Topic 23: Separation and the Hahn–Banach theorem

23.1 Other approaches to separation theorems

The technique use to prove the Strong Separating Hyperplane Theorem 8.3.1 in Lecture 8 has the virtue that it is relatively elementary—it uses only mathematics found in say, Apostol's *Calculus* [2]. But it is not the only approach.

One traditional approach used by convex analysts (e.g., Klee [5], Rockafellar [7], Lay [6]) makes use of Zorn's Lemma 22.2.1 to show that if A and B are nonempty disjoint convex sets in a linear space X, then there are disjoint convex sets $\hat{A}$ and $\hat{B}$ such that $A \subset \hat{A}$, $B \subset \hat{B}$, and $\hat{A} \cup \hat{B} = X$. This implies the existence of a separating hyperplane.

Another approach used by analysts is based on the Hahn–Banach Extension Theorem (e.g., Royden [8]). It is also proved using Zorn's Lemma. Holmes [4] points out that a version of the separating hyperplane theorem can be used to prove the Hahn–Banach Theorem, and vice versa.

23.2 Extension of linear functionals

We first show that linear extensions of linear functionals always exist. This is not the Hahn–Banach Extension Theorem. That theorem imposes additional constraints on the extension.

23.2.1 Theorem *Let X be a vector space, and let $f\colon M \to \boldsymbol{R}$ be linear. Then there is an extension $\hat{f}$ of f to a linear functional on X.*

Proof: We show below in Theorem 23.8.2 that there is a subspace N of X that is complementary to M. That is, for each $x \in X$ there is a unique decomposition

$$x = x_M + x_N, \text{ where } x_M \in M \text{ and } x_N \in N.$$

Define $\hat{f}$ by

$$\hat{f}(x_M + x_N) = f(x_M),$$

so that $\hat{f}(z) = 0$ for every $z \in N$. Then $\hat{f}$ extends f and is linear on X:

$$\begin{aligned}\hat{f}(\alpha x + \beta y) = \hat{f}\Big(\alpha(x_M + x_N) + \beta(y_M + y_N)\Big) &= \hat{f}(\underbrace{\alpha x_M + \beta y_M}_{\in M} + \underbrace{\alpha x_N + \beta y_N}_{\in N}) \\ &= f(\alpha x_M + \beta y_M) = \alpha f(x_M) + \beta f(y_M) = \alpha \hat{f}(x) + \beta \hat{f}(y).\end{aligned}$$

∎

23.3 The Hahn–Banach Extension Theorem

One of the most important and far-reaching results in analysis is the seemingly mild Hahn–Banach Extension Theorem.

23.3.1 Definition *Let $f, g\colon A \to \boldsymbol{R}$. We say that g **dominates** f on A, written*

$$g \geqq f$$

if

$$(\,\forall x \in A\,)\;[\,g(x) \geqslant f(x)\,].$$

*We say that g **strictly dominates** f on A, written*

$$g \gg f$$

if

$$(\,\forall x \in A\,)\;[\,g(x) > f(x)\,].$$

23.3.2 Definition *Let A be a nonempty subset of a set X and let $f\colon A \to Y$. We say that a function $\hat{f}\colon X \to Y$ **extends** f to X, or is an **extension** of f to X, if*

$$(\,\forall x \in A\,)\;\Big[\hat{f}(x) = f(x)\Big].$$

*In other words, f is the **restriction** of $\hat{f}$ to A, often written*

$$f = \hat{f}\,|_A\,.$$

23.3.3 Hahn–Banach Extension Theorem *Let X be a vector space and let $h\colon X \to \boldsymbol{R}$ be a convex function. Let V be a vector subspace of X and let $f\colon V \to \boldsymbol{R}$ be a linear functional dominated by h on V. Then there is a (not generally unique) extension $\hat{f}$ of f to a linear function defined on all of X that is dominated by h on X.*

23.3.4 Remark Some statements of the Hahn–Banach Theorem (e.g., Royden [8, Theorem 10.3.4, p. 233], Dunford and Schwartz [3, Theorem II.3.10, p. 62], Wilanksy [9, Theorem 12.4.1, p. 269]) impose other conditions on h, such as sublinearity (positive homogeneity and subadditivity). Together these imply convexity. It turns out the homogeneity is unnecessary, but it is typically satisfied in the cases they have in mind. We shall see (I hope) that if a linear function f is dominated by a convex function h, there is a sublinear function p (the directional derivative) satisfying $h \geqq p \geqq f$.

Make sure we do this.

Proof of Theorem 23.3.3: The proof is an excellent example of what is known as transfinite induction. It has two parts. The first part says that any dominated extension g of f whose domain is not all of X has a dominated extension to a strictly larger subspace. The second part says that this is enough to conclude that there is a dominated extension defined on all of X.

So let M be a subspace of X that includes V and assume that we have an extension g of f to M that satisfies $g \leqq h$ on M. If $M = X$, then we are done. So suppose there exists $v \in X \setminus M$. Let N be the linear span of $M \cup \{v\}$. For each $x \in N$ there is a unique decomposition

$$x = z + \lambda v \text{ where } z \in M.$$

(To see the uniqueness, suppose $x = z_1 + \lambda_1 v = z_2 + \lambda_2 v$. Then $z_1 - z_2 = (\lambda_2 - \lambda_1)v$. Since $z_1 - z_2 \in N$ and $v \notin N$, it must be the case that $\lambda_2 - \lambda_1 = 0$. But then $\lambda_1 = \lambda_2$ and $z_1 = z_2$.)

Any linear extension $\hat{g}$ of g satisfies $\hat{g}(z+\lambda v) = g(z) + \lambda \hat{g}(v)$, so the requirement that it be dominated by h on N becomes

$$(\,\forall z \in M\,)\;\;(\,\forall \lambda \in \boldsymbol{R}\,)\;[\,g(z) + \lambda \hat{g}(v) \leqslant h(z + \lambda v)\,]. \tag{1}$$

Thus the problem of finding a dominated extension of g to N reduces to showing that we can pick $\hat{g}(v)$ to satisfy (1). For $\lambda = 0$, the inequality in (1) is automatically satisfied. For $\lambda > 0$, the inequality in (1) reduces to

$$\hat{g}(v) \leqslant \frac{h(z + \lambda v) - g(z)}{\lambda},$$

and for $\lambda < 0$, the inequality in (1) reduces to

$$\hat{g}(v) \geqslant \frac{h(z + \lambda v) - g(z)}{\lambda}.$$

Letting $\mu = -\lambda$, we can rewrite this as

$$\hat{g}(v) \geqslant \frac{g(z) - h(z - \mu v)}{\mu}.$$

So we have reduced the problem to showing that there is some real number $\hat{g}(v)$ such that

$$(\,\forall x, y \in M\,)\;\;(\,\forall \mu, \lambda > 0\,)\;\left[\frac{g(x) - h(x - \mu v)}{\mu} \leqslant \hat{g}(v) \leqslant \frac{h(y + \lambda v) - g(y)}{\lambda}\right]. \tag{2}$$

The constant $\hat{g}(v)$ plays no real rôle here, and we may simply omit it. Then multiplying by $\mu\lambda > 0$ and rearranging terms, we see that (2) is equivalent to

$$(\,\forall x, y \in M\,)\;\;(\,\forall \mu, \lambda > 0\,)\;[\,g(\lambda x + \mu y) \leqslant \lambda h(x - \mu v) + \mu h(y + \lambda v)\,]. \tag{3}$$

Thus, a dominated extension of g to N exists if and only if (3) is valid. But proving this is straightforward: if $x, y \in M$ and $\lambda, \mu > 0$, then

$$\begin{aligned}
g(\lambda x + \mu y) &= (\lambda+\mu) g\Big(\frac{\lambda}{\lambda+\mu}x + \frac{\mu}{\lambda+\mu}y\Big) && \text{(linearity of } g\text{)}\\
&\leqslant (\lambda+\mu) h\Big(\frac{\lambda}{\lambda+\mu}x + \frac{\mu}{\lambda+\mu}y\Big) && \text{(} h \text{ dominates } g\text{)}\\
&= (\lambda+\mu) h\Big(\frac{\lambda}{\lambda+\mu}[x-\mu v] + \frac{\mu}{\lambda+\mu}[y+\lambda v]\Big) && \text{(algebra)}\\
&\leqslant (\lambda+\mu)\Big[\frac{\lambda}{\lambda+\mu}h(x-\mu v) + \frac{\mu}{\lambda+\mu}h(y+\lambda v)\Big] && \text{(convexity of } h\text{)}\\
&= \lambda h(x-\mu v) + \mu h(y+\lambda v).
\end{aligned}$$

This shows that as long as there is some $v \notin M$, there is a further extension $\hat{g}$ of g to the larger subspace $N = \operatorname{span}(M \cup \{v\})$ that satisfies $\hat{g} \leqq h$ on N.

To conclude the proof, consider the set $\mathcal{G}$ of all pairs (g, M) of such that M is a linear subspace of X with $V \subset M$, $g \colon M \to \boldsymbol{R}$ is a linear functional, $g|_V = f$, and $g(x) \leqslant h(x)$ for all $x \in M$. Introduce the partial order $\geq$ on $\mathcal{G}$ by $(g', M') \geq (g, M)$ if $M' \supset M$ and $g'|_M = g$. Note that this relation is indeed a partial order.

If $\{(g_\alpha, M_\alpha)\}$ is a chain (a linearly ordered subset) in $\mathcal{G}$, then the function g defined on the linear subspace $M = \bigcup_\alpha M_\alpha$ by $g(x) = g_\alpha(x)$ for $x \in M_\alpha$ is well defined and linear, $g(x) \leqslant h(x)$ for all $x \in M$, and $(g, M) \geqslant (g_\alpha, M_\alpha)$ for each α. By Zorn's Lemma 22.2.1, there is a maximal extension $(\hat{f}, M)$ in $\mathcal{G}$. By the first part of the argument, an extension is not maximal unless $M = X$. ∎

The next result tells us when a sublinear functional is actually linear.

23.3.5 Theorem *A sublinear function $h \colon X \to \boldsymbol{R}$ on a vector space is linear if and only if it dominates exactly one linear functional on X.*

Proof: First let $h \colon X \to \boldsymbol{R}$ be a sublinear functional on a vector space. If h is linear and $f(x) \leqslant h(x)$ for all $x \in X$ and some linear functional $f \colon X \to \boldsymbol{R}$, then $-f(x) = f(-x) \leqslant h(-x) = -h(x)$, so $h(x) \leqslant f(x)$ for all $x \in X$, that is, $f = h$.

Now assume that the sublinear function h dominates exactly one linear functional on X. Note that h is linear if and only if $h(-x) = -h(x)$ for each $x \in X$. So if we assume by way of contradiction that h is not linear, then there exists some $x_0 \neq 0$ such that $-h(-x_0) < h(x_0)$. Let $V = \{\lambda x_0 : \lambda \in \boldsymbol{R}\}$, the vector subspace generated by x_0, and define the linear functionals $f, g \colon V \to \boldsymbol{R}$ by $f(\lambda x_0) = \lambda h(x_0)$ and $g(\lambda x_0) = -\lambda h(-x_0)$. From $f(x_0) = h(x_0)$ and $g(x_0) = -h(-x_0)$, we see that $f \neq g$. Next, notice that $f(z) \leqslant h(z)$ and $g(z) \leqslant h(z)$ for each $z \in V$, that is, h dominates both f and g on the subspace V. Now by the Hahn–Banach Theorem 23.3.3, the two distinct linear functionals f and g have linear extensions to all of X that are dominated by h, a contradiction. ∎

23.4 Another Separating Hyperplane Theorem

23.4.1 Theorem (Algebraic Separating Hyperplane Theorem) *Let A and B be disjoint nonempty convex subsets of X. Assume that* $\operatorname{cor} A \neq \varnothing$. *Then there is a hyperplane $\{\varphi = \alpha\}$ properly separating A and B.*

That is, $A \subset \{\varphi \geqslant \alpha\}$ and $B \subset \{\varphi \leqslant \alpha\}$, and there exists some $a \in A$ and $b \in B$ with $\varphi(a) > \varphi(b)$.

We may also say that the functional φ separates A and B.

23.5 Extension and Separation

Proof of Hahn–Banach 23.3.3 using Separation 23.4.1: Let M be a vector subspace of the vector space X, and let $g\colon X \to \boldsymbol{R}$ be convex. Let $\varphi\colon M \to \boldsymbol{R}$ be a linear functional on M, and assume that g dominates φ on M.

Let

$$A = \{(x, \alpha) \in X \times \boldsymbol{R} : \alpha > g(x)\}, \qquad B = \{(x, \alpha) \in M \times \boldsymbol{R} : \alpha \leqslant \varphi(x)\},$$

Then A and B are disjoint (since $g \geqq \varphi$) and convex. Moreover it is easy to see that every point in A is a core point.

Thus by the Algebraic Separating Hyperplane Theorem 23.4.1, there is a separating hyperplane, which must be non-vertical. It is thus the graph of an affine function $\hat{\varphi} + \beta$, which satisfies Elaborate on this.

$$\hat{\varphi} + \beta \geqq \varphi \text{ on } M,$$

so

$$\hat{\varphi} = \varphi \text{ on } M \text{ and } \beta \geqslant 0.$$

On the other hand

$$\hat{\varphi} \leqq \hat{\varphi} + \beta \leqq g \text{ everywhere.}$$

Thus $\hat{\varphi}$ is the desired extension of φ. ∎

The next proof is standard, and is taken from the Hitchhiker's Guide [1, Theorem 5.61].

Proof of Separation 23.4.1 using Hahn–Banach 23.3.3: Let A and B be disjoint nonempty convex sets in a vector space X, and suppose A has an internal point.

Then the nonempty convex set $A - B$ has an internal point. Let z be an internal point of $A - B$. Clearly, $z \neq 0$ and the set $C = A - B - z$ is nonempty, convex, absorbing, and satisfies $-z \notin C$. (Why?) By part (2) of Lemma 9.3.6, the gauge p_C of C is a sublinear function.

We claim that $p_C(-z) \geq 1$. Indeed, if $p_C(-z) < 1$, then there exist $0 \leq \alpha < 1$ and $c \in C$ such that $-z = \alpha c$. Since $0 \in C$, it follows that $-z = \alpha c + (1-\alpha)0 \in C$, a contradiction. Hence $p_C(-z) \geq 1$.

Let $M = \left\{\alpha(-z) : \alpha \in \boldsymbol{R}\right\}$, the one-dimensional subspace generated by $-z$, and define $f\colon M \to \boldsymbol{R}$ by $f(\alpha(-z)) = \alpha$. Clearly, f is linear and moreover $f \leq p_C$ on M, since for each $\alpha \geqslant 0$ we have $p_C(\alpha(-z)) = \alpha p_C(-z) \geq \alpha = f(\alpha(-z))$, and $\alpha < 0$ yields $f(\alpha(-z)) < 0 \leq p_C(\alpha(-z))$. By the Hahn–Banach Extension Theorem 23.3.3, f extends to $\hat{f}$ defined on all of X satisfying $\hat{f}(x) \leq p_C(x)$ for all $x \in X$. Note that $\hat{f}(z) = -1$, so $\hat{f}$ is nonzero.

To see that $\hat{f}$ separates A and B let $a \in A$ and $b \in B$. Then we have

$$\begin{aligned}\hat{f}(a) &= \hat{f}(a-b-z) + \hat{f}(z) + \hat{f}(b) \leq p_C(a-b-z) + \hat{f}(z) + \hat{f}(b)\\ &= p_C(a-b-z) - 1 + \hat{f}(b) \leq 1 - 1 + \hat{f}(b) = \hat{f}(b).\end{aligned}$$

This shows that the nonzero linear functional $\hat{f}$ separates the convex sets A and B.

To see that the separation is proper, let $z = a - b$, where $a \in A$ and $b \in B$. Since $\hat{f}(z) = -1$, we have $\hat{f}(a) \neq \hat{f}(b)$, so A and B cannot lie in the same hyperplane. ∎

23.6 Other equivalent propositions

23.6.1 Extension of Positive Operators

23.6.1 Theorem (Krein–Rutman Theorem) *Let P be the positive cone of X, and M a linear subspace of X. Assume that $P \cap M$ contains a core point of P. Consider M to be an ordered vector space with positive cone $P \cap M$.*

If $\varphi\colon M \to \boldsymbol{R}$ is a positive linear functional on M, then there is an extension $\hat{\varphi}$ of φ to X so that $\hat{\varphi}$ is a positive functional on X

23.6.2 Existence of Positive Operators

23.6.2 Definition *A **wedge** P is a convex cone in X. It defines a partial order $\geqq$ on X by*

$$x \geqq y \iff x - y \in P.$$

*$P = \{x : x \geqslant 0\}$ is the **positive cone**. A linear functional $\varphi\colon X \to \boldsymbol{R}$ is **positive** if*

$$(\,\forall x \geqq 0\,)\ [\,\varphi(x) \geqslant 0\,].$$

Note that the zero functional is positive.

23.6.3 Theorem (Existence of Positive Operators) *If P is a proper subset of X and $\operatorname{cor} P \neq \varnothing$, then there exists a nonzero positive linear functional on X.*

23.6.3 Support Theorem

23.6.4 Definition *Let $H = \{\varphi = \alpha\}$ and let A be a convex set in X. H **supports** A at x if A lies in a half-space determined by H, and $x \in A \cap H$. The support is **proper** if A is not a subset of H. The point x is called a **support point** of A.*

23.6.5 Lemma *Let A be a convex subset of X and assume $\operatorname{icr} A \neq \varnothing$. If $x \notin \operatorname{icr} A$, then there is a linear functional φ on X such that $(\forall y \in \operatorname{icr} A)\ [\,\varphi(x) > \varphi(y)\,]$.*

23.6.6 Corollary (Support Theorem) *Let A be a convex subset of X and assume that $\operatorname{icr} A \neq \varnothing$. Then x is a proper support point of A if and only if $x \in A \setminus (\operatorname{cor} A)$.*

23.6.4 Subdifferentiability

23.6.7 Definition *Let $g \colon A \to \boldsymbol{R}$ be a convex function. A linear functional φ is a **subgradient** of g at $a \in A$ if*

$$(\forall x \in A)\ [\,g(x) \geqslant g(a) + \varphi(x - a)\,].$$

*If g has a subgradient at a, then we say that g is **subdifferentiable** at a.*

23.6.8 Theorem (Subdifferentiability Theorem) *If g is a convex function on the convex set A in X, and $a \in \operatorname{icr} A$, then g is subdifferentiable at a.*

23.6.5 The plan

$$\begin{tikzcd}
 & \text{Separation Theorem} \arrow[dl, Rightarrow] & & \\
\text{Hahn–Banach} \arrow[dd, Rightarrow] & & & \\
 & & \text{Support Theorem} \arrow[uul, Rightarrow] \arrow[r, Leftrightarrow] & \text{Subgradients} \\
\text{Krein–Rutman} \arrow[dr, Rightarrow] & & & \\
 & \text{Nonzero Positive Functionals} \arrow[uur, Rightarrow] & &
\end{tikzcd}$$

23.6.6 Ancillary concepts

23.6.9 Fact (cf. [4, pp. 2–3]) *If M is a linear subspace of a vector space X, there is a (not unique) subspace N that is complementary to M. That is, $M \cap N = \{0\}$, and every $x \in X$ has a unique representation as $x = x_M + x_N$, where $x_M \in M$ and $x_N \in N$. This is expressed as $X = M \oplus N$.*

23.6.10 Corollary *If φ is a linear functional on M, then it can be extended to all of X via $\hat\varphi(x) = \varphi(x_M)$.*

23.6.7 Hahn–Banach implies Krein–Rutman

Let $\geqq$ be the order induced by P, and let φ be positive on M. Let Y be the span of $P \cup M$, and let $Y = M \oplus N$. For $y \in Y$, we may write

$$y = p_1 - p_2 + x,$$

where $p_1, p_2 \in P$ and $x \in M$.

Let
$$g(y) = \inf\{\varphi(x) : x \in M \ \&\ x \geqq y\}.$$
Then g is sublinear and $\varphi \leqq g$ on M. Extend φ to $\hat\varphi \leqq g$ on Y by Hahn–Banach. Now show that $\hat\varphi$ is positive:

Let $x \in P$ and let $\bar x \in P \cap M$. For $\lambda \geqslant 0$, we have

$$\bar x + \lambda x \in P.$$

Thus
$$\bar x/\lambda \geqq -x.$$
But $\bar x/\lambda \in M$, so by definition of g,

$$g(-x) \leqslant \varphi(\bar x/\lambda).$$

Thus
$$\hat\varphi(-x) \leqslant g(-x) \leqslant \varphi(\bar x)/\lambda.$$
Let $\lambda \to \infty$ to get $\hat\varphi(-x) \leqslant 0$, which proves that $\hat\varphi$ is positive on Y. Then use complementary subspaces to extend $\hat\varphi$ to all of X.

23.6.8 Krein–Rutman implies nonzero positive functionals

This is easy. Let $0 \neq \bar x \in \operatorname{cor} P$, and let $M = \operatorname{span}\{x\}$. Define φ on M by $\varphi(\lambda \bar x) = \lambda$ and apply Krein–Rutman.

23.6.9 Nonzero positive functionals imply support points

Assume $0 \in A \setminus \operatorname{cor}(A)$, and let P the cone generated by $\operatorname{cor} A$. Any positive functional supports A at 0.

23.6.10 Support points imply subdifferentiability

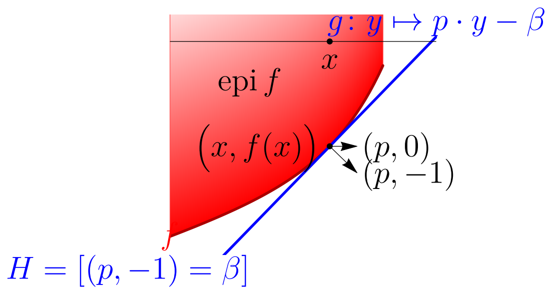

Figure 23.6.1. The affine function $g\colon y \mapsto p \cdot y - \beta$ satisfies $g \leqq f$ and $g(x) = f(x)$. Equivalently, The hyperplane $H = \{(y, \alpha) \in X \times \boldsymbol{R} : (p, -1) \cdot (y, \alpha) = \beta\}$ supports $\operatorname{epi} f$ at the point $\bigl(x, f(x)\bigr)$, which maximizes $(p, -1)$ over $\operatorname{epi} f$ and the maximum values is β.

23.6.11 Subdifferentiability implies support points

Let $x \in A \setminus (\operatorname{cor} A)$, and let ρ be the gauge function of A. The epigraph of ρ is a convex cone. Let φ be a subgradient of ρ at x. It supports the epigraph at $\bigl(x, \rho(x)\bigr)$. Slice through $X \times \boldsymbol{R}$ with the horizontal plane $\{(x, \alpha) : \alpha = 1\}$. WTF? This makes no sense.

23.6.12 Support points implies separating hyperplanes

23.6.11 Lemma *A linear functional φ properly separates convex sets A and B if and only if it properly supports $A - B$ at 0.*

23.7 ⋆ Digression: Quotient spaces

An **equivalence relation** $\sim$ on a set X is a binary relation that is transitive, symmetric, and reflexive. In other words, for all $x, y, z \in X$,

$$(x \sim y \;\&\; y \sim z) \implies x \sim z; \quad x \sim y \implies y \sim x; \quad \text{and} \quad x \sim x.$$

The **equivalence class** $[x]$ of x is defined by

$$[x] = \{y : y \sim x\}.$$

Observe that

$$x \sim y \iff [x] = [y]; \quad \text{and} \quad x \not\sim y \iff [x] \cap [y] = \varnothing.$$

Thus the $\sim$-equivalence classes form a partition of X into disjoint sets. The collection of $\sim$-equivalence classes of X is called the **quotient of X modulo $\sim$**, often written as $X/\sim$. The function $x \mapsto [x]$ from X to $X/\sim$ is called the **quotient map**.

In many contexts, mathematicians say that they **identify** the members of an equivalence class. What they mean by this is that they write X instead of $X/\sim$, and they write x instead of $[x]$.

Given any function f with domain X, we can define an equivalence relation $\sim$ on X by $x \sim y$ whenever $f(x) = f(y)$. This is one of the most common ways to define equivalence relations.

23.8 ⋆ Digression: Complementary subspaces

23.8.1 Definition *Let M and N be linear subspaces of a vector space X. We say that M and N are **complementary subspaces** if each x in X can be written in a* unique *way as*

$$x = x_M + x_N, \text{ where } x_M \in M \text{ and } x_N \in N.$$

*In this case we write $X = M \oplus N$ and say that X is the **direct sum** of M and N.*

It is well-known that every linear subspace M of $\boldsymbol{R}^{\mathrm{m}}$ has an orthogonal complement $M_\perp = \left\{x \in \boldsymbol{R}^{\mathrm{m}} : (\forall z \in M)\,[\,x \cdot z = 0\,]\right\}$. In more general linear subspaces there may not be an inner product, but nevertheless we still have the following.

23.8.2 Theorem *Every linear subspace of a vector space has a complementary subspace.*

Proof: (cf. Holmes[4, § C, pp. 2–3]) Let M be a linear subspace of the vector space X. Define the relation $\sim_M$ on X by

$$x \sim_M y \text{ if } x - y \in M.$$

Exercise 4.2.2 proves that $\sim_M$ is an equivalence relation. Let X/M denote the set of equivalence classes of $\sim_M$, and let $[x]$ denote the equivalence class of x. Then

$$[x] = x + M.$$

(See Exercise 4.2.2.)

We can turn X/M into a vector space by defining vector addition and scalar multiplication via

$$\alpha[x] + \beta[y] = [\alpha x + \beta y].$$

To verify that this is well defined, we need to show that if $[x] = [x']$ and $[y] = [y']$, then $[\alpha x + \beta y] = [\alpha x' + \beta y'.]$ That is, we need to show that

$$\Big(x - x' \in M \;\&\; y - y' \in M\Big) \implies (\alpha x + \beta y) - (\alpha x' + \beta y') \in M,$$

but this is clearly true. As a result $[0] = M$ is the zero of the vector space X/M.

We now have to show that X/M can be identified with a complementary subspace of X. Since X/M is a linear space, it has a basis $\{[b_i] : i \in I\}$ where I is some index set, and each b_i is a fixed representative of its $\sim_M$-equivalence class. It follows that $\{b_i : i \in I\}$ is a linearly independent subset of X. Moreover, $N = \operatorname{span}\{b_i : i \in I\}$ is complementary to M: Let $x \in X$. Then we can uniquely write

$$[x] = \sum_{i=1}^{k} \alpha_i [b_i]$$

since $\{[b_i] : i \in I\}$ is a basis for X/M. This means

$$\sum_{i=1}^{k} \alpha_i b_i \in [x] = x + M.$$

Let

$$x_N = \sum_{i=1}^{k} \alpha_i b_i \quad \text{and} \quad x_M = x_N - x.$$

Then $x_M \in M$, and $x = x_M + x_N$. This decomposition is unique. ∎

References

[1] C. D. Aliprantis and K. C. Border. 2006. *Infinite dimensional analysis: A hitchhiker's guide*, 3d. ed. Berlin: Springer–Verlag.

[2] T. M. Apostol. 1967–69. *Calculus*, 2d. ed., volume 1–2. Waltham, Massachusetts: Blaisdell.

[3] N. Dunford and J. T. Schwartz. 1957. *Linear operators: Part I*. New York: Interscience.

[4] R. B. Holmes. 1975. *Geometric functional analysis and its applications*. Number 24 in Graduate Texts in Mathematics. Berlin: Springer–Verlag.

[5] V. L. Klee, Jr. 1951. Convex sets in linear spaces. *Duke Mathematical Journal* 18(2):443–466. DOI: 10.1215/S0012-7094-51-01835-2

[6] R. Lay, Steven. 1982. *Convex sets and their applications*. Pure and Applied Mathematics: A Wiley-Interscience Series of Texts, Monographs, and Tracts. NY: Wiley–Interscience.

[7] R. T. Rockafellar. 1970. *Convex analysis*. Number 28 in Princeton Mathematical Series. Princeton: Princeton University Press.

[8] H. L. Royden. 1988. *Real analysis*, 3d. ed. New York: Macmillan.

[9] A. Wilansky. 1998. *Topology for analysis*. Mineola, NY: Dover. Unabridged republication of the work originally published by Ginn and Company, Waltham, MA in 1970.