On the Hahn-Banach theorem

S. Kesavan
Department of Mathematics,
Indian Institute of Technology, Madras,
Chennai - 600 036
email: **kesh@iitm.ac.in**

Abstract

The Hahn-Banach theorem is one of the major theorems proved in any first course on Functional Analysis. It has plenty of applications, not only within the subject itself, but also in other areas of mathematics like optimization, partial differential equations and so on. This article will give a brief overview of the Hahn-Banach theorem, its ramifications and indicate some applications.

1 Introduction

One of the major theorems that we encounter in a first course on Functional Analysis is the Hahn-Banach theorem. Together with the Banach-Steinhaus theorem, the open mapping theorem and the closed graph theorem, we have a very powerful set of theorems with a wide range of applications. The latter three theorems are all dependent on the completeness of the spaces involved, whereas the Hahn-Banach theorem does not demand this. It comes in two versions, analytic and geometric. The analytic version concerns the extension of continuous linear functionals from a subspace to the whole space with prescribed properties while the geometric version deals with the separation of disjoint convex sets using hyperplanes. Both these versions have applications outside Functional Analysis, for example in the theory of optimization and in the theory of partial differential equations, to name a few.

In this article we will give an overview of the Hahn-Banach theorem and some of its applications. We will not give detailed proofs of the main theorems since they can be found in any book on Functional Analysis.

2 The Hahn-Banach theorems

The analytic and geometric versions of the Hahn-Banach theorem follow from a general theorem on the extension of linear functionals on a real vector space.

Theorem 2.1 (**Hahn-Banach Theorem**) *Let V be a vector space over $\mathbb{R}$. Let $p : V \to \mathbb{R}$ be a mapping such that*

$$\left.\begin{array}{rcl} p(\alpha x) & = & \alpha p(x) \\ p(x+y) & \leq & p(x)+p(y) \end{array}\right\} \tag{2.1}$$

for all x and $y \in V$ and for all $\alpha > 0$ in $\mathbb{R}$. Let W be a subspace of V and let $g : W \to \mathbb{R}$ be a linear map such that

$$g(x) \leq p(x)$$

for all $x \in W$. Then, there exists a linear extension $f : V \to \mathbb{R}$ of g (i.e. $f(x) = g(x)$ for all $x \in W$) which is such that

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

for all $x \in V$. ■

The above theorem is proved using Zorn's lemma. Let $\mathcal{P}$ denote the collection of all pairs (Y, h), where Y is a subspace of V containing W and $h : Y \to \mathbb{R}$ a linear map which is an extension of g and which is also such that

$$h(x) \;\leq\; p(x)$$

for all $x \in Y$. Clearly $\mathcal{P}$ is non-empty, since $(W, g) \in \mathcal{P}$. Consider the partial order defined on $\mathcal{P}$ by

$$(Y, h) \;\preceq\; (\widetilde{Y}, \widetilde{h})$$

if $Y \subset \widetilde{Y}$ and $\widetilde{h}$ is a linear extension of h.

We show that any chain in $\mathcal{P}$ has an upper bound and so, by Zorn's lemma, there exists a maximal element (Z, f) in $\mathcal{P}$. We then show that $Z = V$ which will complete the proof.

There are two principal examples of the mapping p mentioned in the above theorem.

Given a normed linear space X, we will denote its dual, the space of continuous linear functionals on X, by the symbol X^*.

Example 2.1 Let V be a normed linear space and let W be a subspace of V. Let $g \in W^*$ be a continuous linear functional. Then $p(x) = \|g\|_{W^*}\|x\|$ satisfies (2.1). Applying the theorem above to this case, we get the following result.

Theorem 2.2 *(***Hahn-Banach Theorem***) Let V be a normed linear space over $\mathbb{R}$. Let W be a subspace of V and let $g : W \to \mathbb{R}$ be a continuous linear functional on W. Then there exists a continuous linear extension $f : V \to \mathbb{R}$ of g such that*

$$\|f\|_{V^*} \;=\; \|g\|_{W^*}. \;\blacksquare$$

This result is also true for normed linear spaces over $\mathbb{C}$. If X is a normed linear space over the field of complex numbers, then let us write it in terms of its real and imaginary parts:

$$f \;=\; g + ih,$$

where i stands for a square root of -1. Then g and h are linear functionals, as long as we restrict ourselves to scalar multiplication by reals. Now, since

$f(ix) = if(x)$ for any $x \in X$, it follows easily that $h(x) = -g(ix)$. Thus, for any $x \in X$, we have,

$$f(x) = g(x) - ig(ix).$$

In other words, the real part of a linear functional over $\mathbb{C}$ is enough to describe the functional fully. If we restrict ourselves to scalar multiplication by reals only and consider X as a real normed linear space, then g is a continuous linear functional and we can also show that $\|g\| = \|f\|_{X^*}$.

Thus, given a normed linear space V over $\mathbb{C}$, a subspace W and a continuous linear functional g on W, we can write $g(x) = h(x) - ih(ix)$ for any $x \in X$, where h is a real valued linear functional over $\mathbb{R}$. Then, by the previous theorem, we can find an extension $\widetilde{h}$ of h and define

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

Then it is easy to check that f is a norm preserving extension of g to all of V. Thus we have the following theorem.

Theorem 2.3 *(***Hahn-Banach Theorem***) Let V be a normed linear space over $\mathbb{C}$. Let W be a subspace of V and let $g : W \to \mathbb{C}$ be a continuous linear functional on W. Then there exists a continuous linear extension $f : V \to \mathbb{C}$ of g such that*

$$\|f\|_{V^*} = \|g\|_{W^*}. \blacksquare$$

Example 2.2 The next example of a mapping p satisfying (2.1) is called the **Minkowski functional** of an open convex subset of a real normed linear space which containes the origin.

Proposition 2.1 *Let C be an open and convex set in a real normed linear space V such that $\mathbf{0} \in C$. For $x \in V$, set*

$$p(x) = \inf\{\alpha > 0 \mid \alpha^{-1}x \in C\}.$$

Then, there exists $M > 0$ such that

$$0 \le p(x) \le M\|x\| \tag{2.2}$$

for all $x \in V$. We also have

$$C = \{x \in V \mid p(x) < 1\}. \tag{2.3}$$

Further, p satisfies (2.1).

Proof: Since $\mathbf{0} \in C$ and C is open, there exists an open ball $B(\mathbf{0}; 2r)$, centered at $\mathbf{0}$ and of radius $2r$, contained in C. Now, if $x \in V$, we have $rx/\|x\| \in C$ and so, by definition, $p(x) \leq \frac{1}{r}\|x\|$ which proves (2.2).

Let $x \in C$. Since C is open, and since $\mathbf{0} \in C$, there exists $\varepsilon > 0$ such that $(1+\varepsilon)x \in C$. Thus, $p(x) \leq (1+\varepsilon)^{-1} < 1$. Conversely, let $x \in V$ such that $p(x) < 1$. Then, there exists $0 < t < 1$ such that $\frac{1}{t}x \in C$. Then, as C is convex, we also have $t\frac{1}{t}x + (1-t)\mathbf{0} \in C$, *i.e.* $x \in C$. This proves (2.3).

If $\alpha > 0$, it is easy to see that $p(\alpha x) = \alpha p(x)$. This is the first relation in (2.1). Now, let x and $y \in V$. Let $\varepsilon > 0$. Then

$$\frac{1}{p(x)+\varepsilon}x \in C \quad \text{and} \quad \frac{1}{p(y)+\varepsilon}y \in C.$$

Set

$$t \;=\; \frac{p(x)+\varepsilon}{p(x)+p(y)+2\varepsilon}$$

so that $0 < t < 1$. Then, as C is convex,

$$t\frac{1}{p(x)+\varepsilon}x + (1-t)\frac{1}{p(y)+\varepsilon}y \;=\; \frac{1}{p(x)+p(y)+2\varepsilon}(x+y) \in C$$

which implies that

$$p(x+y) \;\leq\; p(x)+p(y)+2\varepsilon$$

from which the second relation in (2.1) follows since ε was chosen arbitrarily.
■

Let C be a non-empty open convex set in a real normed linear space V and assume that $0 \in C$ and $x_0 \notin C$. Let W be the one-dimensional space spanned by x_0. Define $g : W \to \mathbb{R}$ by

$$g(tx_0) \;=\; t.$$

By definition of the Minkowski functional, since $\frac{1}{t}tx_0 = x_0 \notin C$, we have that

$$g(tx_0) \;=\; t \;\leq p(tx_0)$$

for $t > 0$. Since the Minkowski functional is non-negative, this inequality holds trivially for $t \leq 0$ as well. Thus, by the Hahn-Banach theorem, there exists a linear extension f of g to the whole of V such that, for all $x \in V$,

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

(cf. (2.2)) which yields $|f(x)| \leq M\|x\|$, and so f is continuous as well. Now, if $x \in C$,

$$f(x) \ \leq \ p(x) \ < \ 1 \ = \ g(x_0) \ = \ f(x_0)$$

by (2.3).

It is easy to see that the condition $0 \in C$ can be relaxed, once we have the above result, to prove the existence of $f \in V^*$ such that $f(x) < f(x_0)$ for every $x \in C$. Using this one can prove the following first geometric version of the Hahn-Banach theorem.

Theorem 2.4 *(***Hahn-Banach Theorem***) Let A and B be two non-empty and disjoint convex subsets of a real normed linear space V. Assume that A is open. Then, there exists a closed hyperplane which separates A and B, i.e. there exists $f \in V^*$ and $\alpha \in \mathbb{R}$ such that*

$$f(x) \ \leq \ \alpha \ \leq \ f(y)$$

for all $x \in A$ and $y \in B$.

In other words, given two disjoint non-empty convex sets, one of them being open, we can separate the two by means of a (closed) hyperplane.

Consider, for example, the plane $\mathbb{R}^2$. Let

$$A = \{(x, y) \mid x > 0, y > 0, xy > 1\}$$

and let

$$B = \{(x, y) \mid y \leq 0\}.$$

These are two disjoint, oonvex sets and while A is open, we have that B is closed. They can be separated by the line $\{(x, y) \mid y = 0\}$.

Notice we cannot separate them strictly, *i.e.* there is no line in the plane separating the two sets and at a positive distance from both. On the other hand, if we set $A = \{(x, y) \mid xy \geq 1, 1 \leq x \leq 2\}$, then the line $\{(x, y) \mid y = \frac{1}{2}\}$ separates the two sets srictly.

The following theorem can be proved starting from the preceding one.

Theorem 2.5 (**Hahn-Banach Theorem**) *Let A and B be non-empty and disjoint convex sets in a real normed linear space V. Assume that A is closed and that B is compact. Then A and B can be separated* strictly *by a closed hyperplane, i.e. there exists $f \in V^*$, $\alpha \in \mathbb{R}$ and $\varepsilon > 0$ such that*

$$f(x) \ \leq \ \alpha - \varepsilon \ \ \textit{and} \ \ f(y) \ \geq \ \alpha + \varepsilon$$

for all $x \in A$ and $y \in B$ ■.

In case of complex normed linear spaces, the above geometric versions are true with f being replaced by $\mathrm{Re}(f)$, the real part of f.

The geometric versions are also true in more general settings. A topological vector space is a vector space equipped with a Hausdorff topology such that vector addition and scalar multiplication are continuous operations. Such a space is called *locally convex* if each point admits a neighbourhood system made up of convex sets. The geometric versions of the Hahn-Banach theorems mentioned above are true for locally convex topological vector spaces (cf. Rudin [5]).

3 Richness of the dual space

One of the main consequences of the Hahn-banach theorem(s) is the fact that the dual of a normed linear space is well endowed with functionals and hence merits careful study.

Proposition 3.1 *Let V be a normed linear space and $x_0 \in V$ a non-zero vector. Then, there exists $f \in V^*$ such that $\|f\| = 1$ and $f(x_0) = \|x_0\|$.*

Proof: Let W be the one-dimensional space spanned by x_0. Define $g(\alpha x_0) = \alpha\|x_0\|$. Then $\|g\|_{W^*} = 1$. Hence, there exists $f \in V^*$ such that $\|f\|_{V^*} = 1$ and which extends g. Hence $f(x_0) = g(x_0) = \|x_0\|$. ■

If V is a normed linear space and if x and y are distinct points in V, then, clearly, there exists $f \in V^*$ such that $f(x) \neq f(y)$ (consider $x_0 = x - y \neq \mathbf{0}$). We say that V^* **separates points** of V.

Proposition 3.2 *Let W be a subspace of a normed linear space V. Assume that $\overline{W} \neq V$. Then, there exists $f \in V^*$ such that $f \not\equiv 0$ and such that $f(x) = 0$ for all $x \in W$.*

Proof: Let $x_0 \in V \backslash \overline{W}$. Let $A = \overline{W}$ and $B = \{x_0\}$. Then A is closed, B is compact and they are non-empty and disjoint convex sets. Thus, there exists $f \in V^*$ and $\alpha \in \mathbb{R}$ such that for all $x \in \overline{W}$,

$$\mathrm{Re}(f)(x) \ < \ \alpha \ < \ \mathrm{Re}(f)(x_0).$$

(We assume here that the base field is $\mathbb{C}$; if it is $\mathbb{R}$, then we can write f instead of $\mathrm{Re}(f)$.) Since W is a linear subspace, it follows that for all $\lambda \in \mathbb{R}$, we have $\lambda f(x) < \alpha$ for all $x \in W$. Now, since $\mathbf{0} \in W$, we have $\alpha > 0$. On the other hand, setting $\lambda = n$, we get that, for any $x \in W$,

$$\mathrm{Re}(f)(x) \ < \ \frac{\alpha}{n}$$

whence we see that $\mathrm{Re}(f)(x) \leq 0$ for all $x \in W$. Again, if $x \in W$, we also have $-x \in W$ and so $\mathrm{Re}(f)(-x) \leq 0$ as well and so $\mathrm{Re}(f)(x) = 0$ for all $x \in W$ and $\mathrm{Re}f(x_0) > \alpha > 0$. As already observed, the real part of a functional determines the functional and so the proof is complete. ■

The above proposition gives us a very powerful tool for determining the density of subspaces of a normed linear space. Let V be a normed linear space and let W be a subspace. Assume that we have a continuous linear functional on V which vanishes on W. If we can show that it then vanishes on all of V, it follows from the preceding proposition that $\overline{W} = V$, or, in other words, that W is dense in V. This is frequently used in several situations.

Proposition 3.3 *Let V be a normed linear space. Let $x \in V$. Then*

$$\|x\| \ = \ \sup_{f \in V^*, \ \|f\| \leq 1} |f(x)| \ = \ \max_{f \in V^*, \ \|f\| \leq 1} |f(x)|. \tag{3.1}$$

Proof: Clearly, $|f(x)| \leq \|f\| \ \|x\| \leq \|x\|$ when $\|f\| \leq 1$. On the other hand, by Proposition 3.1, there exists $f \in V^*$ such that $\|f\| = 1$ and $f(x) = \|x\|$ when x is non-zero. Thus the result is established for non-zero vectors and is trivially true for the null vector. ■

Compare the relation

$$\|f\| \ = \ \sup_{x \in V, \ \|x\| \leq 1} |f(x)|, \tag{3.2}$$

which is a *definition*, with the relation (3.1), which is a *theorem*. In the former, the supremum need not be attained, while in the latter the supremum is always attained and hence is a maximum.

4 Reflexive spaces

The relation (3.1) is the starting point for the investigation of a very nice property of Banach spaces called **reflexivity**.

Let $x \in V$ and define

$$J_x(f) \;=\; f(x)$$

for $f \in V^*$. Then, by virtue of (3.1), it follows that $J_x \in (V^*)^* = V^{**}$ and that, in fact,

$$\|J_x\|_{V^{**}} \;=\; \|x\|_V.$$

Thus $J : V \to V^{**}$ given by $x \mapsto J_x$ is a norm preserving linear transformation. Such a map is called an **isometry**. The map J is clearly injective and maps V isometrically onto a subspace of V^{**}.

Definition 4.1 *A Banach space V is said to be* **reflexive** *if the canonical imbedding $J : V \to V^{**}$, given above, is surjective.* ■

Thus, if V is reflexive, we can identify the spaces V and V^{**}, using the isometry, J. Since V^{**}, being a dual space, is always complete, the notion of reflexivity makes sense only for Banach spaces. By applying Proposition 3.3 to V^*, it is readily seen that the supremum in (3.2) is attained for reflexive Banach spaces. A deep result due to R. C. James is that the converse is also true: *if V is a Banach space such that the supremum is attained in (3.2) for all $f \in V^*$, then V is reflexive.*

We saw above that the map J, being an isometry, is injective. If V is finite dimensional, then

$$\dim(V) \;=\; \dim(V^*) \;=\; \dim(V^{**})$$

and so J is surjective as well. Thus every finite dimensional space is reflexive.

Examples of infinite dimensional reflexive spaces are the sequence spaces

$$\ell_p \;=\; \left\{ x = (x_i) \mid \sum_{i=1}^{\infty} |x_i|^p < +\infty \right\}$$

with the norm

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

where $1 < p < \infty$. The space ℓ_1 is *not* reflexive. Nor is the space of bounded sequences, ℓ_∞, with the norm

$$\|x\|_\infty = \sup_{1 \leq i < \infty} |x_i|.$$

The closed subspaces c of convergent sequences and c_0 of sequences converging to zero, of ℓ_∞ are not reflexive either. The space $\mathcal{C}[0,1]$ is also not reflexive.

One of the nice consequences of the Riesz representation theorem is that every Hilbert space is reflexive.

5 Vector Valued Integration

Let us consider the unit interval $[0,1]$ endowed with the Lebesgue measure. Let V be a normed linear space over $\mathbb{R}$. Let $\varphi : [0,1] \to V$ be a continuous mapping. We would like to give a meaning to the integral

$$\int_0^1 \varphi(t)\ dt$$

as a vector in V in a manner that the familier properties of integrals are preserved.

Using our experience with the integral of a continuous real valued function, one could introduce a partition

$$0 = x_0 < x_1 < \cdots < x_n = 1$$

and form Riemann sums of the form

$$\sum_{i=1}^{n} (x_i - x_{i-1})\varphi(\xi_i)$$

where $\xi_i \in [x_{i-1}, x_i]$ for $1 \leq i \leq n$, and define the integral (if it exists) as a suitable limit of such sums. Assume that such a limit exists and denote it by $y \in V$. Let $f \in V^*$. Then, by the continuity and linearity of f, it will follow that $f(y)$ will be the limit of the Riemann sums of the form

$$\sum_{i=1}^{n} (x_i - x_{i-1}) f(\varphi(\xi_i)).$$

But since $f \circ \varphi : [0,1] \to \mathbb{R}$ is continuous, the above limit of Riemann sums is none other than

$$\int_0^1 f(\varphi(t))\ dt.$$

Thus the integral of φ must satisfy the relation

$$f\left(\int_0^1 \varphi(t)\ dt\right) \;=\; \int_0^1 f(\varphi(t))\ dt \tag{5.1}$$

for all $f \in V^*$.

Notice that since V^* separates points of V, such a vector, if it exists, must be unique.

We use this to define the integral of a vector valued function. Let X be a set and let $\mathcal{S}$ be a σ-algebra of subsets of X, on which we have a meaure μ. We say that $(X, \mathcal{S}, \mu)$ is a measure space. A mapping $\varphi : X \to V$ is said to be *weakly measurable* if $f \circ \varphi : X \to \mathbb{R}$ (or $\mathbb{C}$, if the space V is a complex normed linear space) is measurable for every $f \in V^*$.

Definition 5.1 *Let $(X, \mathcal{S}, \mu)$ be a measure space. Let V be a real normed linear space and let $\varphi : X \to \mathbb{R}$ be a weakly measurable mapping. The integral of φ over X, denoted*

$$\int_X \varphi(x)\ d\mu(x),$$

is that vector $y \in V$ which satisfies

$$f(y) \;=\; \int_X f(\varphi(x))\ d\mu(x)$$

for all $f \in V^*$. ■

Proposition 5.1 *Let $(X, \mathcal{S}, \mu)$ be a measure space and let V be a reflexive space. Let $\varphi : X \to V$ be a weakly measurable maping such that*

$$\int_X \|\varphi(x)\|\ d\mu(x) \;<\; +\infty.$$

Then the integral $\int_X \varphi(x)\ d\mu(x)$ exists.

Proof: For $f \in V^*$, define

$$\lambda(f) \;=\; \int_X f(\varphi(x))\; d\mu(x).$$

Then,

$$|\lambda(f)| \;\leq\; \|f\| \int_X \|\varphi(x)\|\; d\mu(x),$$

which, by hypothesis, is finite. This shows that λ is well-defined and that it is also a continuous linear functional on V^*. But V is reflexive and so there exists $y \in V$ such that $J_y = \lambda$. It now follows from the definition that y is indeed the required integral. ■

Proposition 5.2 *Let $\varphi : [0,1] \to V$ be a continuous mapping into a real Banach space V. Then the integral of φ over $[0,1]$ exists.*

Proof: (Sketch) Since $[0,1]$ is compact, the set $\overline{H}$ which is the closure (in V) of the set H which is the convex hull of $\varphi([0,1])$ (*i.e.* the smallest convex set containing $\varphi([0,1])$), is compact, by the completeness of V.

Let L be an arbitrary finite collection of continuous linear functionals on V. Define

$$E_L \;=\; \left\{ y \in \overline{H} \mid f(y) = \int_0^1 f(\varphi(t))\; dt \;\; \text{for all} \;\; f \in L \right\}.$$

It is immediate to see that E_L is a closed set.

Using the geometric version of the Hahn-Banach theorem, we can show that for any such finite collection L of continuous linear functionals, $E_L \neq \emptyset$.

Let I be a finite indexing set and let L_i be finite collections of elements in V^* for each $i \in I$. Then $L = \cup_{i\in I} L_i$ is still finite and further, it is easy to see that

$$\cap_{i\in I} E_{L_i} \;=\; E_L.$$

It now follows from the previous step that the class of closed sets

$$\{E_L \mid L \;\; \text{a finite subset of} \;\; V^*\}$$

has finite intersection property. Since $\overline{H}$ is compact, it now follows that

$$\cap_{L,\ \text{finite subset of }\ V^*} E_L \ \neq \emptyset.$$

In particular, there exists y such that $y \in E_{\{f\}}$ for every $f \in V^*$, *i.e.* y satisfies

$$f(y) \ = \ \int_0^1 f(\varphi(t))\ dt$$

for every $f \in V^*$. Thus $y = \int_0^1 \varphi(t)\ dt$. This completes the proof. ■

6 Lagrange multipliers

Let V be a normed linear space and let $U \subset V$ be an open set. Let $J : U \to \mathbb{R}$ be a given functional.

Definition 6.1 *The* **(Fréchet) derivative** *of J at a point $u \in U$, if it exists, is denoted $J'(u)$, and is a continuous linear functional on V such that*

$$|J(u+h) - J(u) - J'(u)(h)| \ = \ o(\|h\|),$$

where $o(\|h\|)$ is a vector $\varepsilon(h)$ such that

$$\frac{\|\varepsilon(h)\|}{\|h\|} \ \to \ 0 \text{ as } \|h\| \ \to \ 0. \ ■$$

If $V = \mathbb{R}^N$, then $J'(u)$ is just the gradient vector

$$\nabla J(u) = \left(\frac{\partial J}{\partial x_1}, \cdots, \frac{\partial J}{\partial x_N}\right).$$

The functional J is said to admit a relative maximum (repectively, minimum) at a point $u \in U$ if there exists a neighbourhood W of U such that $J(u) \geq (J(w)$ (respectively, $J(u) \leq J(w)$), for all $w \in W$. It is not difficult to see that if J attains a relative maximum or minimum at $u \in U$, and if J is differentiable at u, then $J'(u) = 0$.

Let us now consider the problem of finding extrema under constraints. Let $J : V \to \mathbb{R}$ be functional and let $\{\varphi_i\}_{i=1}^m$ be continuously differentiable functionals defined on V. We look for a relative extremum u of J subject to the constraints $\varphi_i(u) = 0, 1 \leq i \leq m$. Let

$$K = \{v \in V \mid \varphi_i(v) = 0, 1 \leq i \leq m\}.$$

The set K is not open in general and so we cannot apply the previous result and so we cannot deduce that $J'(u) = 0$.

Let us assume that J attains a relative extremum at a point $u \in K$ and that the gradient vectors $\{\varphi'(u)\}_{i=1}^m$ are all linearly independent. Then, by an application of the implicit function theorem, one can show that if $v \in V$ such that $\varphi_i'(u)(v) = 0$ for all $1 \leq i \leq m$, then $J'(u)(v) = 0$ as well (cf. Kesavan [3]).

Proposition 6.1 *Let V be a vector space and let $\{f_i\}_{i=0}^k$ be linear functionals on V such that*

$$\cap_{i=1}^k Ker(f_i) \subset Ker(f_0).$$

Then, there exist scalars $\{\lambda_i\}_{i=1}^k$ such that

$$f_0 = \sum_{i=1}^k \lambda_i f_i.$$

Proof: Consider the linear map $A : V \to \mathbb{R}^{k+1}$ (we assume here that the base field is $\mathbb{R}$) defined by

$$A(x) = (f_0(x), f_1(x), \cdots, f_k(x)).$$

The range is a subspace of $\mathbb{R}^{k+1}$ and , by hypothesis, $(1, 0, \cdots, 0)$ does not belong to this range. Hence (cf. Proposition 3.2) there exist scalars $\beta_i, 0 \leq i \leq k$ such that $\beta_1 \neq 0$ and $\sum_{i=0}^k \beta_i f_i(x) = 0$ for all $x \in V$. The result now follows with $\lambda_i = -\frac{\beta_i}{\beta_1}$. ■

Using the above proposition, we deduce that there exist scalars $\lambda_i, 1 \leq i \leq m$ such that

$$J'(u) = \sum_{i=1}^m \lambda_i \varphi'(u).$$

Let $V = \mathbb{R}^n$ and let J and $\varphi_i, 1 \leq i \leq m$ be continuously differentiable functions. Assume that we have a relative extremum u of J such that $\varphi_i(u) =$

0 for $1 \leq i \leq m$. Then, if $\{\nabla\varphi_i(u)\}_{i=1}^m$ are linearly independent, then we have m scalars $\lambda_i, 1 \leq i \leq m$ such that $J'(u) = \sum_{i=1}^m \lambda_i \nabla\varphi_i(u) = 0$. Let $u = (u_1, \cdots, u_n)$. then we solve for the $n+m$ unknowns $\{u_i \mid 1 \leq i \leq n\} \cup \{\lambda_i \mid 1 \leq i \leq m\}$ by solving the following set of $n+m$ equations:

$$\begin{array}{rcl}
\frac{\partial J}{\partial x_1} - \lambda_1 \frac{\partial \varphi_1}{\partial x_1} - \cdots - \lambda_m \frac{\partial \varphi_m}{\partial x_1} & = & 0, \\
\cdots\ \cdots\ \cdots & = & 0, \\
\frac{\partial J}{\partial x_n} - \lambda_1 \frac{\partial \varphi_1}{\partial x_n} - \cdots - \lambda_m \frac{\partial \varphi_m}{\partial x_n} & = & 0, \\
\varphi_1(u) & = & 0, \\
\cdots & = & 0, \\
\varphi_m(u) & = & 0,
\end{array}$$

which is exactly the method of *Lagrange multipliers* in the calculus of several variables.

BOX 1

For those familiar with the theory of distributions, Proposition 6.1, is applied there as well. Let $\alpha = (\alpha_1, \cdots, \alpha_N)$ be a multi-index, *i.e.* a N-tuple of non-negative integers. We set $|\alpha| = \alpha_1 + \cdot + \alpha_N$ and define

$$D^\alpha \;=\; \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_N^{\alpha_N}}.$$

The Dirac distribution δ and all its derivatives have support $\{0\}$. Proposition 6.1 is used to show the converse: if T is a disribution of $\mathbb{R}^N$ whose support is $\{0\}$, then, there exists a positive integer k and constants $c_\alpha, |\alpha| \leq k$, such that

$$T \;=\; \sum_{|\alpha| \leq k} c_\alpha D^\alpha \delta.$$

(see, Kesavan [2]).

Another application of the Hahn-Banach theorem in the theory of distributions is the *Malgrange-Ehrenpreis theorem.* Let P denote a polynomial in N variables. If α is a multi-index, then replace the monomial $x_1^{\alpha_1} \cdots x_N^{\alpha_N}$ by D^α to get the corresponding differential operator $P(D)$. The theorem states that there is a distribution T on $\mathbb{R}^N$ such that $P(D)(T) = \delta$, where δ is the Dirac distribution (cf. Rudin [5], for a proof). Such a distribution is called a *fundamental solution* of the partial differential operator $P(D)$. For example $\frac{1}{2\pi}\log|x|$ is a fundamental solution of the Laplace operator when $N = 2$ and $-\frac{1}{4\pi|x|}$ is a fundamental solution of the same operator in $\mathbb{R}^3$. Fundamental solutions have several uses; they can be used to solve partial differential equations using the technique of convolution and they also give important information about the smoothness of solutions of partial differential equations when the data is smooth.

7 Convex programming

In the previous section, we considered the relative extrema of functionals in the presence of constarits of th form $\varphi_i(x) = 0, 1 \leq i \leq m$. We now consider the relative extrema of functionals under inequality constraints, *i.e.* constraints of the form $\varphi(x) \leq 0, 1 \leq i \leq m$. The key step to this study is the analogue of Proposition 6.1, which is called the **Farkas-Minkowski lemma**.

Definition 7.1 *A* **cone** *in a real vector space V is a set C such that:*
*(*i*) $\mathbf{0} \in C$;*
*(*ii*) if $x \in C$ and $\lambda \geq 0$, then $\lambda x \in C$.* ■

Lemma 7.1 *Let v_i, $1 \leq i \leq n$ be elements in a normed linear space V. Define*

$$C \ = \ \left\{ \sum_{i=1}^{n} \lambda_i v_i \mid \lambda_i \geq 0, \ 1 \leq i \leq n \right\}.$$

Then C is a closed convex cone. ■

It is easy to see that C is a convex cone. It is also not difficult to see that it is closed when the vectors are all linearly independent. It is possible to reduce the general case to the linearly independent case to get a complete proof of this lemma.

Proposition 7.1 *(***Farkas-Minkowski Lemma***) Let V be a real reflexive Banach space and let $\{f_0, f_1, \cdots, f_n\}$ be elements of V^* such that if for some $x \in V$ we have $f_i(x) \geq 0$ for all $1 \leq i \leq n$, then $f_0(x) \geq 0$ as well. Then, there exists scalars $\lambda_i \geq 0$, $1 \leq i \leq n$ such that*

$$f_0 \ = \ \sum_{i=1}^{n} \lambda_i f_i.$$

Proof: Let
$$C \;=\; \left\{ \sum_{i=1}^{n} \lambda_i f_i \mid \lambda_i \geq 0,\ 1 \leq i \leq n \right\}$$
which is a closed convex cone in V^* by the preceding lemma. Assume that $f_0 \notin C$. Then, by the Hahn-Banach Theorem (cf. Theorem 2.5) there exist $\varphi \in V^{**}$ and $\alpha \in \mathbb{R}$ such that
$$\varphi(f_0) \;<\; \alpha \;<\; \varphi(f)$$
for all $f \in C$. Since $\mathbf{0} \in C$, it follows that $\alpha < 0$. Thus $\varphi(f_0) < 0$ as well.

Now, since V is reflexive, there exists $x \in V$ such that $\varphi = J_x$ and so $f_0(x) < 0$. On the other hand, since C is a cone, for all $\lambda > 0$, and for all $f \in C$, we have $\lambda f \in C$ and so $\varphi(\lambda f) > \alpha$ or, $\varphi(f) > \alpha/\lambda$ whence we deduce, on letting λ tend to infinity, that $\varphi(f) \geq 0$, *i.e.* $f(x) \geq 0$ for all $f \in C$. In particular $f_i(x) \geq 0$ for all $1 \leq i \leq n$ while $f_0(x) < 0$, which is a contradiction. Thus $f_0 \in C$ and the proof is complete. ■

Let V be a real normed linear space and let $J : V \to \mathbb{R}$ be a given functional. Let $K \subset V$ be a closed and convex subset. Then, if J attains a minimum over K at $u \in K$ and if J is differentiable at u, a necessary condition is that
$$J'(u)(v-u) \;\geq\; 0$$
for all $v \in K$ (Exercise!). We would like to generalize this to sets K which are not necessarily convex.

Let $\varphi_i, 1 \leq i \leq m$ be a finite set of functionals on V. Set
$$U \;=\; \{v \in V \mid \varphi_i(u) \leq 0,\ 1 \leq i \leq m\}. \tag{7.1}$$
Of particular interest is the case when the functionals φ_i are affine linear, *i.e.* there exist $f_i \in V^*$ and $d_i \in \mathbb{R}$ for $1 \leq i \leq m$ such that
$$\varphi_i(u) \;=\; f_i(u) + d_i \tag{7.2}$$
for $1 \leq i \leq m$. In this case, we can prove the following result.

Proposition 7.2 *Let U be as given by (7.1) and let the constraints φ_i be affine linear, given by (7.2). Then, for any $u \in U$, set*
$$C(u) \;=\; \{w \in V \mid f_i(w) \leq 0,\ i \in I(u)\} \tag{7.3}$$

where

$$I(u) \;=\; \{i \mid 1 \leq i \leq m,\ \varphi_i(u) = 0\}.$$

If J attains a relative extremum at $u \in U$, and if J is differentiable at u, then

$$J'(u)(w) \;\geq\; 0,\ \textit{for all } w \in C(u). \ \blacksquare$$

Theorem 7.1 *(***Kuhn-Tucker Conditions***) Let V be a real, reflexive Banach space. Let φ_i, $1 \leq i \leq m$ be as in (7.2) and let U be as in (7.1). Let $J : V \to \mathbb{R}$ be a functional which attains a relative minimum at $u \in U$. Assume that J is differentiable at u. Then, there exist constants $\lambda_i(u)$ such that*

$$\left.\begin{array}{rcl} J'(u) + \sum_{i=1}^m \lambda_i(u)\varphi_i'(u) & = & \mathbf{0} \\[2ex] \sum_{i=1}^m \lambda_i(u)\varphi_i(u) & = & 0 \\[2ex] \lambda_i(u) & \geq & 0,\ 1 \leq i \leq m. \end{array}\right\} \qquad (7.4)$$

Proof: By Proposition 7.2, we have that for all w such that $\varphi_i'(u)w \leq 0,\ i \in I(u)$, we have $J'(u)w \geq 0$. Thus, by the Farkas-Minkowski lemma, there exist $\lambda_i(u) \geq 0$ for $i \in I(u)$ such that

$$J'(u) \;=\; -\sum_{i \in I(u)} \lambda_i(u)\varphi_i'(u).$$

Setting $\lambda_i(u) = 0$ for all $i \notin I(u)$, we get (7.4). This completes the proof. $\blacksquare$

The above theorem can be generalized to cases when the φ_i are not affine. In this situation, in addition to differentiability at u, we need to assume another technical condition of 'admissibility' on the constraints at u. In particular, when the constraints φ_i, $1 \leq i \leq m$ are all convex, the admissibility condition reads as follows:

- either, all the φ_i are affine and the set U given by (7.1) is non-empty;
- or, there exists an element $v^* \in V$ such that $\varphi_i(v^*) \leq 0$ for all $1 \leq i \leq m$ and $\varphi_i(v^*) < 0$ whenever φ_i is not affine linear.

If J is differentiable at u and the constraints are differentiable and admissible (at u), then (7.4) is a necessary condition for u to be a relative minimum of J at u. In addition, if J and the constraints φ_i are all convex, then (7.4) is both necessary and sufficient. Interested readers can find further details in the book by Ciarlet [1].

References

[1] Ciarlet, P. G. Introduction à l'analyse numérique matricielle et à l'optimisation, Masson, Paris, France, 1982; English translation, Cambridge University Press, Cambridge, UK, 1989.

[2] Kesavan, S. Topics in Functional Analysis and Applications, New Age International (P) Limited (originally Wiley-Eastern), 1989.

[3] Kesavan, S. Nonlinear Functional Analysis: A First Course, Texts and readings in Mathematics (TRIM), **28**, Hindustan Book Agency, 2004.

[4] Kesavan, S. Functional Analysis, Texts and readings in Mathematics (TRIM), **52**, Hindustan Book Agency, 2009.

[5] Rudin, W. Functional Analysis, McGraw-Hill, 1973.