Functional Analysis

Michael Nelson

Contents

Part I
Class Notes

1 Introduction

Given a measure μ, the nth **moment** is by definition $\int_I t^n \mathrm{d}\mu(t)$ where Ij is a subinterval of $\mathbb{R}$. The moment problem says that if we are given a sequence (a_n) of real numbers, can we find a measure μ such that

$$a_n = \int_I t^n \mathrm{d}\mu(t).$$

for all $n \in \mathbb{N}$. If $I = [0,1]$, then this is called the Hausdorff moment problem. If $I = [0,\infty)$, then this is called the Stieltjes moment problem. If $I = (-\infty,\infty)$, then this is called the Hamburger moment problem.

Let us start with some intuition on how we can solve this problem. For a function f and a measure μ, let us denote

$$\langle f, \mu \rangle = \int_I f \mathrm{d}\mu \tag{1}$$

In some sense, (1) behaves like an inner-product. Of course, f and μ are different types of mathematical objects; one is a function and the other is a measure. So for all functions f and measures μ.

1.1 Convex Cones

Definition 1.1. Let V be an $\mathbb{R}$-vector space. A set $K \subseteq V$ is said to be a **convex cone** if

1. if $x, y \in K$ then $x + y \in K$
2. if $x \in K$ and $\alpha \geq 0$, then $\alpha x \in K$.

Given a convex cone $K \subseteq V$, if we have the additional axiom $-K \cap K = \{0\}$, then we can define a partial order on V as follows: if $x, y \in V$, then we say $x \leq_K y$ if $y - x \in K$. In this case, we will have $0 \leq_K x$ for all $x \in K$. Thus it makes sense to call the elements of K the **positive** elements with respect to $\leq_K$.

1.1.1 Extending Linear Functionals Which Satisfy Positivity Condition

Theorem 1.1. *(Marcel Extension Theorem) Let V be an $\mathbb{R}$-vector space, let $W \subseteq V$ be a subspace of V, and let $K \subseteq V$ be a convex cone. Suppose $V = W + K$ and $\psi \colon W \to \mathbb{R}$ is a linear functional such that $\psi(x) \geq 0$ for all $x \in K \cap W$. Then there exists $\widetilde{\psi} \colon V \to \mathbb{R}$ such that $\widetilde{\psi}$ is a linear functional such that $\widetilde{\psi}|_W = \psi$ and such that $\widetilde{\psi}(x) \geq 0$ for all $x \in K$.*

Proof. Let $v \in V \backslash W$. We will first show that we can extend ψ to a linear functional $\widetilde{\psi} \colon W + \mathbb{R}v \to \mathbb{R}$ where

$$W + \mathbb{R}v = \{w + \lambda v \mid w \in W \text{ and } \lambda \in \mathbb{R}\}.$$

Define two sets

$$A = \{x \in W \mid x + v \in K\} \quad \text{and} \quad B = \{y \in W \mid y - v \in K\}.$$

We claim that

$$\sup\{-\psi(x) \mid x \in A\} \leq \inf\{\psi(y) \mid y \in B\}. \tag{2}$$

Indeed, let $x \in A$ and let $y \in B$. Then note that

$$x + y = (x + v) + (y - v)$$

shows us that $x + y \in K$. It follows that

$$\begin{aligned} 0 &\leq \psi(x + y) \\ &= \psi(x) + \psi(y) \end{aligned}$$

which implies $-\psi(x) \leq \psi(y)$, and hence we have (2). We set $\widetilde{\psi}(v)$ to be any number between $\sup\{-\psi(x) \mid x \in A\}$ and $\inf\{\psi(y) \mid y \in B\}$ and we define we define $\widetilde{\psi} \colon W + \mathbb{R}v \to \mathbb{R}$ by

$$\widetilde{\psi}(w + \lambda v) = \psi(w) + \lambda \widetilde{\psi}(v) \tag{3}$$

for all $w + \lambda v \in W + \mathbb{R}v$. Note that (3) is well-defined since v is linearly independent from W. It is easy to check that (3) gives us a linear functional $\widetilde{\psi} \colon W + \mathbb{R}v \to \mathbb{R}$ such that $\widetilde{\psi}|_W = \psi$. Furthermore we have

$$-\psi(x) \leq \widetilde{\psi}(v) \leq \psi(y)$$

for all $x \in A$ and $y \in B$. The only thing left is to check that $\widetilde{\psi}$ satisfies the positivity condition. Let $w + \lambda v \in K \cap (W + \mathbb{R}v)$. We consider the folloiwng cases:

Case 1: Assume that $\lambda > 0$. Then note that

$$(1/\lambda)w + v = (1/\lambda)(w + \lambda v) \in K$$

since K is a convex cone. This implies $(1/\lambda)w \in A$. Thus

$$\begin{aligned} 0 &\leq \lambda(\psi((1/\lambda)w) + \widetilde{\psi}(v)) \\ &= \psi(w) + \lambda\widetilde{\psi}(v) \\ &= \widetilde{\psi}(w + \lambda v). \end{aligned}$$

Case 2: Assume that $\lambda < 0$. Then note that

$$(-1/\lambda)w - v = (-1/\lambda)(w + \lambda v) \in K$$

since K is a convex cone. This implies $(-1/\lambda)w \in B$. Thus

$$\begin{aligned} 0 &\leq -\lambda(\psi((-1/\lambda)w) - \widetilde{\psi}(v)) \\ &= \psi(w) + \lambda\widetilde{\psi}(v) \\ &= \widetilde{\psi}(w + \lambda v). \end{aligned}$$

Case 3: Assume $\lambda = 0$. Then $w \in K \cap W$, and hence

$$\begin{aligned} 0 &\leq \psi(w) \\ &= \widetilde{\psi}(w). \end{aligned}$$

Thus the positivity condition is satisfied.

Now to extend ψ to all of V, we must appeal to Zorn's Lemma. More specifically, we define a partially ordered set $(\mathcal{F}, \leq)$ as follows: the underlying set $\mathcal{F}$ is given by

$$\mathcal{F} = \{\text{linear functionals } \psi' \colon W' \to \mathbb{R} \mid W' \supseteq W,\ \psi'|_W = \psi, \text{ and } \psi'(x) \geq 0 \text{ for all } x \in W' \cap K\}.$$

A member of $\mathcal{F}$ is denoted by an ordered pair: (ψ', W'). If (ψ_1, W_1) and (ψ_2, W_2) are two members of $\mathcal{F}$ then we say $(\psi_1, W_1) \leq (\psi_2, W_2)$ if $W_2 \supseteq W_1$ and $\psi_2|_{W_1} = \psi_1$. Observe that every totally ordered subset in $(\mathcal{F}, \leq)$ has an upper bound. Indeed, suppose $\{(\psi_i, W_i)\}_{i \in I}$ is a totally ordered subset in $(\mathcal{F}, \leq)$. Then if we set $W' = \bigcup_{i \in I} W_i$ and if we define $\psi' \colon W \to \mathbb{R}$ as follows: if $x \in W$, then $x \in W_i$ for some i and we set $\psi'(x) = \psi_i(x)$. Then it is easy to check that (ψ', W') is a member of $\mathcal{F}$ which is an upper bound of $\{(\psi_i, W_i)\}_{i \in I}$. Therefore by Zorn's Lemma, there exists a *maximal* element in $(\mathcal{F}, \leq)$. This maximal element *must* be defined on all of V, otherwise we can extend it to a larger subspace as shown above. □

1.2 Hausdorff Moment Problem

Now we consider $\mathcal{M} = C[0,1]$, $\mathcal{N} = P[0,1]$, and $\mathcal{P} = \{\text{nonnegative continuous functions on } [0,1]\}$. Thus $f \in \mathcal{P}$ if and only if $f(x) \geq 0$ for all $x \in [0,1]$. Clearly $\mathcal{P}$ is a convex cone. For $p \in \mathcal{N}$ we write it as

$$p(x) = b_n x^n + b_{n-1}x^{n-1} + \cdots + b_1 x + b_0,$$

and we define

$$\psi(p) = b_n a_n + b_{n-1}a_{n-1} + \cdots + b_1 a_1 + b_0 a_0.$$

Note that $\psi(x^i) = a_i$. This is clearly a linear functional on $\mathcal{N}$. The first crucial step is to show $\psi(p) \geq 0$ for all $p \in \mathcal{P} \cap \mathcal{N}$. We'll need to use the following theorem of Bernstein:

Theorem 1.2. *(S. Bernstein) A polynomial p is non-negative on $[0,1]$ if and only if it can be represented as*

$$p(x) = A_0 x^n + A_1 x^{n-1}(1-x) + A_2 x^{n-2}(1-x)^2 + \cdots + A_{n-1}x(1-x)^{n-1} + A_n(1-x)^n$$

with $A_0, A_1, \ldots, A_n \geq 0$.

If $\psi(x^i(1-x)^j) \geq 0$ for all $i, j \geq 0$ then by the previous theorem of Bernstein, we will have $\psi(p) \geq 0$ for all $p \in \mathcal{P} \cap \mathcal{N}$. It turns out that this is a sufficient condition too. We write

$$x^i(1-x)^j = x^i \sum_{k=0}^{j} \binom{j}{k}(-1)^k x^k = \sum_{k=0}^{j} \binom{j}{k}(-1)^k x^{i+k}.$$

Thus

$$\begin{aligned}\psi(x^i(1-x)^j) &= \sum_{k=0}^{j} \binom{j}{k} (-1)^k \psi(x^{i+k}) \\ &= \sum_{k=0}^{j} \binom{j}{k} (-1)^k a_{i+k}.\end{aligned}$$

So we need to impose the condition

$$\sum_{k=0}^{j} \binom{j}{k} (-1)^k a_{i+k} \geq 0$$

for all $i, j \geq 0$. Under this condition, we have that all conditions of the Marcel Riesz extension theorem are satisfied, namely we need to check that $\mathcal{M} = \mathcal{P} + \mathcal{N}$. However this is clear: if $f \in \mathcal{M}$, then f is bounded, say $f \leq M$. Then

$$f = (f - M) + M,$$

where $f - M \in \mathcal{P}$ and $M \in \mathcal{N}$. So applying the Marcel Riesz extension theorem, there exists $\widetilde{\psi}\colon \mathcal{M} \to \mathbb{R}$ such that $\widetilde{\psi}(p) = \psi(p)$ for any polynomial p and $\widetilde{\psi}(f) \geq 0$ whenever $f \in \mathcal{P}$. The final important ingredient is the Riesz Representation Theorem:

1.2.1 Riesz Representation Theorem

Lemma 1.3. *(Dini's Theorem) Let X be a compact topological space and let $(f_n\colon X \to \mathbb{R})$ be an increasing sequence of continuous functions which converges pointwise to a continuous function $f\colon X \to \mathbb{R}$. Then (f_n) converges uniformly to f.*

Proof. Let $\varepsilon > 0$. For each $n \in \mathbb{N}$, let $g_n = f - f_n$ and let $E_n = \{g_n < \varepsilon\}$. Each g_n is continuous and thus each E_n is open. Since (f_n) is increasing, each (g_n) is decreasing, and thus the sequence of sets (E_n) is ascending. Since (f_n) converges pointwise to f, it follows that the collection $\{E_n\}$ forms an open cover of X. By compactness of X, we can choose a finite subcover of $\{E_n\}$, and since (E_n) is ascending, this means that there is an $N \in \mathbb{N}$ such that $E_N = X$. Choosing such an N, we see that $n \geq N$ implies

$$\begin{aligned}\varepsilon &> g_n(x) \\ &= f(x) - f_n(x) \\ &= |f(x) - f_n(x)|\end{aligned}$$

for all $x \in X$. It follows that (f_n) converges uniformly to f. □

Theorem 1.4. *(Riesz Representation Theorem) For any linear functional $\ell\colon C[0,1] \to \mathbb{R}$ such that $\ell(f) \geq 0$ for all $f \geq 0$, there exists a unique finite (positive) measure μ on $[0,1]$ such that*

$$\ell(f) = \int_0^1 f \mathrm{d}\mu$$

for all $f \in C[0,1]$.

Proof. Uniqueness is clear. Let's prove existence. Let $B[0,1]$ be the space of all bounded functions $f\colon [0,1] \to \mathbb{R}$. Set $\mathcal{M} = B[0,1]$, $\mathcal{N} = C[0,1]$, and $\mathcal{P} = \{\text{nonnegative bounded functions}\}$. Clearly $\mathcal{P}$ is a convex cone. We also have $\mathcal{M} = \mathcal{P} + \mathcal{N}$ by the same reason as before. Indeed, for any bounded function $f \in \mathcal{M}$ there exists a continuous function $g \in \mathcal{N}$ such that $g \leq f$. Then

$$f = (f - g) + g$$

where $f - g \in \mathcal{P}$ and $g \in \mathcal{N}$. We also know $\ell(f) \geq 0$ for all $f \in \mathcal{P} \cap \mathcal{N}$. So by the Marcel Riesz extension theorem, there exists $\widetilde{\ell}\colon B[0,1] \to \mathbb{R}$ such that $\widetilde{\ell}|_{C[0,1]} = \ell$ and $\widetilde{\ell}|_{\mathcal{P}} \geq 0$. Now we define a measure μ on $\mathcal{B}[0,1]$ by

$$\mu(E) = \widetilde{\ell}(1_E)$$

for each $E \in \mathcal{B}[0,1]$. We next show that μ is a measure. Let (E_n) be a sequence of pairwise disjoint sets in $\mathcal{B}[0,1]$. Then

$$\begin{aligned}\mu\left(\bigcup_{n=1}^{\infty} E_n\right) &= \widetilde{\ell}\left(1_{\bigcup_{n=1}^{\infty} E_n}\right) \\ &=\end{aligned}$$

Observe

$$f_n - f, f - f_n \leq |f_n - f| \leq \|f_n - f\|_{\sup}$$

By the positivity of $\widetilde{\ell}$ we have

$$\widetilde{\ell}(f_n - f), \widetilde{\ell}(f - f_n) \leq \widetilde{\ell}(\|f_n - f\|_{\sup}).$$

Equivalently,

$$|\widetilde{\ell}(f_n - f)| \leq \widetilde{\ell}(\|f_n - f\|_{\sup}) = \|f_n - f\|_{\sup}\widetilde{\ell}(1).$$

Therefore if $f_n \to f$ uniformly. Thus $\widetilde{\ell}$ is continuous with respect to the sup norm.

Now if (f_n) is an increasing sequence which converges pointwise to f, then $f_n \to f$ uniformly (Dini's Theorem). Thus if (f_n) is increasing and converges pointwise to f, then $\widetilde{\ell}(f_n) \to \widetilde{\ell}(f)$. Observe that $(1_{\bigcup_{n=1}^N E_n})$ is increasing and converges pointwise to $1_{\bigcup_{n=1}^\infty E_n}$. It follows that

$$\begin{aligned}
\mu\left(\bigcup_{n=1}^\infty E_n\right) &= \widetilde{\ell}\left(1_{\bigcup_{n=1}^\infty E_n}\right) \\
&= \lim_{N\to\infty} \widetilde{\ell}\left(1_{\bigcup_{n=1}^N E_n}\right) \\
&= \lim_{N\to\infty} \widetilde{\ell}\left(\sum_{n=1}^N 1_{E_n}\right) \\
&= \lim_{N\to\infty} \sum_{n=1}^N \widetilde{\ell}(E_n) \\
&= \lim_{N\to\infty} \sum_{n=1}^N \mu(E_n) \\
&= \sum_{n=1}^\infty \mu(E_n).
\end{aligned}$$

Thus μ is a Borel measure on $[0,1]$. It is finite since $\mu([0,1]) = \widetilde{\ell}(1_{[0,1]}) < \infty$. Let $f \in C[0,1]$. Choose an increasing sequence (φ_n) of simple functions which converges pointwise to f. Then by MCT we have

$$\int_0^1 \varphi_n \mathrm{d}\mu \to \int_0^1 f \mathrm{d}\mu.$$

If $\varphi = \sum_{k=1}^n a_k 1_{A_k}$, then

$$\begin{aligned}
\int_0^1 \varphi \mathrm{d}\mu &= \sum_{k=1}^n a_k \mu(A_k) \\
&= \sum_{k=1}^n a_k \widetilde{\ell}(1_{A_k}) \\
&= \widetilde{\ell}\left(\sum_{k=1}^n a_k 1_{A_k}\right) \\
&= \widetilde{\ell}(\varphi).
\end{aligned}$$

So $\widetilde{\ell}(\varphi_n) \to \widetilde{\ell}(f) = \ell(f)$. We have

$$\int_0^1 \varphi_n \mathrm{d}\mu \to \ell(f)$$

Thus $\widetilde{\ell}(f) = \int f \mathrm{d}\mu$ for any f continuous. □

Another formulation of the Riesz Representation Theorem is given by:

Theorem 1.5. *(Riesz Representation Theorem) For any bounded (with respect to the supremum norm) linear functional $\ell\colon C[0,1] \to \mathbb{R}$ such that $\ell(f) \geq 0$ for all $f \geq 0$, there exists a unique finite (signed) measure μ on $[0,1]$ such that*

$$\ell(f) = \int_0^1 f \mathrm{d}\mu.$$

And a more general version of the Riesz Representation Theorem is given by:

Theorem 1.6. *(Kakutani general version of the Riesz Representation Theorem) Let X be a compact Hausdorff topological space and let $C(X)$ be the Banach space of all continuous functions $f\colon X \to \mathbb{R}$ equipped with the supremum norm:*

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

For any bounded linear functional $\ell\colon C(X) \to \mathbb{R}$ there exists a unique Borel regular measure μ on X such that

$$\ell(f) = \int_X f \mathrm{d}\mu.$$

Let $f \in C[0,1]$. Then f is uniformly continuous. For each $n \in \mathbb{N}$ define a partition

$$0 < x_0^{(n)} < x_1^{(n)} < \cdots < x_n^{(n)} = 1$$

of $[0,1]$ such that none of these points are discontinuities of f and such that

$$|x_{i+1}^{(n)} - x_i^{(n)}| < \frac{2}{n}$$

for all $i = 0, 1, \ldots, n$. Now define $\varphi_n\colon [0,1] \to \mathbb{R}$ by

$$\varphi_n(x) = \sum_{i=0}^{n-1} f(x_i^{(n)}) 1_{(x_i^{(n)}, x_{i+1}^{(n)}]}$$

for all $x \in [0,1]$. Since f is uniformly continuous, we see that (φ_n) converges uniformly to f. Therefore $\widetilde{\ell}(\varphi_n) \to \widetilde{\ell}(f)$ and $\int_0^1 \varphi_n \mathrm{d}\mu \to \int_0^1 f \mathrm{d}\mu$. So it suffices to show

$$\int_0^1 \varphi_n \mathrm{d}\mu = \widetilde{\ell}(\varphi_n).$$

Thus

$$\begin{aligned}
\widetilde{\ell}(\varphi_n) &= \widetilde{\ell}\Big(\sum_{i=0}^{n-1} f(x_i^{(n)}) 1_{(x_i^{(n)}, x_{i+1}^{(n)}]}\Big) \\
&= \sum_{i=0}^{n-1} f(x_i^{(n)}) \widetilde{\ell}\big(1_{(x_i^{(n)}, x_{i+1}^{(n)}]}\big) \\
&= \int_0^1 \varphi_n \mathrm{d}\mu
\end{aligned}$$

for all $n \in \mathbb{N}$.

Theorem 1.7. *(Hausdorff) A sequence (a_n) is a moment sequence of some finite Borel measure μ on $[0,1]$, that is,*

$$a_n = \int_0^1 x^n \mathrm{d}\mu$$

if and only if $(-1)^k (\Delta^k a)_n \geq 0$ for all $k, n \geq 0$ where $(\Delta a)_n = a_{n+1} - a_n$.

We have

$$\begin{aligned}
\Delta^2 a &= \Delta(\Delta a) \\
&= (a_{n+2} - 2a_{n+1} + a_n)_n
\end{aligned}$$

More generally

$$\Delta^k a = \Big(\sum_{i=n}^{n+k} (-1)^i \binom{n}{i} a_{n+i}\Big).$$

Sequences satisfying this condition

$$((-1)^k \Delta^k a)_n \geq 0$$

are called monotone sequences. Observe that

$$(-1)^k (\Delta^k a)_n = \int_0^1 x^n (1-x)^k \mathrm{d}\mu \geq 0.$$

1.3 Hahn-Banach Theorem

Definition 1.2. Let V be an $\mathbb{R}$-vector space. A **partial-seminorm** is a function $p \colon V \to \mathbb{R}$ which satisfies

1. (nonnegativity) $p(x) \geq 0$ for all $x \in V$.
2. (nonnegative homogeneity) $p(\lambda x) = \lambda p(x)$ for all $\lambda \geq 0$ and $x \in V$.
3. (subadditivity) $p(x+y) \leq p(x) + p(y)$ for all $x, y \in V$.

Remark 1. The terminology "partial-seminorm" is made up by me. Recall that a **seminorm** is a function $p \colon V \to \mathbb{R}$ which satisfies

1. (absolute homogeneity) $p(\lambda x) = |\lambda| p(x)$ for all $\lambda \in \mathbb{R}$ and $x \in V$.
2. (subadditivity) $p(x+y) \leq p(x) + p(y)$ for all $x, y \in V$.

It is easy to check that a seminorm is necessarily nonnegative. Thus every seminorm is a partial-seminorm. On the other hand, there are partial-seminorms which are not seminorms.

Theorem 1.8. *Let V be an $\mathbb{R}$-vector space equipped with a partial-seminorm $p \colon V \to \mathbb{R}$ and let U be a subspace of V. Then every linear functional $\varphi \colon U \to \mathbb{R}$ such that $|\varphi| \leq p|_U$ can be extended to a linear functional $\widetilde{\varphi} \colon V \to \mathbb{R}$ such that $\widetilde{\varphi}|_U = \varphi$ and $|\widetilde{\varphi}| \leq p$.*

Remark 2. Note that by $|\varphi| \leq p|_U$, we mean $|\varphi(u)| \leq p(u)$ for all $u \in U$.

Proof. Let $\varphi \colon U \to \mathbb{R}$ be a linear functional such that $|\varphi| \leq p|_U$. We will construct an extension of φ using Marcel Riesz's Extension Theorem. Let

$$P = \{(\lambda, v) \in \mathbb{R} \times V \mid p(v) \leq \lambda\}.$$

Then observe that P is a convex cone contained in the space $\mathbb{R} \times V$. Indeed, if $\alpha > 0$ and $(\lambda, v) \in P$, then $(\alpha\lambda, \alpha v) \in P$ since

$$\begin{aligned} p(\alpha v) &= \alpha p(v) \\ &\leq \alpha\lambda \end{aligned}$$

Also if $(\lambda_1, v_1), (\lambda_2, v_2) \in P$, then $(\lambda_1 + \lambda_2, v_1 + v_2) \in P$ since

$$\begin{aligned} p(v_1 + v_2) &\leq p(v_1) + p(v_2) \\ &= \lambda_1 + \lambda_2. \end{aligned}$$

Furthermore, we have $\mathbb{R} \times V = (\mathbb{R} \times U) + P$, since if $(\lambda, v) \in \mathbb{R} \times V$, then

$$(\lambda, v) = (\lambda - p(v), 0) + (p(v), v)$$

with $(\lambda - p(v), 0) \in \mathbb{R} \times U$ and $(p(v), v) \in P$. Finally define $\psi \colon \mathbb{R} \times U \to \mathbb{R}$ by

$$\psi(\lambda, u) = \lambda - \varphi(u)$$

for all $(\lambda, u) \in \mathbb{R} \times U$. Observe that $\psi|_{(\mathbb{R} \times U) \cap P} \geq 0$. Indeed, if $(\lambda, v) \in (\mathbb{R} \times U) \cap P$, then

$$\begin{aligned} \psi(\lambda, v) &= \lambda - \varphi(v) \\ &\geq \lambda - p(v) \\ &\geq 0 \end{aligned}$$

Thus we have all of the ingredients to apply the Marcel Riesz Extension Theorem: choose $\widetilde{\psi} \colon \mathbb{R} \times V \to \mathbb{R}$ such that $\widetilde{\psi}|_{\mathbb{R} \times U} = \psi$ and $\widetilde{\psi}|_P \geq 0$. Define $\widetilde{\varphi} \colon V \to \mathbb{R}$ by

$$\widetilde{\varphi}(v) = -\widetilde{\psi}(0, v)$$

for all $v \in V$. Note that if $u \in U$, then

$$\begin{aligned} \widetilde{\varphi}(u) &= -\widetilde{\psi}(0, u) \\ &= -\psi(0, u) \\ &= \varphi(u). \end{aligned}$$

Thus $\widetilde{\varphi}|_U = \varphi$. We claim $|\widetilde{\varphi}| \leq p$. To see this, assume for a contradiction that $v_0 \in V$ such that

$$\widetilde{\varphi}(v_0) > p(v_0).$$

Then using that $(p(x_0), x_0) \in P$, we have

$$\begin{aligned}0 &\leq \widetilde{\psi}(p(x_0), x_0) \\ &= \widetilde{\psi}(0, x_0) + \widetilde{\psi}(p(x_0), 0) \\ &= -\widetilde{\varphi}(x_0) + \psi(p(x_0), 0) \\ &= -\widetilde{\varphi}(x_0) + p(x_0) \\ &< -p(x_0) + p(x_0) \\ &= 0,\end{aligned}$$

which is a contradiction. This establishes our claim and we are done. □

In the setting of normed linear spaces, the Hahn-Banach Theorem says that any linear functional ℓ defined on a subspace $\mathcal{Y} \subseteq \mathcal{X}$ which is bounded on $\mathcal{Y}$ can be extended to a bounded linear functional $\widetilde{\ell}$ on $\mathcal{X}$ such that $\widetilde{\ell}|_{\mathcal{Y}} = \ell$ and $\|\widetilde{\ell}\|_{\mathcal{X}} = \|\ell\|_{\mathcal{Y}}$. This is an immediate consequence of our more general version that we have just proved.

Proposition 1.1. *Let $\mathcal{X}$ be a normed linear space and let x_0 be a nonzero vector in $\mathcal{X}$. Then there exists a bounded linear functional $\ell \colon \mathcal{X} \to \mathbb{R}$ with $\|\ell\| = 1$ such that $\ell(x_0) = \|x_0\|$.*

So if you have two points $a \neq b$ in $\mathcal{X}$, then there exists a bounded linear functional $\ell \in \mathcal{X}^*$ such that $\ell(a) \neq \ell(b)$.

Theorem 1.9. *Let $\mathcal{X}$ be a reflexive Banach space and let $\mathcal{Y}$ be a closed subspace of $\mathcal{X}$. Then for every $x \in \mathcal{X}$ there exists $y_0 \in \mathcal{Y}$ such that $\mathrm{d}(x, \mathcal{Y}) = \|x - y_0\|$.*

Remark 3. We can replace $\mathcal{Y}$ with a convex set.

Proof. Define a function $\varphi \colon \mathcal{Y} \to \mathbb{R}$ by

$$\varphi(y) = \|y - x\|$$

for all $y \in \mathcal{Y}$. □

2 Geometric Form of the Hahn-Banach Theorem

2.1 Gauge Functional

Definition 2.1. Let V be an $\mathbb{R}$-vector space and let S be a subset of V. A point $x \in S$ is said to be an **interior point** of S if for any $y \in V$, there exists $\varepsilon_{x,y} > 0$ such that $|t| < \varepsilon_{x,y}$ implies $x + ty \in S$. We denote by $\operatorname{int} S$ to be the set of all interior points of S. Note that if $0 \in \operatorname{int} S$, then $0 \in S$. Indeed, assuming $0 \in \operatorname{int} S$, then there exists $\varepsilon_{0,0} > 0$ such that $|t| < \varepsilon_{0,0}$ implies $0 = 0 + t \cdot 0 \in S$. The converse of course isn't true (take $S = \{0\}$).

Remark 4. Here we write $\varepsilon_{x,y}$ to emphasize that $\varepsilon_{x,y}$ *depends* on x and y. Usually we will just write ε instead of $\varepsilon_{x,y}$.

Definition 2.2. Let V be an $\mathbb{R}$-vector space and let $C \subseteq V$ be a convex set with 0 as an interior point. Define $\mathrm{p}_C \colon V \to \mathbb{R}$ by

$$\mathrm{p}_C(x) = \inf\{\alpha > 0 \mid (1/\alpha)x \in C\}$$

for all $x \in V$. This is called the **gauge functional** of C.

Example 2.1. Let $(V, \|\cdot\|)$ be a normed linear space and let $C = \mathrm{B}_1[0]$ be the closed unit ball centered at 0 with radius 1. Then $\mathrm{p}_C(x) = \|x\|$ for all $x \in V$.

2.1.1 Gauge Functional is a Partial-Seminorm

Proposition 2.1. *Let V be an $\mathbb{R}$-vector space and let $C \subseteq V$ be a convex set with 0 as an interior point. Then the gauge functional p_C is a partial-seminorm.*

Proof. We first show p_C is subadditive. Let $\varepsilon > 0$ and let $x, y \in V$. Set $a = \mathrm{p}_C(x) + \varepsilon/2$ and set $b = \mathrm{p}_C(y) + \varepsilon/2$. Then $a, b > 0$ and $(1/a)x, (1/b)x \in C$. Since C is convex, we see that

$$\frac{1}{a+b}(x+y) = \frac{a}{a+b}\left(\frac{1}{a}x\right) + \frac{b}{a+b}\left(\frac{1}{b}y\right) \in C.$$

It follows that

$$\begin{aligned}\mathrm{p}_C(x) + \mathrm{p}_C(y) + \varepsilon &= a + b \\ &\geq \mathrm{p}_C(x+y).\end{aligned}$$

Taking $\varepsilon \to 0$ shows that p_C is subadditive.

Next we show that p_C satisfies nonnegative homogeneity. Let $\lambda \geq 0$ and let $x \in V$. First note that if $\lambda = 0$, then since

$$p_C(0) = \inf\{\alpha > 0 \mid (1/\alpha) \cdot 0 \in C\} = 0,$$

we have $0 = 0 \cdot p_C(x) = p_C(0 \cdot x)$. Thus we may assume $\lambda > 0$. Then

$$\begin{aligned} p_C(\lambda x) &= \inf\{\alpha > 0 \mid (1/\alpha)\lambda x \in C\} \\ &= \lambda \inf\{\alpha > 0 \mid (1/\alpha) x \in C\} \\ &= \lambda p_C(x). \end{aligned}$$

Finally note that p_C is nonnegative by definition. Thus p_C is a partial-seminorm. □

2.1.2 Properties of Gauge Functional

Proposition 2.2. *Let V be an $\mathbb{R}$-vector space and let $C \subseteq V$ be a convex set with 0 as an interior point. We have*

1. $C \subseteq \{p_C \leq 1\}$.
2. $\operatorname{int} C = \{p_C < 1\}$.

Proof. 1. Let $x \in C$. Then $(1/1)x \in C$ and hence $p_C(x) \leq 1$.

2. Let $x \in \operatorname{int} C$. Then there exists $\varepsilon > 0$ such that $x + \varepsilon x \in C$. So

$$\begin{aligned} x + \varepsilon x &= (1+\varepsilon)x \\ &= \frac{1}{1/(1+\varepsilon)} x \end{aligned}$$

shows $p_C(x) \leq 1/(1+\varepsilon) < 1$. Conversely, let $x \in V$ such that $p_C(x) < 1$. Then there exists $0 < \alpha < 1$ such that $(1/\alpha)x \in C$. Now let $y \in V$. Since $0 \in \operatorname{int}(C)$, there exists $\varepsilon > 0$ such that $|t| < \varepsilon$ implies $ty \in C$. Then $|t| < \varepsilon$ implies

$$x + (1-\alpha)ty = \alpha(1/\alpha)x + (1-\alpha)ty \in C$$

since C is convex. In particular, setting $\delta = (1-\alpha)\varepsilon$, we see that $|t| < \delta$ implies $x + ty \in C$. □

2.1.3 Gauge Functional Induced from Partial-Seminorm

Recall from Proposition (2.1) that is C is a convex subset of a real vector space V such that $0 \in \operatorname{int} C$, then the gauge functional $p_C \colon V \to \mathbb{R}$ is a partial-seminorm. We will now show a converse to this.

Proposition 2.3. *Let V be an $\mathbb{R}$-vector space, let $p \colon V \to \mathbb{R}$ be a partial-seminorm, and set $C = \{p \leq 1\}$. Then C is a convex set, and moreover, we have $p_C = p$.*

Proof. Let $x, y \in C$ and $\alpha \in [0,1]$. Then

$$\begin{aligned} p((1-\alpha)x + \alpha y) &\leq p((1-\alpha)x) + p(\alpha y) \\ &= (1-\alpha)p(x) + \alpha p(y) \\ &\leq (1-\alpha) + \alpha \\ &= 1 \end{aligned}$$

implies $(1-\alpha)x + \alpha y \in C$. Thus C is a convex set.

Now assume there exists $x_0 \in V$ such that $p_C(x_0) < p(x_0)$. Then there exists $\alpha \in \mathbb{R}$ such that

$$p_C(x_0) \leq \alpha < p(x_0)$$

and such that $(1/\alpha)x_0 \in C$. Then $p((1/\alpha)x_0) \leq 1$ which is equivalent to $(1/\alpha)p(x_0) \leq 1$ which implies $p(x_0) \leq \alpha$. This is a contradiction. So $p_C(x) \geq p(x)$ for all $x \in V$. Now assume there exists $x_0 \in V$ such that $p(x_0) < p_C(x_0)$. Then there exists $\alpha \in \mathbb{R}$ such that

$$p(x_0) \leq \alpha < p_C(x_0).$$

Then $(1/\alpha)p(x_0) \leq 1$. In other words, $p((1/\alpha)x_0) \leq 1$ which is equivalent to $(1/\alpha)x_0 \in C$. This contradicts the fact that $p_C(x_0)$ is the infimum of all such $\alpha > 0$. Therefore $p(x) \geq p_C(x)$ for all $x \in V$. It follows that $p = p_C$. □

Theorem 2.1. *Let V be an $\mathbb{R}$-vector space and let C be a nonempty convex subset of V such that $C = \operatorname{int} C$. Then for any $y \notin C$, there exists a hyperplane $\{\ell = \alpha\}$ where $\ell \colon V \to \mathbb{R}$ is some linear functional and $\alpha \in \mathbb{R}$ such that $y \in \{\ell = \alpha\}$ and $C \subseteq \{\ell < \alpha\}$.*

Proof. By translating if necessary, we may assume that $0 \in \operatorname{int} C$. This means it is possible to define the gauge potential p_C of C. Define $\ell \colon \mathbb{R}y \to \mathbb{R}$ by $\ell(ay) = a$ for all $ay \in \mathbb{R}y$. Notice if $a < 0$, then

$$\begin{aligned}\ell(ay) &= a \\ &< 0 \\ &\leq \mathrm{p}_C(ay),\end{aligned}$$

and if $a > 0$, then

$$\begin{aligned}\ell(ay) &= a \\ &\leq a\mathrm{p}_C(y) \\ &= \mathrm{p}_C(ay),\end{aligned}$$

where we used the fact that $\mathrm{p}_C(y) \geq 1$ since $y \notin \operatorname{int} C = C$. So we see that $\ell \leq \mathrm{p}_C|_{\mathbb{R}y}$. Therefore by the Hahn-Banach Theorem, we can extend ℓ to $\widetilde{\ell} \colon V \to \mathbb{R}$ such that $\widetilde{\ell}|_{\mathbb{R}y} = \ell$ and $\widetilde{\ell} \leq \mathrm{p}_C$. In particular, if $x \in C$, then

$$\widetilde{\ell}(x) \leq \mathrm{p}_C(x) < 1.$$

Thus $C \subseteq \{\widetilde{\ell} < \alpha\}$ where $\alpha = 1$. Also clearly $\widetilde{\ell}(y) = 1$, and so we are done. □

2.1.4 First Geometric Form of Hahn-Banach

Theorem 2.2. *(first geometric form of Hahn-Banach) Let V be an $\mathbb{R}$-vector space and let $A, B \subseteq V$ be nonempty convex sets such that $A \cap B = \emptyset$. Suppose A satisfies $A = \operatorname{int} A$. Then there exists a hyperplane that separates A and B. More precisely, there exists a linear functional $\ell \colon V \to \mathbb{R}$ and $\alpha \in \mathbb{R}$ such that $A \subseteq \{\ell \leq \alpha\}$ and $B \subseteq \{\ell \geq \alpha\}$.*

Proof. Set $C = A - B = \{a - b \mid a \in A, b \in B\}$. It's easy to see that C is a nonemtpy convex set. It's also easy to see that $\operatorname{int} C = C$. Indeed,

$$C = \bigcup_{b \in B} A - \{b\}.$$

Also $0 \notin C$ since A and B are disjoint from one another. By the previous result, there exists a linear functional $\ell \colon V \to \mathbb{R}$ such that $0 \in \{\ell = 0\}$ and $C \subseteq \{\ell < 0\}$.

Now let $a \in A$ and $b \in B$. Since $a - b \in C$, we have

$$\begin{aligned}0 &> \ell(a - b) \\ &= \ell(a) - \ell(b),\end{aligned}$$

that is, $\ell(a) < \ell(b)$. Therefore

$$\sup\{\ell(a) \mid a \in A\} \leq \inf\{\ell(b) \mid b \in B\}.$$

So choose α between $\sup\{\ell(a) \mid a \in A\}$ and $\inf\{\ell(b) \mid b \in B\}$. Then $A \subseteq \{\ell \leq \alpha\}$ and $B \subseteq \{\ell \geq \alpha\}$. □

2.1.5 Second Geometric Form of Hahn-Banach

Theorem 2.3. *(second geometric form of Hahn-Banach) Let $\mathcal{X}$ be a normed linear space and let $A, B \subseteq \mathcal{X}$ be two nonempty convex sets such that $A \cap B = \emptyset$. Suppose A is closed and B is compact. Then there exists a closed hyperplane that strictly separates A and B. More precisely, there exists a bounded linear functional $\ell \colon V \to \mathbb{R}$ and $\alpha \in \mathbb{R}$ such that $A \subseteq \{\ell < \alpha\}$ and $B \subseteq \{\ell > \alpha\}$.*

Proof. Set $C = A - B = \{a - b \mid a \in A, b \in B\}$. It's easy to see that C is a nonemtpy convex set. It's also easy to see that C is closed. Also $0 \notin C$ since A and B are disjoint from one another. Then C^c is open so there exists $r > 0$ such that $\mathrm{B}_r(0) \subseteq C^c$, that is, $\mathrm{B}_r(0) \cap C = \emptyset$. By the previous first geometric form of Hahn-Banach, we can separate $\mathrm{B}_r(0)$ and C by a hyperplane, say $\{\ell = \alpha\}$. Then $\ell(a - b) \leq \ell(rx)$ for all $a \in A$, $b \in B$ and $x \in \mathrm{B}_1(0)$. It can be shown that $\ell \colon \mathcal{X} \to \mathbb{R}$ is bounded. Therefore

$$\ell(a - b) \leq \inf\{\ell(rx) \mid x \in \mathrm{B}_1(0)\} = -r\|\ell\|.$$

Now take $\varepsilon = (1/2)r\|\ell\| > 0$. Then

$$\ell(a) + \varepsilon \leq \ell(b) - \varepsilon$$

for all $a \in A$ and $b \in B$. This implies

$$\sup\{\ell(a) \mid a \in A\} < \inf\{\ell(b) \mid b \in B\}.$$

So choose α strictly between $\sup\{\ell(a) \mid a \in A\}$ and $\inf\{\ell(b) \mid b \in B\}$. Then $A \subseteq \{\ell < \alpha\}$ and $B \subseteq \{\ell > \alpha\}$. □

Part II
Homework

Part III
Appendix