Chapter 12

The Hahn-Banach Theorem

In this chapter V is a real or complex vector space. The scalars will be taken to be real until the very last result, the comlex-version of the Hahn-Banach theorem.

12.1 The geometric setting

If A is a subset of V then the *translate* of A by a vector $x \in V$ is the set

$$x + A = \{x + a : a \in A\}$$

If A and B are subsets of V and t any real number we use the notation

$$tA = \{ta : a \in A\} \qquad \text{and} \qquad A + B = \{a + b : \ a \in A, b \in B\}$$

If $x, y \in V$ then the segment xy is the set of all points on the line running from x to y:

$$xy = \{tx + (1 - t)y : 0 \leq t \leq 1\}$$

A subset C of V is convex if for any two point $P, Q \in C$ the segment PQ is contained in C. Equivalently, C is convex if

$$\lambda C + (1 - \lambda)C \subset C$$

for every $\lambda \in [0, 1]$.

It is clear that the translate of any convex set is convex, and indeed if C is a convex set then so is $a + tC$ for any $a \in V$ and $t \in \mathbf{R}$.

A subspace W of V has *codimension* 1 if there is a vector $x \in V \setminus W$ such that $W + \mathbf{R}x = V$. This is equivalent to saying that the quotient space V/W has dimension 1.

A hyperplane is a set of the form $W + x$ where W is any codimension one subspace and x is any vector.

Let W be a codimension 1 subspace of V, and v any vector outside W. Then V can be expressed as the union of W with two open half-spaces:

$$V = W \cup (W + \{tv : t > 0\}) \cup (W - \{tv : t > 0\})$$

If x is any vector in V then the hyperplane $W + x$ specifies two *closed half-spaces*:

$$W + x + \{tv : t \geq 0\} \text{ and } W + x - \{tv : t \geq 0\}$$

whose intersection is the hyperplane $W + x$ and whose union is all of V. We shall refer to these closed half-spaces as the two *sides* of the hyperplane.

We will prove the following geometrically intuitive fact:

- If C is a convex subset of V and $p \in V$ a point outside C then there is a hyperplane H such that C is a subset of one side of H and p lies on the other side.

Though it is possible to prove this by "purely geometric" reasoning, it will be both more convenient and more useful for our purposes to use an algebraic approach.

It will be convenient to use the infinities ∞ and $-\infty$. We require that $-\infty < \infty$, and $-\infty < x < \infty$ for all real numbers x. The following arithmetic operations with ∞ will be defined:

$$t + \infty = \infty + t = \infty, \qquad k\infty = \infty k = \infty, \qquad 0\infty = \infty 0 = 0$$

for all $k > 0$ and all $t \in \mathbf{R} \cup \{\infty\}$.

12.2 The algebraic formulation

Let C be a non-empty convex subset of V. If C is non-empty then we can translate C appropriately to ensure that $0 \in C$. For this section, we assume that the origin 0 belongs to C, i.e. $0 \in C$. The "size" of a vector $v \in V$

relative to C is the "smallest" non-negative number $t \geq 0$ such that v lies in the tC; more precisely, define

$$p_C(v) = \inf\{t \geq 0 : v \in tC\}$$

where the infimum of the empty set is taken to be ∞. The function

$$p_C : V \to [0, \infty]$$

is the Minkowski functional for the set C.

Note that

$$p_C(0) = 0$$

but it might be the case that $p_C(v)$ is 0 for some non-zero v.

Proposition 1 *Let C be a convex set containing 0. Then*

(i) *for any $v \in V$, the set $\{t \geq 0 : v \in tC\}$ is an interval, either equal to $[p_C(v), \infty)$ or $(p_C(v), \infty)$*

(ii) *$p_C(tv) = tp_C(v)$ for every $v \in V$ and $t \geq 0$*

(iii) *if $x, y \in V$ then*

$$p_C(x + y) \leq p_C(x) + p_C(y)$$

This is the "triangle inequality."

Proof. (i) Suppose s is a real number $> p_C(v)$. We want to show that v lies in sC. The definition of $p_C(v)$ implies that there is some $t \in [0, s)$ such that $v \in tC$ and so $v = tx$ for some $x \in C$. Hence

$$v = s\left[\frac{t}{s}x + \left(1 - \frac{t}{s}\right)0\right] \in sC$$

by convexity of C.

(ii) For any real $s > 0$ we have

$$\{t \geq 0 : sv \in tC\} = \{sr : r \geq 0, \text{and}\, v \in rC\} = s\{r \geq 0 : v \in rC\}$$

which implies $p_C(sv) = sp_C(v)$. If $s = 0$ then this is clear (note that $0\infty = 0$).

(iii) If either $p_C(x)$ or $p_C(y)$ is ∞ then $p_C(x+y)$ is automatically $\leq p_C(x) + p_C(y)$. So suppose $p_C(x) < \infty$ and $p_C(y) < \infty$. Let t, s be real numbers with $t > p_C(x)$ and $s > p_C(y)$. Then $x \in tC$ and $y \in sC$ and so

$$x + y \in tC + sC = (t+s)\left(\frac{t}{t+s}C + \frac{s}{t+s}C\right) \subset (t+s)C,$$

the last subset relation following from convexity of C. So $p_C(x+y) \leq t+s$. Taking inf over t and then over s gives $p_C(x+y) \leq p_C(x) + p_C(y)$. $\boxed{\text{QED}}$

The definition of the Minkowski functional p_C implies that $p_C(x) \leq 1$ for every $x \in C$, i.e.

$$C \subset \{v \in V : p_C(v) \leq 1\}$$

but C may not actually be equal to the "closed ball" $\{v \in V : p_C(v) \leq 1\}$.

There is a converse construction of a convex set from a functional p:

Proposition 2 *Suppose $p : V \to [0, \infty]$ is a map satisfying the following conditions:*

(a) $p(0) = 0$

(b) $p(tx) = tp(x)$ *for all* $x \in V$ *and real* $t \geq 0$

(c) $p(x+y) \leq p(x) + p(y)$ *for every* $x, y \in V$

Let

$$C = \{v \in V : p(v) \leq 1\}$$

Then C is a convex set containing 0 and the Minkowski functional of C is p, i.e. $p_C = p$.

Proof. It is readily checked that C is a convex set and contains 0.

Note that for any real $t > 0$ we have

$$x \in tC \qquad \text{if and only if} \qquad p(x) \leq t$$

(However, the set $\{x \in V : p(x) = 0\}$ may contain non-zero elements). So

$$\{t \in \mathbf{R} : t > 0, x \in tC\} = \{t : p(x) \leq t\} = [p(x), \infty)$$

Taking the infimum gives

$$p_C(x) = p(x) \qquad \boxed{\text{QED}}$$

Next we formulate an algebraic equivalent of a hyperplane.

Let W be a codimension 1 subspace of V. Then there is a vector $n \notin W$ such that

$$V = W + \mathbf{R}n$$

This sum is a direct sum, for if $x \in W \cap \mathbf{R}n$ then $x = tn \in W$ for some real number t, and so t must be zero for otherwise $n = t^{-1}x$ would be in W. So the map

$$W \oplus \mathbf{R}n \to V : (w, tn) \mapsto w + tn$$

is a linear isomorphism. Thus the map

$$L : V \to \mathbf{R} : w + tn \mapsto t$$

is linear and its kernel is exactly the subspace W.

Conversely, if $f : V \to \mathbf{R}$ is a non-zero linear map then the kernel $\ker f$ is a codimension 1 subspace of V, for $\ker f + \mathbf{R}n = V$, where n is any vector for which $f(n) = 1$. If f and g are two non-zero linear maps $V \to \mathbf{R}$, then $\ker f = \ker g$ if and only if f is a non-zero multiple of g.

A hyperplane in V is specified by the level set of a linear functional, i.e. if H is a hyperplane then there is a non-zero linear functional $L : V \to \mathbf{R}$ and a real number $t \in \mathbf{R}$ such that $L^{-1}(t) = H$.

Thus we have found algebraic equivalents for the geometric notions of convex sets and hyperplanes.

12.3 The Hahn-Banach Theorem

Theorem 1 . *Let V be a real vector space. Suppose $p : V \to [0, \infty]$ is a mapping satisfying the following conditions:*

(a) $p(0) = 0$

(b) $p(tx) = tp(x)$ *for all* $x \in V$ *and real* $t \geq 0$

(c) $p(x + y) \leq p(x) + p(y)$ *for every* $x, y \in V$

Assume, furthermore, that for each $x \in V$, either both $p(x)$ and $p(-x)$ are ∞ or that both are finite.

Let $a \in V$ and α a real number with $0 \leq \alpha \leq p(a)$. Then there is a linear functional

$$f : V \to \mathbf{R}$$

such that $f(a) = \alpha$ *and*

$$f(x) \leq p(x)$$

for all $x \in V$.

More generally, if W *is a subspace of* V *and* $g : W \to \mathbf{R}$ *a linear mapping satisfying* $g(x) \leq p(x)$ *for all* $x \in W$ *then there is a linear mapping* $f : V \to \mathbf{R}$ *such that* $f \leq p$ *and* $f|W = g$.

Proof Let us first show that the first statement follows from the second. Let $W = \mathbf{R}a$ and on W define $g : W \to \mathbf{R} : ta \mapsto t\alpha$. This is a linear map and satisfies $g(ta) = t\alpha \leq tp(a) = p(ta)$ when $t \geq 0$ and $g(ta) = t\alpha \leq 0 \leq p(ta)$ when $t < 0$.

To prove the existence of the extension f we can use a Zorn's lemma argument applied to all extensions of g which are bounded by p. If V is finite dimensional then of course this reduces to an induction argument.

Thus it will suffice to show how to extend g to a subspace $W + \mathbf{R}v$, where v is any vector outside W. Note that each element of $W + \mathbf{R}v$ can be expressed uniquely in the form $w + tv$ with $w \in W$ and $t \in \mathbf{R}$.

If p is finite-valued then the proof proceeds smoothly, but taking into account points where p is ∞ makes the argument complicated.

The linear map $h : W + \mathbf{R}v \to \mathbf{R} : w + tv \mapsto g(w)$ restricts to g on W. If it so happens that $p(x) = \infty$ for all $x \in W + \mathbf{R}v$ outside W then $h \leq p$ holds automatically on $W + \mathbf{R}v$. So we may and shall assume that the particular vector v outside W is chosen such that $p(v) < \infty$.

Define $h : W + \mathbf{R}v \to \mathbf{R}$ by

$$h(w + tv) = g(w) + th(v)$$

for all $w \in W$ and $t \in \mathbf{R}$, where $h(v)$ is a real number chosen to satisfy

$$h(w - sv) = g(w) - sh(v) \leq p(w - sv) \qquad \text{for all } w \in W \text{ and real } s \geq 0 \tag{12.1}$$

and

$$h(w + tv) = g(w) + th(v) \leq p(w + tv) \qquad \text{for all } w \in W \text{ and real } t \geq 0 \tag{12.2}$$

That such a choice of $h(v)$ is possible will be shown below. Note that $h(w) = g(w)$ for all $w \in W$. The preceding inequalities together imply

$$h(x) \leq p(x) \qquad \text{holds for all } x \in W + \mathbf{R}v.$$

Thus h gives the desired extension of g.

To complete the proof, we need to show that a real number $h(v)$ can be chosen which is $\geq \frac{g(w)-p(w-sv)}{s}$ and $\leq \frac{p(w+tv)-g(w)}{t}$ for every $w \in W$, and all real numbers $s, t > 0$. This means that we have to show that

$$\sup_{s>0, w\in W} \frac{g(w) - p(w - sv)}{s} \leq \inf_{t>0, w\in W} \frac{p(w + tv) - g(w)}{t} \tag{12.3}$$

and that this inequality isn't reading $\infty \leq \infty$ or $-\infty \leq -\infty$.

The inequality (12.3) is equivalent to

$$\frac{g(w) - p(w - sv)}{s} \leq \frac{p(w + tv) - g(w)}{t} \tag{12.4}$$

holding for all $w \in W$ and all $s, t > 0$. If either $p(w - sv)$ or $p(w + tv)$ is ∞ then (12.4) holds automatically. So suppose both $p(w - sv)$ and $p(w + tv)$ are finite. Then, after rearranging terms, (12.4) is equivalent to

$$g((t + s)w) \leq p(sw + stv) + p(tw - tsv)$$

and this is indeed true since

$$g((t + s)w) \leq p((t + s)w) \leq p(sw + stv) + p(tw - tsv)$$

The infinites don't occur: for (i) the right side of (12.3) is seen, upon taking $w = 0$, to be bounded above by $p(v)$ which has been assumed to be finite; and (ii) the left side of (12.3) is bounded below by $-p(-v)$.

So we may choose a real number $h(v)$ satisfying:

$$\sup_{s>0, w\in W} \frac{g(w) - p(w - sv)}{s} \leq h(v) \leq \inf_{t>0, w\in W} \frac{p(w + tv) - g(w)}{t} \tag{12.5}$$

This proves the existence of a linear function

$$h : W + \mathbf{R}v \to \mathbf{R}$$

which restricts to g on W and which satisfies $h \leq p$ on $W + \mathbf{R}v$.

The Zorn's lemma procedure is now routine. Let X be the set of all pairs

$$(S, h)$$

where S is a subspace of V with $W \subset S$ and $h : S \to \mathbf{R}$ is a linear mapping satisfying $h|W = g$ and $h \leq p|S$. Define the relation $<$ on X by

$$(S_1, h_1) < (S_2, h_2)$$

to mean $S_1 \subset S_2$ and $h_2|S_1$. This is a partial ordering with the property that any totally ordered subset has a maximal element. Zorn's lemma says then that X has a maximal element (S_*, h_*). If S_* were a proper subspace of V, i.e. $S_* \neq V$, then the argument given before produces an extension of h to a larger subspace $S_* + \mathbf{R}v$, contradicting the maximality of (S_*, h_*). $\boxed{\text{QED}}$

There is a complex-scalars version of the Hahn-Banach theorem. Before looking at this we make a quick observation:

Lemma 1 *Let V be a complex vector space and suppose $G : V \to \mathbf{R}$ is a real–linear mapping. Then there is a unique complex–linear mapping $f : V \to \mathbf{C}$ such that $G = \mathrm{Re}(f)$; explicitly,*

$$f(x) = G(x) - iG(ix)$$

for every $x \in V$.

Proof. Suppose $h : V \to \mathbf{C}$ is a complex–linear mapping, and denote by u its real part and by v its imaginary part:

$$h(x) = u(x) + iv(x)$$

for all $x \in V$. Using complex-linearity of h we have

$$u(ix) + iv(ix) = h(ix) = -v(x) + iu(x)$$

which yields v in terms of u:

$$v(x) = -u(ix)$$

Thus a complex–linear mapping h is uniquely determined by its real part. We can reconstruct the full map from the real part:

$$h(x) = u(x) + iv(x) = u(x) - iu(ix)$$

Using this as guide, we define $f : V \to \mathbf{C}$ by

$$f(x) = G(x) - iG(ix)$$

for all $x \in V$. Then f is clearly real–linear and we check readily that

$$f(ix) = G(ix) - iG(-x) = i\left[G(x) - iG(ix)\right] = if(x)$$

for every $x \in V$. So f is complex–linear. $\boxed{\text{QED}}$

Theorem 2 *Let V be a complex vector space. Suppose $p : V \to [0, \infty]$ is a mapping satisfying the following conditions:*

(a) $p(0) = 0$

(b) $p(\alpha x) = |\alpha| p(x)$ *for all* $x \in V$ *and all* $\alpha \in \mathbf{C}$

(c) $p(x + y) \leq p(x) + p(y)$ *for every* $x, y \in V$

Let $a \in V$ and s a complex number with $0 \leq |s| \leq p(a)$. Then there is a complex–linear functional

$$f : V \to \mathbf{C}$$

such that $f(a) = s$ and

$$|f(x)| \leq p(x)$$

for all $x \in V$.

More generally, if W is a subspace of V and $g : W \to \mathbf{C}$ a complex–linear mapping satisfying $g(x) \leq p(x)$ for all $x \in W$ then there is a complex–linear mapping $f : V \to \mathbf{C}$ such that $|f(x)| \leq p(x)$ for all $x \in V$ and $f|W = g$.

Proof. Viewing V as a real vector space, the previous theorem gives us a real–linear mapping $G : V \to \mathbf{R}$ satisfying $G|W = \mathrm{Re}(g)$ and $G \leq p$. Let $f : V \to \mathbf{C}$ be the complex–linear map whose real part is G. Then $f|W = g$. Let $x \in V$. Then there is a complex number λ of modulus 1 for which

$$\lambda f(x) = |f(x)|$$

So

$$|f(x)| = \lambda f(x) = f(\lambda x)$$

which implies that $f(\lambda x)$ is real and hence equal to $G(\lambda x)$. So

$$|f(x)| = G(\lambda x) \leq p(\lambda x) = |\lambda| p(x) = p(x)$$

Thus f satisfies all the desired properties. $\boxed{\text{QED}}$