i=1}^\infty$ and $b = (b_i)_{i=1}^\infty$.

Theorem D.1. *Suppose that* $1 < p, q < +\infty$ *satisfy*

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Then the map

$$T : l^q \to (l^p)^* : b \mapsto f_b,$$

where

$$f_b(a) = \sum_{i=1}^\infty a_i b_i,$$

is an isometric isomorphism, so $(l^p)^* = l^q$.

Proof. In Lemma 3.3, we showed that, for $b \in l^q$, $T(b) = f_b$ is a well-defined element of $(l^p)^*$ and that $\|T(b)\| = \|f_b\| = \|b\|_q$. Furthermore, it is easy to see that the map

$$T : l^q \to (l^p)^* : b \mapsto f_b$$

is linear. (Check this.) Therefore T is an isometry and this also implies that T is injective. (Why?)

It remains to show that $T : l^q \to (l^p)^*$ is surjective. Given $f \in (l^p)^*$, write $f_i = f(e_i)$, $n \geq 1$, where

$$e_i = (\underbrace{0, \dots, 0}_{i-1}, 1, 0, \dots).$$

It is clear from the construction that $T((f_i)_{i=1}^\infty) = f$. Thus, to show that T is surjective, it will be enough to show that $(f_i)_{i=1}^\infty \in l^q$.

Claim: $x = (f_i)_{i=1}^\infty \in l^q$. Assume for a contradiction that

$$\sum_{i=1}^\infty |f_i|^q = +\infty.$$

Let $x^{(n)} = (f_1, f_2, \dots, f_n, 0, \dots)$, then, by assumption,

$$\|x^{(n)}\|_q = \left(\sum_{i=1}^n |f_i|^q \right)^{1/q} \to +\infty, \text{ as } n \to +\infty.$$

Fix n and let

$$y^{(n)} = \left(\frac{|f_1|}{f_1} \left(\frac{|f_1|^q}{\|x^{(n)}\|_q^q} \right)^{1/p}, \frac{|f_2|}{f_2} \left(\frac{|f_2|^q}{\|x^{(n)}\|_q^q} \right)^{1/p}, \dots, \frac{|f_n|}{f_n} \left(\frac{|f_n|^q}{\|x^{(n)}\|_q^q} \right)^{1/p}, 0, \dots \right).$$

(If $f_i = 0$ then we omit the term $|f_i|/f_i$.) Then

$$\|y^{(n)}\|_p = \left(\sum_{i=1}^{n} \frac{|f_i|^q}{\|x^{(n)}\|_q^q}\right)^{1/p} = \left(\frac{\|x^{(n)}\|_q^q}{\|x^{(n)}\|_q^q}\right)^{1/p} = 1$$

and

$$|y_i^{(n)}|^p = c|f_i|^q,$$

where $c = \|x^{(n)}\|_q^{-q}$. Furthermore, $y_i^{(n)} f_i = |y_i^{(n)}||f_i|$.

By the equality statement in Hölder's inequality, we have

$$\begin{aligned} f(y^{(n)}) = f\left(\sum_{i=1}^{n} y_i^{(n)} e_i\right) &= \sum_{i=1}^{n} y_i^{(n)} f_i \\ &= \sum_{i=1}^{n} |y_i^{(n)}||f_i| \\ &= \sum_{i=1}^{n} |y_i^{(n)}||x_i^{(n)}| \\ &= \|x^{(n)}\|_q \|y^{(n)}\|_p = \|x^{(n)}\|_q \end{aligned}$$

(since $\|y^{(n)}\|_p = 1$). Thus

$$\|x^{(n)}\|_q = |f(y^{(n)})| \le \|f\|,$$

which contradicts $\|x^{(n)}\|_q \to +\infty$, as $n \to +\infty$.

Thus we conclude that $(f_i)_{i=1}^{\infty} \in l^q$, and the proof is complete. ☐

Theorem D.2. *Then the map*

$$T : l^\infty \to (l^1)^* : b \mapsto f_b,$$

where

$$f_b(a) = \sum_{i=1}^{\infty} a_i b_i,$$

is an isometric isomorphism, so $(l^1)^* = l^\infty$.

Proof. In Lemma 3.4, we showed that, for $b \in l^\infty$, $T(b) = f_b$ is a well-defined element of $(l^1)^*$ and that $\|T(b)\| = \|f_b\| = \|b\|_\infty$. Furthermore, it is easy to see that the map

$$T : l^\infty \to (l^1)^* : b \mapsto f_b$$

is linear. (Check this.) Therefore T is an isometry and this also implies that T is injective. (Why?)

It remains to show that $T : l^\infty \to (l^1)^*$ is surjective. Choose $f \in (l^1)^*$ and consider $f_n = f(e_n)$, $n \ge 1$, where

$$e_n = (\underbrace{0, \dots, 0}_{n-1}, 1, 0, \dots).$$

It is clear from the construction that $T((f_n)_{n=1}^{\infty}) = f$. Thus, for T to be surjective, it will be enough to show that $(f_i)_{i=1}^{\infty} \in l^{\infty}$. We have

$$|f_n| = |f(e_n)| \leq \|f\| \|e_n\|_1 = \|f\|,$$

so f_n is a bounded sequence, giving $(f_n)_{n=1}^{\infty} \in l^{\infty}$, as required $\square$

E. THE HAHN-BANACH THEOREM

Let V be a normed vector space over $\mathbb{R}$ (we deal with complex spaces later) and let $W \subset V$ be a linear subspace. Then, given $\widetilde{f} \in V^*$, we may define a linear functional $f \in W^*$ by restriction, i.e.,

$$f(x) = \widetilde{f}(x), \qquad \text{for } x \in W.$$

The next result gives the converse.

Theoem E.1 = Theorem 3.7 (Hahn-Banach Theorem). *Let V be a normed vector space over $\mathbb{R}$ and let $W \subset V$ be a linear subspace. Suppose that $f \in W^*$ then f can be extended to a linear functional $\widetilde{f} \in V^*$ with $\|\widetilde{f}\| = \|f\|$. (Here, "extended" means that $\widetilde{f}(x) = f(x)$ for $x \in W$.)*

To prove this result we need to introduce the concept of a partially ordered set.

Definition. A set S with a relation $\succ$ is called *partially ordered* (and $\succ$ is called a *partial ordering*) if

(1) $x \succ x$;
(2) $x \succ y$ and $y \succ z \implies x \succ z$;
(3) $x \succ y$ and $y \succ x \implies x = y$.

If either $x \succ y$ then $y \succ x$ then we say that x and y are *comparable.* (In general, two elements need not be comparable.)

A subset $T \subset S$ is called a *chain* if every pair of elements in T are comparable.

Let $U \subset S$. We say that $x \in S$ is an *upper bound* for U if $x \succ u$, for all $u \in U$.

We say that $x \in S$ is a *maximal element* (for S) if $y \in S$, $y \succ x \implies y = x$.

Example 1. Let $S = \mathbb{R}$ and let $\succ$ be $\geq$ (the usual inequality). In this example every pair of elements is comparable.

Example 2. (This is a less trivial example.) Let X be a set and let $S = \{A : A \subset X\}$ be the set of all subsets of X. For $A, B \in S$, define $A \succ B \iff A \supset B$. It is easy to find pairs of sets which are not comparable (provided X has at least two elements!).

We shall use the following.

Zorn's Lemma. *Let S be a partially ordered set in which every chain has an upper bound. Then S has a maximal element.*

Zorn's Lemma is equivalent to the so-called Axiom of Choice below. This is an axiom of mathematics and is independent of all the usual axioms; thus it cannot be "proved". However, assuming it gives a consistent theory. Not assuming it also gives a consistent – but perhaps less interesting – theory. (Actually, the Hahn-Banach Theorem follows from weaker – but more complicated – assumptions than Zorn's Lemma/Axiom of Choice.)

Axiom of Choice. *Let $\mathcal{C}$ be any collection of non-empty sets. Then we can choose a set consisting of exactly one element from each of these sets. (More precisely, there is a function $f : \mathcal{C} \to \bigcup_{A \in \mathcal{C}} A$, such that $f(A) \in A$, for all $A \in \mathcal{C}$.)*

To prove the Hahn-Banach Theorem we need the following lemma. This tells us that we can extend a linear functional on a subspace if we add in one extra dimension.

Lemma E.2. *Let V be a vector space over $\mathbb{R}$ and let $W \subset V$ be a vector subspace. Let $x_0 \in V$ but $x_0 \notin W$. Then a functional $f \in W^*$ extends to the span of $\{x_0, W\}$ without increasing the norm (i.e., there exists $\widetilde{f} \in (\operatorname{span}\{x_0, W\})^*$ with $\widetilde{f}|_W = f$ and $\|\widetilde{f}\| = \|f\|$).*

Proof. If $f = 0$ they we may take $\widetilde{f} = 0$ to be the extension, so suppose that $f \neq 0$ and hence that $\|f\| > 0$.

Any vector in $\operatorname{span}\{x_0, W\}$ is of the form $\lambda x_0 + w$, where $\lambda \in \mathbb{R}$ and $w \in W$. Since the extension $\widetilde{f}$ has to be linear, it must satisfy the equation

$$\widetilde{f}(\lambda x_0 + w) = \lambda r_0 + \widetilde{f}(w) = \lambda r_0 + f(w),$$

for some $r_0 \in \mathbb{R}$. However, we need to choose r_0 in such a way that we do not increase the norm (so that $\|\widetilde{f}\| = \|f\|$). Without loss of generality, we may suppose that $\|f\| = 1$ (for otherwise we can consider $f/\|f\|$). We may also suppose that $\lambda \neq 0$, since otherwise $\lambda x_0 + w = w \in W$. Then we need to choose r_0 so that

$$|\widetilde{f}(\lambda x_0 + w)| \leq \|\lambda x_0 + w\|, \qquad \text{for all } \lambda \in \mathbb{R}, w \in W.$$

We may rewrite the inequality as

$$-\|\lambda x_0 + w\| \leq \widetilde{f}(\lambda x_0 + w) \leq \|\lambda x_0 + w\|$$

and hence as

$$-f(w) - \|\lambda x_0 + w\| \leq \lambda r_0 \leq -f(w) + \|\lambda x_0 + w\|.$$

Dividing by λ (recall $\lambda \neq 0$), this becomes

$$-f\left(\frac{w}{\lambda}\right) - \left\|x_0 + \frac{w}{\lambda}\right\| \leq r_0 \leq -f\left(\frac{w}{\lambda}\right) + \left\|x_0 + \frac{w}{\lambda}\right\|$$

(for all $\lambda \in \mathbb{R}\backslash\{0\}$, $w \in W$). However w/λ is an arbitrary element of W, so we may finally rewrite the inequality as

$$-f(w) - \|x_0 + w\| \leq r_0 \leq -f(w) + \|x_0 + w\|, \qquad \text{for all } w \in W.$$

We need to show that such an r_0 can be chosen.

Now, if $w_1, w_2 \in W$ then

$$\begin{aligned} f(w_2) - f(w_1) &\leq |f(w_2) - f(w_1)| \\ &\leq \|w_2 - w_1\| = \|(w_2 + x_0) - (w_1 + x_0)\| \\ &\leq \|w_2 + x_0\| + \|w_1 + x_0\|, \end{aligned}$$

so that

$$-f(w_1) - \|w_1 + x_0\| \leq -f(w_2) + \|w_2 + x_0\|.$$

Thus,

$$A = \sup_{w_1 \in W} (-f(w_1) - \|w_1 + x_0\|) \leq \inf_{w_2 \in W} (-f(w_2) + \|w_2 + x_0\|) = B.$$

So we only need to choose $r_0 \in [A, B]$ to ensure that $\|\widetilde{f}\| = \|f\|$. $\square$

Lemma E.2 tells us that it is possible to extend functionals one dimension at a time. However there is no guarantee that we can go from a subspace to the whole space V in this way. We get over this difficulty by employing the full force of Zorn's Lemma (but Lemma E.2 will still play a key role).

Proof of the Hahn-Banach Theorem. We shall define a relation on pairs (U, g), where $U \supset W$ is a subspace of V containing W and $g \in U^*$ such that $\|g\| = \|f\|$, by

$$(U, g) \succ (U', g') \iff \begin{cases} U \supset U' \\ g|_{U'} = g' \end{cases}.$$

It is easy to check that $\succ$ is a partial ordering on

$$S = \{(U, g) : U \supset W \text{ is a subspace of } V,\ g \in U^*,\ \|g\| = \|f\|\}.$$

Let $\{(U_\alpha, g_\alpha)\}_{\alpha \in \mathcal{A}}$ be a chain in S (where $\mathcal{A}$ is an arbitrary indexing set). Define $U = \bigcup_{\alpha \in \mathcal{A}} U_\alpha$ and $g : U \to \mathbb{R}$ by $g(u) = g_\alpha(u)$, whenever $u \in U_\alpha$. We claim that $g \in U^*$ and we shall now show that this is true.

g is well defined: Suppose that $u \in U_\alpha$ and $u \in U_\beta$. Since they are in a chain, either $(U_\alpha, g_\alpha) \succ (U_\beta, g_\beta)$ or $(U_\beta, g_\beta) \succ (U_\alpha, g_\alpha)$; without loss of generality, we shall assume the former. This means that $U_\alpha \supset U_\beta$ and that $g_\alpha|_{U_\beta} = g_\beta$, i.e., $g_\alpha(u) = g_\beta(u)$, whenever $u \in U_\beta$. Therefore, $g(u)$ is well-defined.

g is linear: Suppose that $u, v \in U$. Then $u \in U_\alpha$, $v \in U_\beta$, for some α, β. Assume $U_\alpha \supset U_\beta$ (again using the chain property) then

$$\begin{aligned} g(\lambda u + \mu v) &= g_\alpha(\lambda u + \mu v) \\ &= \lambda g_\alpha(u) + \mu g_\alpha(v) \\ &= \lambda g(u) + \mu g(v). \end{aligned}$$

$\|g\| = \|f\|$: Suppose that $u \in U$. Then $u \in \mathcal{U}_\alpha$, for some α. We have

$$\|g\| = \sup_{u \in U \setminus \{0\}} \frac{|g(u)|}{\|u\|} = \sup_{\alpha \in \mathcal{A}} \sup_{u \in U_\alpha \setminus \{0\}} \frac{|g_\alpha(u)|}{\|u\|} = \sup_{\alpha \in \mathcal{A}} \|g_\alpha\| = \|f\|.$$

Thus $(U, g) \in S$ and, by construction, $(U, g) \succ (U_\alpha, g_\alpha)$, for all $\alpha \in \mathcal{A}$, i.e., (U, g) is an upper bound for the chain $\{(U_\alpha, g_\alpha)\}_{\alpha \in \mathcal{A}}$. Therefore, by Zorn's Lemma, S has a maximal element (U_∞, g_∞). We claim that $U_\infty = V$ and thus that g_∞ is the desired extension of f to V.

Claim: $U_\infty = V$. Suppose, for a contradiction, that $U_\infty \neq V$. Choose $x_0 \notin U_\infty$. By Lemma E.2, we may extend g_∞ to $h \in (\operatorname{span}\{x_0, U_\infty\})^*$ with $\|h\| = \|f\|$. However, $(\operatorname{span}\{x_0, U_\infty\}, h) \in S$ and $(\operatorname{span}\{x_0, U_\infty\}, h) \succ (U_\infty, g_\infty)$ (but $(\operatorname{span}\{x_0, U_\infty\}, h) \neq (U_\infty, g_\infty)$). This contracdicts the maximality of (U_∞, g_∞). Therefore $U_\infty = V$ and $\widetilde{f} = g_\infty \in V^*$ has $\|\widetilde{f}\| = \|f\|$.

We now state the Hahn-Banach Theorem for complex vector spaces.

Theorem E.3 = Theorem 3.8 (Complex Hahn-Banach Theorem). *Let V be a normed vector space over $\mathbb{C}$ and let $W \subset V$ be a linear subspace. Suppose that $f \in W^*$ then f can be extended to a linear functional $\widetilde{f} \in V^*$ with $\|\widetilde{f}\| = \|f\|$.*

Before we begin the proof, note that any vector space over $\mathbb{C}$ may also be regarded as a vector space over $\mathbb{R}$. The simplest example is $\mathbb{C}$ itself, which is a 2-dimensional vector space over $\mathbb{R}$.

Proof. Consider V as a vector space over $\mathbb{R}$. Given $f \in W^*$, so $f : W \to \mathbb{C}$, we may write it in the form $f(x) = g(x) + ih(x)$, where $g = \Re f : W \to \mathbb{R}$, $h = \Im f : W \to \mathbb{R}$. Clearly $g : W \to \mathbb{R}$ and $h : W \to \mathbb{R}$ are both bounded linear functionals.

Since $f(ix) = if(x)$, $x \in W$, we have that

$$g(ix) + ih(ix) = -h(x) + ig(x)$$

so that $\Im f = -g(ix)$, i.e.,

$$f(x) = g(x) - ig(ix), \qquad \text{for all } x \in W.$$

Now consider W and V as a vector space over $\mathbb{R}$. We can use the real Hahn-Banach Theorem to extend the bounded linear functional $g : W \to \mathbb{R}$ to a linear functional $\widetilde{g} : V \to \mathbb{R}$. Since we are considering V as a real vector space, this means that $\widetilde{g}(ax) = a\widetilde{g}(x)$ for $a \in \mathbb{R}$ but *not necessarily* for $a \in \mathbb{C}$.

We are going to use $\widetilde{g}$ to define $\widetilde{f} : V \to \mathbb{C}$ by

$$\widetilde{f}(x) = \widetilde{g}(x) - i\widetilde{g}(ix).$$

One easily checks that $\widetilde{f}$ is a linear functional on V, considered as a complex space: for $a, b \in \mathbb{R}$,

$$\begin{aligned}
\widetilde{f}((a+ib)x) &= \widetilde{g}(ax + ibx) - i\widetilde{g}(iax - bx) \\
&= a\widetilde{g}(x) + b\widetilde{g}(ix) - ia\widetilde{g}(ix) + bi\widetilde{g}(x) \\
&= a\widetilde{f}(x) + ib\widetilde{f}(x) \\
&= (a+ib)\widetilde{f}(x).
\end{aligned}$$

Claim: $\|\widetilde{f}\| = \|f\|$.

(a) If $x \in V$ with $f(x) \in \mathbb{R}$ then $\widetilde{f}(x) = g_0(x)$. Thus $|\widetilde{f}(x)| = |g_0(x)| \leq \|g_0\|\|x\| \leq \|f\|\|x\|$.

(b) If $x \in V$ with $\widetilde{f}(x) = re^{i\theta}$ (with $r > 0$, $0 \leq \theta < 2\pi$) then

$$|\widetilde{f}(x)| = r = e^{-i\theta}\widetilde{f}(x) = \widetilde{f}(e^{-i\theta}x).$$

We may apply (a) to $e^{-i\theta}x$ to get

$$|\widetilde{f}(x)| = r \leq \|f\|\|e^{-i\theta}x\| = \|f\|\|x\|.$$

This completes the proof. □

The following are immediate consequences of the Hahn-Banach Theorem.

Proposition E.4. *Let V be a normed vector space over $\mathbb{R}$ (with norm $\|\cdot\|$). Given $x_0 \in V$, there exists $f \in V^*$ such that $f(x_0) = \|x_0\|$ and $\|f\| = 1$.*

Proof. Let $W = \{\lambda x_0 : \lambda \in \mathbb{R}\}$ and define $f_0 \in W^*$ by $f_0(\lambda x_0) = \lambda \|x_0\|$. It is immediate that $\|f_0\| = 1$. Apply the Hahn-Banach Theorem to extend f_0 to $f \in V^*$ with $\|f\| = 1$. $\square$

Proposition E.5. *Let V be a normed vector space over $\mathbb{R}$. Given $x, y \in V$, with $x \neq y$, there exists $f \in V^*$ such that $f(x) \neq f(y)$.*

Proof. Let $x_0 = x - y$. Then Proposition E.4 allows us to find $f \in V^*$ such that

$$f(x) - f(y) = f(x - y) = ||x - y|| \neq 0,$$

as required. $\square$

Of course, there are versions of these results for normed vector spaces over $\mathbb{C}$.

The next result shows that the "norm" can be replaced by a more general function $p : V \to \mathbb{R}$ (called a semi-norm).

Theorem E.6. *Let V be a vector space over $\mathbb{R}$ and let $W \subset V$ be a linear subspace. Suppose that $p : V \to \mathbb{R}$ satisfies*

$$p(x + y) \leq p(x) + p(y) \quad \textit{and} \quad p(\lambda x) = |\lambda| p(x),$$

for all $x, y \in V$, $\lambda \in \mathbb{R}$. Then a linear functional $f : W \to \mathbb{R}$ satisfying $f(x) \leq p(x)$, for all $x \in W$, can be extended to a linear functional $\widetilde{f} : V \to \mathbb{R}$ satisfying $f(x) \leq p(x)$, for all $x \in V$.

Proof. Omitted. It is much the same as for the standard Hahn-Banach Theorem. $\square$

F. SEPARATION THEOREMS

Definition. Let V be a vector space. A set $A \subset V$ is said to be *convex* if, for any $x, y \in A$ and $0 \leq \alpha \leq 1$, $\alpha x + (1-\alpha)y \in A$.

The next theorem says that, given two disjoint, convex and non-empty sets in a normed vector space, one can find a bounded linear functional which separates them in the sense that the functional maps one set inside $(-\infty, \gamma)$ and the other inside $[\gamma, +\infty)$, for some $\gamma \in \mathbb{R}$.

Theorem F.1 (Ham Sandwich Theorem). *Let V be a normed vector space (over $\mathbb{R}$) and let $A, B \subset V$ be disjoint, convex and non-empty. If A is open then there exists a linear functional $f \in V^*$ and a constant $\gamma \in \mathbb{R}$ such that*

$$f(x) < \gamma \leq f(y), \quad \textit{for all } x \in A, y \in B.$$

Proof. Fix $a_0 \in A$ and $b_0 \in B$. Let $x_0 = b_0 - a_0$ and

$$C = A - B + x_0 := \{a - b + x_0 \,:\, a \in A, b \in B\}.$$

Then C is an open neighbourhood of $0 \in V$ and C is convex.

We define $p : V \to \mathbb{R}$ by $p(x) = \inf\{t > 0 \,:\, t^{-1}x \in C\}$. In particular, since $A \cap B = \varnothing$, $x_0 \notin C$, so that $p(x_0) \geq 1$. (Check that p satisfies the hypotheses of the Theorem E.6.)

Let W be the one-dimensional linear subspace spaned by x_0 and define $f : W \to \mathbb{R}$ by $f(\lambda x_0) = \lambda$. If $\lambda \geq 0$ then $f(\lambda x_0) = \lambda \leq \lambda p(x_0) = p(\lambda x_0)$ and if $\lambda < 0$ then $f(\lambda x_0) = \lambda < 0 \leq p(\lambda x_0)$, so $f \leq p$ on W.

By Theorem E.6, the linear functional f extends to a linear functional $\widetilde{f} : V \to \mathbb{R}$, such that $\widetilde{f}(x) \leq p(x)$, for all $x \in V$.

We need to show that $\widetilde{f}$ is bounded (i.e. that $\widetilde{f} \in V^*$). Now, if $x \in C$ then $\widetilde{f}(x) \leq p(x) \leq 1$. Similarly, if $x \in -C$ (so $-x \in C$), $\widetilde{f}(x) = -\widetilde{f}(-x) \geq -1$. Thus, for $x \in C \cap (-C)$, $|\widetilde{f}(x)| \leq 1$. This shows that $\widetilde{f}$ is bounded on a neighbourhood of $0 \in V$ and we may deduce that $\widetilde{f} \in V^*$.

If $a \in A$, $b \in B$ then, since $\widetilde{f}(x_0) = 1$,

$$\begin{aligned} \widetilde{f}(a) - \widetilde{f}(b) + 1 &= \widetilde{f}(a - b + x_0) \\ &\leq p(a - b + x_0) < 1, \end{aligned}$$

since $a - b + x_0 \in C$. (We get the strict inequality < 1 since C is open.) In other words, for all $a \in A, b \in B$,

$$\widetilde{f}(a) < \widetilde{f}(b).$$

Let

$$\gamma = \inf\{\widetilde{f}(b) \,:\, b \in B\},$$

so that, for all $b \in B$,

$$\gamma \leq \widetilde{f}(b).$$

We claim that we also have $\widetilde{f}(a) < \gamma$, for all $a \in A$. Clearly, $\widetilde{f}(a) \leq \gamma$, for all $a \in A$. Suppose, for a contradiction, that there exists $a_1 \in A$ such that $\widetilde{f}(a_1) = \gamma$. Choose $v \in V$ so that $\widetilde{f}(v) > 0$. Since A is open, if $\epsilon > 0$ is sufficiently small then $a_1 + \epsilon v \in A$. However, we have $\widetilde{f}(a_1 + \epsilon v) = \gamma + \epsilon \widetilde{f}(v) > \gamma$, a contradiction.

Therefore, for all $a \in A, b \in B$,

$$\widetilde{f}(a) < \gamma \leq \widetilde{f}(b),$$

as required. $\square$