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 Sets

Proof. Let V be an $\mathbb{R}$-vector space and let C be a subset of V. We say C is **convex** if for all $t \in (0,1)$ and $x, y \in S$, we have $tx + (1-t)y \in C$. □

Proposition 1.1. *Let V be an $\mathbb{R}$-vector space and let C be a convex subset of V. Then for all $n \in \mathbb{N}$, $x_1, \dots, x_n \in C$, and $t_1, \dots, t_n \in (0,1)$ such that $\sum_{i=1}^n t_i = 1$, we have $\sum_{i=1}^n t_i x_i \in C$.*

Proof. Let $x = \sum_{i=1}^n t_i x_i$ and assume that n is minimal in the sense that if $x = \sum_{i'=1}^{n'} t'_{i'} x_{i'}$ is another representation of x, where each $x_{i'} \in S$ and $t'_{i'} \in (0,1)$ such that $\sum_{i'=1}^{n'} t'_{i'} = 1$, then we must have $n \leq n'$. Assume for a contradiction that $x \notin C$, so $n > 2$. Then observe that

$$\begin{aligned}
\sum_{i=1}^{n-1} \frac{t_i}{1-t_n}((1-t_n)x_i + t_n x_n) &= \sum_{i=1}^{n-1} t_i x_i + \left(\sum_{i=1}^{n-1} \frac{t_i}{1-t_n} \right) t_n x_n \\
&= \sum_{i=1}^{n-1} t_i x_i + \left(\frac{1-t_n}{1-t_n} \right) t_n x_n \\
&= \sum_{i=1}^{n-1} t_i x_i + t_n x_n \\
&= \sum_{i=1}^{n} t_i x_i \\
&= x,
\end{aligned}$$

gives another representation with $n-1$ terms, a contradiction. □

1.1.1 Convex Closure and Closed Convex Closure

Definition 1.1. Let V be an $\mathbb{R}$-vector space and let S be a subset of V. The **convex closure** of S is defined by

$$\mathrm{conv}(S) = \{ tx + (1-t)y \mid t \in (0,1) \text{ and } x, y \in S \}.$$

Moreover, suppose $\|\cdot\|$ is a norm on V, so that $(V, \|\cdot\|)$ is a normed linear space. The **closed convex closure** of S is defined to be the smallest closed convex set which contains S and is denoted by $\overline{\mathrm{conv}}(S)$.

Proposition 1.2. *With the notation as above,* $\mathrm{conv}(S)$ *is the smallest convex set which contains S. Furthermore, we have* $\overline{\mathrm{conv}(S)} = \overline{\mathrm{conv}}(S)$.

Proof. Let us first show that $\mathrm{conv}(S)$ is in fact a convex set. Let $s, t, t' \in (0,1)$ and let $x, x', y, y' \in S$. Then observe that

$$s(tx + (1-t)y) + (1-s)(t'x' + (1-t')y') = stx + s(1-t)y + (1-s)t'x' + (1-s)(1-t')y' \in \mathrm{conv}(S),$$

where we used Proposition (1.1) together with the fact that

$$st + s(1-t) + (1-s)t' + (1-s)(1-t') = 1.$$

It follows that $\mathrm{conv}(S)$ is convex. It is also the smallest convex set which contains S since if C is a convex set which contains S, then we must have $tx + (1-t)y \in C$ for all $t \in (0,1)$ and $x, y \in S$, which implies $\mathrm{conv}(S) \subseteq C$.

Now we will show $\overline{\mathrm{conv}(S)} = \overline{\mathrm{conv}}(S)$. To see this, first note that since $\overline{\mathrm{conv}}(S)$ is convex, we have $\mathrm{conv}(S) \subseteq \overline{\mathrm{conv}}(S)$, and hence

$$\begin{aligned} \overline{\mathrm{conv}(S)} &\subseteq \overline{\overline{\mathrm{conv}}(S)} \\ &= \overline{\mathrm{conv}}(S). \end{aligned}$$

For the reverse inclusion, it suffices to show that $\overline{\mathrm{conv}(S)}$ is convex, since then $\overline{\mathrm{conv}(S)}$ would be a closed convex set, and so $\overline{\mathrm{conv}}(S) \subseteq \overline{\mathrm{conv}(S)}$ by definition of $\overline{\mathrm{conv}}(S)$. In fact, we will show that the closure of a convex set is convex. To this end, suppose C is a convex set and let $t \in (0,1)$ and $x, y \in \overline{C}$. Choose sequences (x_n) and (y_n) in C such that $x_n \to x$ and $y_n \to y$. Then $(tx_n + (1-t)y_n)$ is a sequence in C (as C is convex) which converges to $tx + (1-t)y$. It follows that $tx + (1-t)y \in \overline{C}$, and hence $\overline{C}$ is convex. □

1.1.2 Convex Closure Preserves Minkowski Sum

Definition 1.2. Let V be an $\mathbb{R}$-vector space and let S_1, S_2 be subsets of V. We define the **Minkowski sum** of S_1 and S_2 to be the set

$$S_1 + S_2 = \{x_1 + x_2 \mid x_1 \in S_1 \text{ and } x_2 \in S_2\}.$$

Proposition 1.3. *Let V be an $\mathbb{R}$-vector space and let C_1, C_2 be convex subsets of V. Then $C_1 + C_2$ is convex.*

Proof. Let $t \in (0,1)$, let $c_1, c_1' \in C_1$, and let $c_2, c_2' \in C_2$. Then we have

$$t(c_1 + c_2) + (1-t)(c_1' + c_2') = (tc_1 + (1-t)c_1') + (tc_2 + (1-t)c_2') \in C_1 + C_2,$$

since both C_1 and C_2 are convex. It follows that $C_1 + C_2$ is convex. □

Proposition 1.4. *Let V be an $\mathbb{R}$-vector space and let S_1, S_2 be subsets of V. Then we have* $\mathrm{conv}(S_1 + S_2) = \mathrm{conv}(S_1) + \mathrm{conv}(S_2)$.

Proof. Note that $\mathrm{conv}(S_1) + \mathrm{conv}(S_2)$ is a convex set which contains $S_1 + S_2$. Thus

$$\mathrm{conv}(S_1 + S_2) \subseteq \mathrm{conv}(S_1) + \mathrm{conv}(S_2).$$

For the reverse inclusion, let $z_1 \in \mathrm{conv}(S_1)$ and $z_2 \in \mathrm{conv}(S_2)$ and express them as $z_1 = t_1 x_1 + (1-t_1)y_1$ and $z_2 = t_2 x_2 + (1-t_1)y_2$ where $x_1, y_1 \in S_1$, $x_2, y_2 \in S_2$, and $t_1, t_2 \in (0,1)$. Then note that

$$\begin{aligned} z_1 + z_2 &= t_1x_1 + (1-t_1)y_1 + t_2x_2 + (1-t_2)y_2 \\ &= t_1x_1 + t_2x_2 + y_1 - t_1y_1 + y_2 - t_2y_2 \\ &= (t_1 - t_2)(x_1 + y_2) + t_2(x_1 + x_2) + (1-t_1)(y_1 + y_2), \end{aligned}$$

where $(t_1 - t_2) + t_2 + (1 - t_1) = 1$ and where $x_1 + y_2, x_1 + x_2, y_1 + y_2 \in S_1 + S_2$. It follows that $z_1 + z_2 \in \mathrm{conv}(S_1 + S_2)$. Thus we have the reverse inclusion

$$\mathrm{conv}(S_1 + S_2) \supseteq \mathrm{conv}(S_1) + \mathrm{conv}(S_2).$$

□

1.2 Convex Cones

Definition 1.3. Let V be an $\mathbb{R}$-vector space. A set $P \subseteq V$ is said to be a **convex cone** if

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

Given a convex cone $P \subseteq V$, we can define a partial order on V as follows: if $x, y \in V$, then we say $x \leq_P y$ if $y - x \in P$. To see that this is a preorder, note that reflexivity of $\leq_P$ follows from the fact that $0 \in P$. Transitivity of $\leq_P$ follows from the fact that P is closed under addition: if $x \leq_P y$ and $y \leq_P z$, then $z - x = (z-y) + (y-x) \in P$ shows $x \leq_P z$. Thus $\leq_P$ is in fact a preorder. If we assume in addition that $-P \cap P = 0$, then we also have antisymmetry of $\leq_P$. In this case, $\leq_P$ is a partial order. Note that, we will have $0 \leq_P x$ for all $x \in P$, thus it makes sense to call the elements of P the **positive** elements with respect to the preorder $\leq_P$.

1.3 Marcel Riesz Extension Theorem

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

Proof. Let $v \in V \backslash W$. We will first show that we can extend ψ to a linear functional $\widetilde{\psi}\colon W + \mathbb{R}v \to \mathbb{R}$ such that $\widetilde{\psi}$ preserves the positivity condition. Define two sets $A = \{x \in W \mid -x \leq_P v\}$ and $B = \{y \in W \mid v \leq_P y\}$. Note that A and B are nonempty since $V = W + P$. 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 $-x \leq_P v \leq_P y$ implies $x + y \in C$. It follows that

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

In other words, $-\psi(x) \leq \psi(y)$, which implies (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 $w + \lambda v \in P \cap (W + \mathbb{R}v)$. We consider the following cases:

Case 1: Assume that $\lambda > 0$. Then note that $(1/\lambda)w + v = (1/\lambda)(w + \lambda v) \in P$ since P 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 $(-1/\lambda)w - v = (-1/\lambda)(w + \lambda v) \in P$ since P 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 that $\lambda = 0$. Then $w \in P \cap W$, and hence $0 \leq \psi(w) = \widetilde{\psi}(w)$.

In all three cases, we see that the positivity condition is satisfied. Thus we can extend ψ to a linear functional on $W + \mathbb{R}v$ while preserving the positivity condition.

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

$$\mathcal{F} = \{\text{linear functionals } \psi'\colon W' \to \mathbb{R} \mid W' \supseteq W,\ \psi'|_W = \psi, \text{ and } \psi'|_{W'\cap C = P} \geq 0\}.$$

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 $W_1 \subseteq W_2$ 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 $\mathcal{F}$ and that it is an upper bound of $\{(\psi_i, W_i)\}_{i\in I}$. Since $\mathcal{F}$ is nonempty (it contains (ψ, W)) and every totally ordered subset of $\mathcal{F}$ has an upper bound, we can apply Zorn's Lemma to obtain 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 and obtain a contradiction. $\square$

1.4 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:

1.4.1 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. $\square$

Theorem 1.4. *(Riesz Representation Theorem) Let $\ell\colon C[0,1] \to \mathbb{R}$ be a linear functional such that $\ell(f) \geq 0$ for all $f \geq 0$. Then 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 $f\colon [0,1] \to \mathbb{R}$ and let $N[0,1]$ be the space of all nonnegative bounded functions $f\colon [0,1] \to \mathbb{R}$. Clearly $B[0,1]$ contains $C[0,1]$ as subspace and it is easy to see that $B[0,1] = C[0,1] + N[0,1]$. Indeed, for any bounded function $f \in B[0,1]$ there exists a continuous function $g \in C[0,1]$ such that $g \leq f$. Then

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

where $f - g \in N[0,1]$ and $g \in C[0,1]$. Furthermore, $N[0,1]$ is a convex cone and by assumption we have $\ell(f) \geq 0$ for all $f \in C[0,1] \cap N[0,1]$. So by the Marcel Riesz extension theorem, there exists a linear functional $\widetilde{\ell}\colon B[0,1] \to \mathbb{R}$ such that $\widetilde{\ell}|_{C[0,1]} = \ell$ and $\widetilde{\ell}|_{N[0,1]} \geq 0$. 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

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

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

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

Sequences satisfying this condition

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

are called monotone sequences. Observe that

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

1.5 Hahn-Banach Theorem

Definition 1.4. Let V be an $\mathbb{R}$-vector space. A **partial-seminorm** is a function $p\colon V \to \mathbb{R}$ which satisfies

1. (nonnegativity) $p \geq 0$, that is, $p(x) \geq 0$ 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. Indeed, consider the function $p\colon \mathbb{R} \to \mathbb{R}$ defined by

$$p(x) = \begin{cases} x & \text{if } x \geq 0 \\ -x/2 & \text{if } x < 0 \end{cases}$$

for all $x \in \mathbb{R}$. It is easy to check that p is a partial-seminorm which is not a seminorm.

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 1.5. *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 **internal point** of S if for any $y \in V$, there exists $\varepsilon_{x,y} > 0$ such that $|t| < \varepsilon_{x,y}$ implies $x + ty \in S$. The set of all points internal points of S is called the **core** of S and is denoted by $\operatorname{core} S$.

Remark 4. Let us make several remarks about this definition.

1. We write $\varepsilon_{x,y}$ to emphasize that $\varepsilon_{x,y}$ *depends* on x and y. Usually we will just write ε instead of $\varepsilon_{x,y}$.
2. Note that if $0 \in \operatorname{core} S$, then $0 \in S$. Indeed, assuming $0 \in \operatorname{core} S$, 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 $S = \{0\}$).
3. Suppose V is equipped with a metric. Recall that a point $x \in S$ is said to be an **interior point** of S if there exists an $\varepsilon > 0$ such that $\mathrm{B}_\varepsilon(x) \subseteq S$. The set of all interior points of S is denoted $\operatorname{int} S$. It is easy to see that every interior point of S is an internal point of S. Thus $\operatorname{int} S \subseteq \operatorname{core} S$. If S happens to be open, then $S = \operatorname{int} S \subseteq \operatorname{core} S = S$ which forces $\operatorname{int} S = \operatorname{core} S$.

Definition 2.2. Let V be an $\mathbb{R}$-vector space and let $C \subseteq V$ be a convex set with 0 as an internal point. We define the **gauge functional** of C to be the function $\mathrm{p}_C \colon V \to \mathbb{R}$ given by

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

for all $x \in V$.

Note that 0 be an internal point of C guarantees that $\mathrm{p}_C(x) < \infty$. Indeed, since 0 is an internal point, there exists an $\varepsilon > 0$ such that $tx \in C$ for all $|t| < \varepsilon$. In particular, if $\alpha > 1/\varepsilon$, then $1/\alpha < \varepsilon$, and hence $(1/\alpha)x \in C$. Thus we see that $\mathrm{p}_C(x) \leq 1/\varepsilon$. Thus having 0 be an internal point of C guarantees that $\mathrm{p}_C(x) < \infty$.

Example 2.1. Let $\mathcal{X}$ be a normed linear space. Then $\mathrm{p}_{\mathrm{B}_1[0]}(x) = \|x\|$ for all $x \in \mathcal{X}$.

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 internal 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

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

we have $0 = 0 \cdot \mathrm{p}_C(x) = \mathrm{p}_C(0 \cdot x)$. Thus we may assume $\lambda > 0$. Then

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

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 internal point. We have*

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

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

2. Let $x \in \operatorname{core} C$. 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 $0 \in \operatorname{core}(C)$, 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 $0 \in \operatorname{core} C$, 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 $C = \operatorname{core} C$. 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

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

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 $y \notin \operatorname{core} C = C$. 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 $A = \operatorname{core} A$. 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 $C = A - B = \{a - b \mid a \in A, b \in B\}$. Then C is a nonempty convex set. Furthermore we have $\operatorname{int} C = C$. Indeed, let $a - b \in C$ and let $y \in V$. Choose $\varepsilon > 0$ such that $|t| < \varepsilon$ implies $a + ty \in A$. Then $|t| < \varepsilon$ implies $a - b + ty = (a + ty) - b \in C$. Finally note that $0 \notin C$ since A and B are disjoint from one another. By the previous result, there exists a linear functional $\ell\colon V \to \mathbb{R}$ and an $\beta \in \mathbb{R}$ such that $0 \in \{\ell = \beta\}$ and $C \subseteq \{\ell < \beta\}$. Note that since $\ell(0) = \beta$, we must necessarily have $\beta = 0$.

Now let $a \in A$ and $b \in B$. Since $a - b \in C$, we have $0 > \ell(a - b) = \ell(a) - \ell(b)$, 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

Lemma 2.3. *Let $\mathcal{X}$ be a normed linear space, let A be a closed subset of $\mathcal{X}$, and let B be a compact subset of $\mathcal{X}$. Then $A + B$ is closed.*

Proof. Let $x \in \overline{A + B}$ and choose a sequence $(a_n + b_n)$ in $A + B$ such that $a_n + b_n \to x$. Since B is compact, there exist a convergent subsequence of (b_n), say $(b_{\pi(n)})$. In fact, by relabeling indices if necessary, we may assume that (b_n) is convergent, say $b_n \to b$ where $b \in B$. Now since $a_n + b_n \to x$ and $b_n \to b$, it follows easily that $a_n \to x - b$. Since A is closed, we must have $x - b \in A$. Thus $x = (x - b) + b$ shows $x \in A + B$, which implies $A + B = \overline{A + B}$, hence $A + B$ is closed. □

Theorem 2.4. *(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 $C = A - B = \{a - b \mid a \in A, b \in B\}$. Then C is a nonempty convex set. Furthermore, C is closed by Lemma (2.3) since $-B$ is compact and $A - B = A + (-B)$. Also $0 \notin C$ since A and B are disjoint from one another. Thus C^c is open and contains 0, which means there exists $r > 0$ such that $\mathrm{B}_r(0) \subseteq C^c$. In other words, $\mathrm{B}_r(0) \cap C = \emptyset$. 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 $B \subseteq \{\ell > \alpha\}$. □

2.2 Lower Semicontinuity

Definition 2.3. Let $\mathcal{X}$ be a normed linear space. A function $\varphi\colon \mathcal{X} \to (-\infty, \infty]$ is said to be **lower semicontinuous** if for every $c \in \mathbb{R}$ the set $\{\varphi \leq c\}$ is closed.

Here are some basic facts:

1. φ is lower semicontinuous if and only if $\{(x, \lambda) \mid \varphi(x) \leq \lambda\}$ is a closed set in $\mathcal{X} \times \mathbb{R}$ for every $\lambda \in \mathbb{R}$.
2. φ_1 and φ_2 are lower semicontinuous implies $\varphi_1 + \varphi_2$ is lower semicontinuous.
3. $\{\varphi_i\}_{i \in I}$ is a collection of lower semicontinuous functions, then $\sup_{i \in I} \varphi_i$ is also lower semicontinuous.
4. if $K \subseteq \mathcal{X}$ is compact, then $\inf_{x \in \mathcal{K}} \varphi(x)$ is acheived.

2.3 Convexity

Definition 2.4. Let $\mathcal{X}$ be a normed linear space. A function $\varphi\colon \mathcal{X} \to (-\infty, \infty]$ is said to be **convex** if

$$\varphi(tx + (1-t)y) \leq t\varphi(x) + (1-t)\varphi(y)$$

for all $x, y \in \mathcal{X}$ and $t \in [0, 1]$.

Here are some basic facts:

1. φ is convex if and only if $\operatorname{epi}(\varphi) = \{(x, \lambda) \mid \varphi(x) \leq \lambda\}$ is a convex set in $\mathcal{X} \times \mathbb{R}$.
2. If φ_1 and φ_2 are convex, then $\varphi_1 + \varphi_2$ is convex.
3. If $\{\varphi_i\}_{i \in I}$ are all convex, then $\sup_{i \in I} \varphi_i$ is convex.
4. If φ is convex, then $\{\varphi \leq c\}$ is a convex set for all $c \in \mathbb{R}$. The converse is not true in general.

We usually assume both convexity and lower semicontinuity in optimization problems.

2.3.1 Conjugate Function

Definition 2.5. Let $\mathcal{X}$ be a normed linear space and let $\varphi\colon \mathcal{X} \to (-\infty, \infty]$ be a function such that $\varphi \neq \infty$.

1. We define the **conjugate function** of φ to be the function $\varphi^*\colon \mathcal{X}^* \to (-\infty, \infty]$ defined by

   $$\varphi^*(\ell) = \sup_{x \in \mathcal{X}} (\ell(x) - \varphi(x))$$

   for all $\ell \in \mathcal{X}^*$. The conjugate function φ^* is sometimes called a **Fenchel transform** of φ or a **Legendre transform** of φ.

2. We define the **double conjugate function** of φ to be the function $\varphi^{**}\colon \mathcal{X} \to (-\infty, \infty]$ defined by

   $$\varphi^{**}(x) = \sup_{\ell \in \mathcal{X}^*} (\ell(x) - \varphi^*(\ell))$$

   for all $x \in \mathcal{X}$.

Example 2.2. Suppose $\mathcal{X} = \mathbb{R}$ and $\varphi\colon \mathcal{X} \to (-\infty, \infty]$ is given by

$$\varphi(x) = \frac{1}{p}|x|^p$$

for all $x \in \mathcal{X}$ where $1 < p < \infty$. Recall from the Riesz representation theorem for Hilbert, each $\ell \in \mathcal{X}^*$ has the form $\ell = \ell_y$ for a unique $y \in \mathbb{R}$ where $\ell_y(x) = yx$ for all $x \in \mathcal{X}$. Using this fact, suppose $\ell = \ell_y$ is in $\mathcal{X}^*$. Then we have

$$\begin{aligned}
\varphi^*(y) &:= \varphi^*(\ell_y) \\
&= \sup_{x\in\mathbb{R}}(\ell_y(x) - \varphi(x)) \\
&= \sup_{x\in\mathbb{R}}\left(yx - \frac{1}{p}|x|^p\right) \\
&= \sup_{x\in\mathbb{R}}\left(|y||x| - \frac{1}{p}|x|^p\right) \\
&= \frac{1}{q}|y|^q + \frac{1}{p}|y^{p/q}|^p - \frac{1}{p}|y^{p/q}|^p \\
&= \frac{1}{q}|y|^q,
\end{aligned}$$

where $1 < q < \infty$ such that $1/p + 1/q = 1$. Here, we used Young's inequality, which says

$$\frac{a^p}{p} + \frac{b^q}{q} \geq ab$$

for all $a, b \geq 0$, with equality acheived if and only if $a^p = b^q$.

The example above suggests that we have the following generalization of Young's inequality:

$$\varphi^*(\ell) + \varphi(x) \geq \ell(x)$$

for all $\ell \in \mathcal{X}^*$ and $x \in \mathcal{X}$. Indeed, this is a simple consequence of the definition of φ^*: for all $\ell \in \mathcal{X}^*$, we have

$$\begin{aligned}
\varphi^*(\ell) &= \sup_{x\in\mathcal{X}}(\ell(x) - \varphi(x)) \\
&\geq \varphi(x) - \ell(x)
\end{aligned}$$

for all $x \in \mathcal{X}$.

2.3.2 Fenchel-Moreau

Lemma 2.5. *Let $\mathcal{X}$ be a normed linear space and let $\varphi\colon \mathcal{X} \to (-\infty, \infty]$ be a lower semicontinuous convex function such that $\varphi \neq \infty$. Then $\varphi^* \neq \infty$.*

Proof. Choose $x_0 \in \mathcal{X}$ such that $\varphi(x_0) < \infty$ and choose $\lambda_0 \in \mathbb{R}$ such that $\lambda_0 < \varphi(x_0)$. Consider the normed linear space $\mathcal{X} \times \mathbb{R}$ and the subsets $A = \{(x, \lambda) \mid \varphi(x) \leq \lambda\}$ and $B = \{(x_0, \lambda_0)\}$. Then A is a nonempty closed convex set and B is a nonempty compact convex set. Furthermore A and B are disjoint from one another. Thus by the second geometric form of Hahn-Banach, there exists a bounded linear functional $\ell\colon \mathcal{X} \times \mathbb{R} \to \mathbb{R}$ and an $\alpha \in \mathbb{R}$ such that

$$A \subseteq \{\ell > \alpha\} \quad \text{and} \quad B \subseteq \{\ell < \alpha\}. \tag{4}$$

Define $\psi\colon \mathcal{X} \to \mathbb{R}$ by $\psi(x) = \ell(x, 0)$ for all $x \in \mathcal{X}$. Then ψ is a bounded linear functional because ℓ is a bounded linear functional and $\psi = \ell|_{\mathbb{R}\times\{0\}}$. Set $k = \ell(0, 1)$ and note that

$$\begin{aligned}
\ell(x, \lambda) &= \ell(x, 0) + \ell(0, \lambda) \\
&= \psi(x) + \lambda k
\end{aligned}$$

for all $(x, \lambda) \in \mathcal{X} \times \mathbb{R}$.

Now by (4), we have

$$\begin{cases} \psi(x) + \lambda k > \alpha & \text{if } (x, \lambda) \in A \\ \psi(x_0) + \lambda_0 k < \alpha \end{cases}$$

In particular, since $(x_0, \varphi(x_0)) \in A$, we have

$$\begin{aligned}
0 &< \psi(x_0) + \varphi(x_0)k - \alpha \\
&< \psi(x_0) + \varphi(x_0)k - \psi(x_0) - \lambda_0 k \\
&= \varphi(x_0)k - \lambda_0 k \\
&= (\varphi(x_0) - \lambda_0)k.
\end{aligned}$$

Thus $k > 0$ since $\varphi(x_0) > \lambda_0$. Now using the fact that $(x, \varphi(x)) \in A$ for all $x \in \mathcal{X}$, we can divide $\psi(x) + \lambda k > \alpha$ by $-1/k$ to obtain

$$-\frac{1}{k}\psi(x) - \varphi(x) < -\frac{\alpha}{k}.$$

In particular, we see that

$$\begin{aligned}\varphi^*(-\psi/k) &= \sup_{x\in\mathcal{X}}(-\psi(x)/k - \varphi(x)) \\ &\le -\frac{\alpha}{k} \\ &< \infty.\end{aligned}$$

So $\varphi^* \neq \infty$. $\square$

Theorem 2.6. *(Fenchel-Moreau) If $\varphi\colon X \to (-\infty, \infty]$ is lower semicontinuous, convex, and $\varphi \neq \infty$, then $\varphi^{**} = \varphi$.*

Proof. Note that for every $\ell \in \mathcal{X}^*$ and $x \in \mathcal{X}$, we have

$$\begin{aligned}\ell(x) - \varphi^*(\ell) &= \ell(x) - \sup_{y\in\mathcal{X}}(\ell(y) - \varphi(y)) \\ &\le \ell(x) - (\ell(x) - \varphi(x)) \\ &= \varphi(x).\end{aligned}$$

Therefore

$$\begin{aligned}\varphi^{**}(x) &= \sup_{\ell\in\mathcal{X}^*}(\ell(x) - \varphi^*(\ell)) \\ &\le \varphi(x).\end{aligned}$$

It remains to show $\varphi^{**}(x) \ge \varphi(x)$.

Step 1: Suppose $\varphi \ge 0$ and assume for a contradiction that $\varphi^{**}(x_0) < \varphi(x_0)$. We apply the second geometric form of Hahn-Banach again in the space $\mathcal{X} \times \mathbb{R}$ with sets $A = \{(x, \lambda) \mid \varphi(x) \le \lambda\}$ and $B = \{(x_0, \varphi^{**}(x_0))\}$. By the same argument as in the proof of Lemma (2.5), there exists a bounded linear functional $\ell \in \mathcal{X}^*$, an $\alpha \in \mathbb{R}$, and a $k \in \mathbb{R}$ such that

$$\ell(x) + \lambda k > \alpha \tag{5}$$

for all $(x, \lambda) \in A$ and such that

$$\ell(x_0) + k\varphi^{**}(x_0) < \alpha \tag{6}$$

Note that we could have $\varphi(x_0) = \infty$, so we can't plug in $(x_0, \varphi(x_0))$ into (5) to conclude that $k > 0$ as in the proof of Lemma (2.5). However we can still show that $k \ge 0$. Indeed, assume for a contradiction that $k < 0$. Choose $y_0 \in \mathcal{X}$ such that $\varphi(y_0) < \infty$. Since $(y_0, \varphi(y_0)) \in A$, we have

$$\ell(y_0) + k\lambda \ge \ell(y_0) + k\varphi(y_0) > \alpha$$

for all $\lambda \ge \varphi(y_0)$. In particular, taking $\lambda \to \infty$ gives us $-\infty \ge \alpha$, which is a contradiction. So we must have $k \ge 0$. In order to proceed with the proof, we need to make k a little bigger, so choose $\varepsilon > 0$ so that $k + \varepsilon > 0$. Then just as in the proof of Lemma (2.5), we have

$$\varphi^*\left(-\frac{1}{k+\varepsilon}\ell\right) = \sup_{x\in\mathcal{X}}\left(-\frac{1}{k+\varepsilon}\ell(x) - \varphi(x)\right) \le -\frac{\alpha}{k+\varepsilon}$$

and hence

$$\begin{aligned}\ell(x_0) + (k+\varepsilon)\varphi^{**}(x_0) &= \ell(x_0) + (k+\varepsilon)\sup_{\ell\in\mathcal{X}^*}(\ell(x_0) - \varphi^*(\ell)) \\ &\ge \ell(x_0) + (k+\varepsilon)\left(-\frac{1}{k+\varepsilon}\ell(x_0) - \varphi^*\left(-\frac{1}{k+\varepsilon}\ell\right)\right) \\ &\ge \ell(x_0) + (k+\varepsilon)\left(-\frac{1}{k+\varepsilon}\ell(x_0) + \frac{\alpha}{k+\varepsilon}\right) \\ &= \ell(x_0) - \ell(x_0) + \alpha \\ &= \alpha.\end{aligned}$$

By taking $\varepsilon \to 0$, we obtain

$$\ell(x_0) + k\varphi^{**}(x_0) \geq \alpha,$$

which contradicts (6). This contradiction proves that $\varphi^{**} \geq \varphi$, and hence $\varphi^{**} = \varphi$.

Step 2: Now consider the general case where we may not have $\varphi \geq 0$. Choose $\ell_0 \in \mathcal{X}^*$ such that $\varphi^*(\ell_0) < \infty$ (such ℓ_0 exists by Lemma (2.5)). Define $\varphi_1 \colon \mathcal{X} \to (-\infty, \infty]$ by

$$\varphi_1(x) = \varphi(x) - \ell_0(x) + \varphi^*(\ell_0).$$

Then φ_1 is convex, lower semicontinuous, and $\varphi_1 \not\equiv \infty$. In addition, we have $\varphi_1 \geq 0$. So by step 1, we obtain $\varphi_1^{**} = \varphi_1$. Now observe that

$$\begin{aligned}
\varphi_1^*(\ell) &= \sup_{x\in\mathcal{X}}(\ell(x) - \varphi_1(x)) \\
&= \sup_{x\in\mathcal{X}}(\ell(x) - \varphi(x) + \ell_0(x) - \varphi^*(\ell_0)) \\
&= \sup_{x\in\mathcal{X}}((\ell + \ell_0)(x) - \varphi(x)) - \varphi^*(\ell_0) \\
&= \varphi^*(\ell + \ell_0) - \varphi^*(\ell_0).
\end{aligned}$$

Therefore

$$\begin{aligned}
\varphi_1^{**}(x) &= \sup_{\ell\in\mathcal{X}^*}(\ell(x) - \varphi_1^*(\ell)) \\
&= \sup_{\ell\in\mathcal{X}^*}(\ell(x) - \varphi^*(\ell + \ell_0) + \varphi^*(\ell_0)) \\
&= \sup_{\ell+\ell_0\in\mathcal{X}^*}((\ell + \ell_0)(x) - \varphi^*(\ell + \ell_0) - \ell_0(x) + \varphi^*(\ell_0)) \\
&= \varphi^{**}(x) - \ell_0(x) + \varphi^*(\ell_0).
\end{aligned}$$

So

$$\begin{aligned}
\varphi^{**}(x) - \ell_0(x) + \varphi^*(\ell_0) &= \varphi_1^{**}(x) \\
&= \varphi_1(x) \\
&= \varphi(x) - \ell_0(x) + \varphi^*(\ell_0).
\end{aligned}$$

Hence $\varphi^{**} = \varphi$.

□

2.3.3 Example

Example 2.3. Let $\mathcal{X}$ be a normed linear space and consider let $\varphi = \|\cdot\|$ be the norm function. Then φ is lower semicontinuous and convex. Let's compute the conjugate function

$$\begin{aligned}
\varphi^*(\ell) &= \sup_{x\in\mathcal{X}}(\ell(x) - \|x\|) \\
&= \sup_{x\in\mathcal{X}} \|x\| \left(\ell\left(\frac{x}{\|x\|}\right) - 1\right).
\end{aligned}$$

Now if $\|\ell\| > 1$, then there exists $x_0 \in \mathcal{X}$ such that $\|x_0\| = 1$ and $\ell(x_0) > 1$. Then for any $\lambda \in \mathbb{R}$, we have

$$\begin{aligned}
\varphi^*(\ell) &= \sup_{x\in\mathcal{X}} \|x\| \left(\ell\left(\frac{x}{\|x\|}\right) - 1\right) \\
&\geq \|\lambda x_0\| \left(\ell\left(\frac{\lambda x_0}{\|\lambda x_0\|}\right) - 1\right) \\
&= |\lambda| \left(\ell(x_0) - 1\right),
\end{aligned}$$

so by taking $\lambda \to \infty$, we see that $\varphi^*(\ell) = \infty$. On the other hand, if $\|\ell\| \leq 1$, then it is easy to check that $\varphi^*(\ell) = 0$. Thus

$$\varphi^*(\ell) = \begin{cases} 0 & \text{if } \|\ell\| \leq 1 \\ \infty & \text{if } \|\ell\| > 1 \end{cases}$$

For a set $E \subseteq \mathcal{X}$ nonempty we define

$$\mathrm{I}_E(x) = \begin{cases} 0 & \text{if } x \in E \\ \infty & \text{if } x \notin E \end{cases} = \log\left(\frac{1}{1_E(x)}\right)$$

So $\varphi^* = 1_{\mathrm{B}_1[0]}$ where

$$\mathrm{B}_1[0] = \{\ell \in \mathcal{X}^* \mid \|\ell\| \leq 1\}.$$

Now we have

$$\begin{aligned} \|x\| &= \varphi(x) \\ &= \varphi^{**}(x) \\ &= \sup_{\ell \in \mathcal{X}^*} (\ell(x) - \varphi^*(x)) \\ &= \sup_{\ell \in \mathcal{X}^*} (\ell(x) - \mathrm{I}_{\mathrm{B}_1[0]}(x)) \\ &= \sup_{\substack{\ell \in \mathcal{X}^* \\ \|\ell\| \leq 1}} \ell(x). \end{aligned}$$

This identity can be proved in a more elementary way by applying Hahn-Banach.

2.4 Support Functional

Definition 2.6. Let $\mathcal{X}$ be a normed linear space and let S be a subset of $\mathcal{X}$. We define $\mathrm{q}_S \colon \mathcal{X}^* \to (-\infty, \infty]$ by

$$\mathrm{q}_S(\ell) = \sup_{x \in S} \ell(x).$$

We call q_S the **support functional** of C.

2.4.1 Basic Properties of Support Functional

Proposition 2.4. *Let $\mathcal{X}$ be a normed linear space and let S be a subset of $\mathcal{X}$. Then*

1. q_S *is a partial-seminorm.*
2. $\mathrm{q}_S = \mathrm{q}_{\mathrm{conv}(S)} = \mathrm{q}_{\overline{\mathrm{conv}}(S)}$,
3. *Let S_1 and S_2 be subsets of $\mathcal{X}$. Then $\mathrm{q}_{S_1+S_2} = \mathrm{q}_{S_1} + \mathrm{q}_{S_2}$.*
4. *Let $\mathcal{K}$ be a closed subspace of $\mathcal{X}$. Then*

$$\mathrm{q}_{\mathcal{K}}(\ell) = \begin{cases} 0 & \textit{if } \ell \in \mathcal{K}^\perp \\ \infty & \textit{else} \end{cases}$$

where $\mathcal{K}^\perp = \{\ell \in \mathcal{X}^* \mid \ell|_{\mathcal{K}} = 0\}$.

Proof. 1. Clearly q_S is nonnegative since $\ell(0) = 0$ for all linear functionals $\ell \in \mathcal{X}^*$. Next, suppose $\lambda \geq 0$ and $\ell \in \mathcal{X}^*$. Then

$$\begin{aligned} \mathrm{q}_S(\lambda\ell) &= \sup_{x \in S} \ell(\lambda x) \\ &= \sup_{x \in S} \lambda\ell(x) \\ &= \lambda \sup_{x \in S} \ell(x) \\ &= \lambda \mathrm{q}_S(\ell). \end{aligned}$$

Similarly, suppose $\ell_1, \ell_2 \in \mathcal{X}^\times$. Then

$$\begin{aligned} \mathrm{q}_S(\ell_1 + \ell_2) &= \sup_{x \in S}\{(\ell_1 + \ell_2)(x)\} \\ &= \sup_{x \in S}\{\ell_1(x) + \ell_2(x)\} \\ &\leq \sup_{x \in S}\{\ell_1(x)\} + \sup_{x \in S}\{\ell_2(x)\} \\ &= \mathrm{q}_S(\ell_1) + \mathrm{q}_S(\ell_2). \end{aligned}$$

Thus q_S is a partial-seminorm.

2. Since $S \subseteq \mathrm{conv}(S) \subseteq \overline{\mathrm{conv}}(S)$, we clearly have $\mathrm{q}_S \leq \mathrm{q}_{\mathrm{conv}(S)} \leq \mathrm{q}_{\overline{\mathrm{conv}}(S)}$. Conversely, let $\ell \in \mathcal{X}^*$ and let $tx + (1-t)y \in \mathrm{conv}(S)$ where $t \in (0,1)$ and $x, y \in S$. Then observe that

$$\begin{aligned} \ell(tx + (1-t)y) &= t\ell(x) + (1-t)\ell(y) \\ &\leq t \sup_{z \in S} \ell(z) + (1-t) \sup_{z \in S} \ell(z) \\ &= t\mathrm{q}_S(\ell) + (1-t)\mathrm{q}_S(\ell) \\ &= \mathrm{q}_S(\ell). \end{aligned}$$

It follows that $\mathrm{q}_{\mathrm{conv}(S)}(\ell) \leq \mathrm{q}_S(\ell)$, and since ℓ was arbitrary, we have $\mathrm{q}_{\mathrm{conv}(S)} \leq \mathrm{q}_S$. To show $\mathrm{q}_{\overline{\mathrm{conv}}(S)} \leq \mathrm{q}_{\mathrm{conv}(S)}$, we will prove something more general: if E is a subset of $\mathcal{X}$, then $\mathrm{q}_{\overline{E}} \leq \mathrm{q}_E$. Indeed, let $\ell \in \mathcal{X}^*$, let $x \in \overline{E}$, and choose a sequence (x_n) of elements in E such that $x_n \to x$. Then observe that

$$\begin{aligned} \ell(x) &= \lim_{n \to \infty} \ell(x_n) \\ &\leq \sup_{y \in E} \ell(y) \\ &= \mathrm{q}_E(\ell). \end{aligned}$$

It follows that $\mathrm{q}_{\overline{E}}(\ell) \leq \mathrm{q}_E(\ell)$, and since ℓ was arbitrary, we have $\mathrm{q}_{\overline{E}} \leq \mathrm{q}_E$.

3. Let $x_1 + x_2 \in S_1 + S_2$ and let $\ell \in \mathcal{X}^*$. Then observe that

$$\begin{aligned} \ell(x_1 + x_2) &= \ell(x_1) + \ell(x_2) \\ &\leq \sup_{y_1 \in S_1} \ell(y_1) + \sup_{y_2 \in S_2} \ell(y_2) \\ &= \mathrm{q}_{S_1}(\ell) + \mathrm{q}_{S_2}(\ell) \\ &= (\mathrm{q}_{S_1} + \mathrm{q}_{S_2})(\ell) \end{aligned}$$

It follows that $\mathrm{q}_{S_1+S_2}(\ell) \leq (\mathrm{q}_{S_1} + \mathrm{q}_{S_2})(\ell)$, and since ℓ was arbitrary, we have $\mathrm{q}_{S_1+S_2} \leq \mathrm{q}_{S_1} + \mathrm{q}_{S_2}$. Conversely, let $\ell \in \mathcal{X}^*$, let $\varepsilon > 0$, and choose $x_1 \in S_1$ and $x_2 \in S_2$ such that $\ell(x_1) + \varepsilon/2 > \mathrm{q}_{S_1}(\ell)$ and $\ell(x_2) + \varepsilon/2 > \mathrm{q}_{S_2}(\ell)$. Then observe that

$$\begin{aligned} (\mathrm{q}_{S_1} + \mathrm{q}_{S_2})(\ell) &= \mathrm{q}_{S_1}(\ell) + \mathrm{q}_{S_2}(\ell) \\ &< \ell(x_1) + \frac{\varepsilon}{2} + \ell(x_2) + \frac{\varepsilon}{2} \\ &= \ell(x_1) + \ell(x_2) + \varepsilon \\ &= \ell(x_1 + x_2) + \varepsilon \\ &\leq \mathrm{q}_{S_1+S_2}(\ell) + \varepsilon. \end{aligned}$$

By taking $\varepsilon \to 0$, we see that $(\mathrm{q}_{S_1} + \mathrm{q}_{S_2})(\ell) \leq \mathrm{q}_{S_1+S_2}(\ell)$, and since ℓ was arbitrary, we have $\mathrm{q}_{S_1} + \mathrm{q}_{S_2} \leq \mathrm{q}_{S_1+S_2}$.

4. Let $\ell \in \mathcal{X}^*$. First suppose that $\ell \in \mathcal{K}^\perp$. Then $\ell(x) = 0$ for all $x \in \mathcal{K}$. Thus

$$\begin{aligned} \mathrm{q}_{\mathcal{K}}(\ell) &= \sup_{x \in \mathcal{K}} \ell(x) \\ &= \sup_{x \in \mathcal{K}} 0 \\ &= 0. \end{aligned}$$

Now suppose that $\ell \notin \mathcal{K}^\perp$. Choose $x \in \mathcal{K}$ such that $\ell(x) \neq 0$ and let $\lambda \geq 0$. Then observe that

$$\begin{aligned} \lambda\ell(x) &= \ell(\lambda x) \\ &\leq \sup_{y \in \mathcal{K}} \ell(y) \\ &= \mathrm{q}_{\mathcal{K}}(\ell). \end{aligned}$$

Taking $\lambda \to \infty$ gives us $\mathrm{q}_{\mathcal{K}}(\ell) = \infty$. □

2.4.2 Examples of Support Functionals

Example 2.4. Suppose $C = \{x_0\}$. Then $\mathrm{q}_{\{x_0\}}(\ell) = \ell(x_0)$.

Example 2.5. Suppose $C = \mathrm{B}_1[0]$, then $\mathrm{q}_{\mathrm{B}_1[0]} = \|\ell\|$.

Example 2.6. Suppose $C = \mathrm{B}_R[0]$, then $\mathrm{q}_{\mathrm{B}_R[0]} = R\|\ell\|$. Recall that the gauge functional in this case is $\mathrm{p}_{\mathrm{B}_R[0]}(x) = \|x\|/R$. More generally, we have

$$\begin{aligned} \mathrm{q}_{\mathrm{B}_R[x_0]}(x) &= \mathrm{q}_{\{x_0\}+\mathrm{B}_R[0]}(x) \\ &= \mathrm{q}_{\{x_0\}}(x) + \mathrm{q}_{\mathrm{B}_R[0]}(x) \\ &= \ell(x_0) + R\|\ell\|. \end{aligned}$$

If $\mathcal{M}$ is a closed subspace of $\mathcal{X}$, then

$$\mathrm{q}_{\mathcal{M}}(\ell) = \begin{cases} 0 & \text{if } \ell \in \mathcal{M}^\perp \\ \infty & \text{else} \end{cases}$$

Let $\varphi(x) = \mathrm{I}_E(x)$ for some set $E \subseteq \mathcal{X}$. Then

$$\begin{aligned} \varphi^*(\ell) &= \sup_{x\in\mathcal{X}}(\ell(x) - \mathrm{I}_E(x)) \\ &= \sup_{x\in E} \ell(x) \\ &= \mathrm{q}_E(\ell). \end{aligned}$$

Notice $\varphi^*(\ell) = \mathrm{q}_{\overline{\mathrm{conv}}(E)}(\ell)$. Then

$$\begin{aligned} \varphi^{**}(x) &= \sup_{\ell\in\mathcal{X}^*}(\ell(x) - \varphi^*(\ell)) \\ &= \sup_{\ell\in\mathcal{X}^*}(\ell(x) - \mathrm{q}_E(\ell)) \\ &= \sup_{\ell\in\mathcal{X}^*}(\ell(x) - \mathrm{q}_{\overline{\mathrm{conv}}(E)}(\ell)) \end{aligned}$$

It can be shown that I_E is convex if and only if E is convex. It can also be shown that I_E is lower semicontinuous if and only if E is closed. So if E is clsoed and convex, then Fenchel-Moreau applies and we get

$$\mathrm{I}_E(x) = \sup_{\ell\in\mathcal{X}^*}(\ell(x) - \mathrm{q}_E(\ell)).$$

In some sense, the gauge (Minkowski) functional p_C plays the role of a norm if we want C convex to play the role of the unit ball. In that sense, the support functional q_C plays the role of the norm in the dual space $\mathcal{X}^*$. In this direct, the Cauchy-Schwarz inequality $|\ell(x)| \leq \|\ell\|\|x\|$ is replaced by

$$|\ell(x)| \leq \mathrm{q}_C(\ell)\mathrm{p}_C(x) \tag{7}$$

for all $\ell \in \mathcal{X}^*$ and $x \in \mathcal{X}$. Indeed, for any $x \in \mathcal{X}$ and $\varepsilon > 0$ we have $x/(\mathrm{p}_C(x) + \varepsilon) \in C$ by definition of $\mathrm{p}_C(x)$, and thus

$$\ell\left(\frac{1}{\mathrm{p}_C(x)+\varepsilon}x\right) \leq \sup_{y\in C}\ell(y) = \mathrm{q}_C(\ell)$$

which implies (7).

Proposition 2.5. *$x \in \overline{\mathrm{conv}}(E)$ if and only if $\ell(x) \leq \mathrm{q}_E(\ell)$ for all $\ell \in \mathcal{X}^*$.*

Proof. Recall that $\mathrm{I}_E^*(\ell) = \mathrm{q}_E(\ell) = \mathrm{q}_{\overline{\mathrm{conv}}(E)}(\ell) = \mathrm{I}^*_{\overline{\mathrm{conv}}(E)}$. We have

$$\begin{aligned} \mathrm{I}_E^{**}(x) &= \sup_{\ell\in\mathcal{X}^*}(\ell(x) - \mathrm{I}_E^*(\ell)) \\ &= \sup_{\ell\in\mathcal{X}^*}(\ell(x) - \mathrm{q}_E(\ell)). \end{aligned}$$

On the other hand, we have

$$\begin{aligned} \mathrm{I}_E^{**}(x) &= \sup_{\ell\in\mathcal{X}^*}(\ell(x) - \mathrm{I}^*_{\overline{\mathrm{conv}}(E)}(\ell)) \\ &= \mathrm{I}^{**}_{\overline{\mathrm{conv}}(E)}(x) \end{aligned}$$

We can apply Fenchel-Moreau to $\mathrm{I}_{\overline{\mathrm{conv}}(E)}$ which is convex and lowersemicontinuous and obtain

$$\mathrm{I}_{\overline{\mathrm{conv}}(E)}(x) = \sup_{\ell \in \mathcal{X}^*} (\ell(x) - \mathrm{q}_E(\ell))$$

So

$$\begin{aligned} x \in \overline{\mathrm{conv}}(E) &\iff \mathrm{I}_{\overline{\mathrm{conv}}(E)}(x) = 0 \\ &\iff \sup_{\ell \in \mathcal{X}^*} (\ell(x) - \mathrm{q}_E(\ell)) = 0 \\ &\iff \ell(x) \le \mathrm{q}_E(\ell) \text{ for all } \ell \in \mathcal{X}^*. \end{aligned}$$

□

2.5 Another Application

For a subspace $\mathcal{M} \subseteq \mathcal{X}$, we define its **annihilator** by

$$\mathcal{M}^\perp = \{\ell \in \mathcal{X}^* \mid \ell|_{\mathcal{M}} = 0\}.$$

For a closed subspace $\mathcal{N} \subseteq \mathcal{X}^*$, we define

$$\mathcal{N}_\perp = \{x \in \mathcal{X} \mid \ell(x) = 0 \text{ for all } \ell \in \mathcal{N}\}.$$

Proposition 2.6. *If $\mathcal{M} \subseteq \mathcal{X}$ is a closed subspace, then $(\mathcal{M}^\perp)_\perp = \mathcal{M}$.*

Proof. We have $\mathrm{I}^*_{\mathcal{M}}(\ell) = \mathrm{q}_{\mathcal{M}}(\ell) = \mathrm{I}_{\mathcal{M}^\perp}(\ell)$. So

$$\begin{aligned} \mathrm{I}_{\mathcal{M}}(x) &= \mathrm{I}^{**}_{\mathcal{M}}(x) \\ &= \sup_{\ell \in \mathcal{X}^*} (\ell(x) - \mathrm{I}^*_{\mathcal{M}}(\ell)) \\ &= \sup_{\ell \in \mathcal{X}^*} (\ell(x) - \mathrm{I}_{\mathcal{M}^\perp}(\ell)) \\ &= \sup_{\ell \in \mathcal{M}^\perp} (\ell(x)) \\ &= \mathrm{I}_{(\mathcal{M}^\perp)_\perp}(x) \end{aligned}$$

□

2.6 Fenchel-Rockafeller

Theorem 2.7. *(Fenchel-Rockafellar) Let $\varphi, \psi \colon \mathcal{X} \to (-\infty, \infty]$ be two convex functions. Suppose there exists $x_0 \in \mathcal{X}$ such that $\varphi(x_0), \psi(x_0) < \infty$ and φ is continuous at x_0. Then*

$$\inf_{x \in \mathcal{X}} (\varphi(x) + \psi(x)) = \sup_{\ell \in \mathcal{X}^*} (-\varphi^*(-\ell) - \psi^*(\ell)) = \max_{\ell \in \mathcal{X}^*} (-\varphi^*(-\ell) - \psi^*(\ell)).$$

Proof. (Sketch) Let $a = \inf_{x \in \mathcal{X}} (\varphi(x) + \psi(x))$ and let $b = \sup_{\ell \in \mathcal{X}^*} (-\varphi^*(-\ell) - \psi^*(\ell))$. It's easy to see that $b \le a$. Indeed,

$$\begin{aligned} -\varphi^*(\ell) - \psi^*(\ell) &= -\varphi^*(-\ell) - (-\ell(x)) - \psi^*(\ell) - \ell(x) \\ &\le \varphi(x) + \psi(x) \end{aligned}$$

for all $x \in \mathcal{X}$ and $\ell \in \mathcal{X}^*$. For the reverse direction, let $C = \operatorname{epi} \varphi$, let $B = \{(x, \lambda) \mid \lambda \le a - \psi(x)\}$, and let $A = \operatorname{int} C$. Then A and B are both nonempty convex sets. Furthemore we have $A \cap B = \emptyset$ (otherwise we'll have $(x, \lambda) \in \mathcal{X} \times \mathbb{R}$ such that $\varphi(x) < \lambda \le a - \psi(x)$ which implies $\varphi(x) + \psi(x) < a$, giving a contradiction). Applying Hahn-Banach, we obtain a linear functional $\Phi \colon \mathcal{X} \times \mathbb{R} \to \mathbb{R}$ such that $\overline{C} = \overline{A} \subseteq \{\Phi \ge \alpha\}$ and $B \subseteq \{\Phi \le \alpha\}$. Let $\ell(x) = \Phi(x, 0)$ and $k = \Phi(0, 1) \in \mathbb{R}$. Then

$$\begin{aligned} \ell(x) + k\lambda &\ge \alpha \text{ for } (x, \lambda) \in \overline{A} = \overline{C} \\ \ell(x) + k\lambda &\le \alpha \text{ for } (x, \lambda) \in B. \end{aligned}$$

Similarly as before, one can show that $k > 0$. □

2.6.1 Application

Let $C \subseteq \mathcal{X}$ be non-empty and convex. Then

$$\mathrm{d}(x_0, C) = \inf_{x \in C} \|x_0 - x\| = \sup_{\substack{\ell \in \mathcal{X}^* \\ \|\ell\| \leq 1}} (\ell(x_0) - \mathrm{q}_C(\ell)).$$

Then $\varphi(x) = \|x - x_0\|$ is convex and $\psi(x) = \mathrm{I}_C(x)$ is convex if C is convex. Then

$$\begin{aligned}
\varphi^*(\ell) &= \sup_{x \in \mathcal{X}} (\ell(x) - \|x - x_0\|)) \\
&= \sup_{x \in \mathcal{X}} (\ell(x - x_0) - \|x - x_0\| + \ell(x_0)) \\
&= \varphi^*(\ell) \\
&= \mathrm{I}_{\mathrm{B}_1[0]}(\ell) + \ell(x_0).
\end{aligned}$$

So by Fenchel-Rockafellar, we have

$$\inf_{x \in \mathcal{X}} (\|x - x_0\| + \mathrm{I}_C(x)) = \sup_{\ell \in \mathcal{X}^*} (\ell(x_0) - \mathrm{I}_{\mathrm{B}_1[0]}(-\ell) - \mathrm{q}_C(\ell))$$

Before starting the proof, recall that we proved last time using Fenchel-Rockafellar that if $C \neq \emptyset$ is convex, then

$$\mathrm{d}(x_0, C) = \sup_{\substack{\ell \in \mathcal{X}^* \\ \|\ell\| \leq 1}} (\ell(x_0) - \mathrm{q}_C(\ell)).$$

Note that when $C = \mathcal{M}$ is a subspace, we have

$$\mathrm{d}(x_0, C) = \sup_{\substack{\ell \in \mathcal{X}^* \\ \|\ell\| \leq 1}} (\ell(x_0) - \mathrm{I}_{\mathcal{M}^\perp}(\ell)) = \sup_{\substack{\ell \in \mathcal{M}^\perp \\ \|\ell\| \leq 1}} \ell(x).$$

3 Baire Category Theorem

Theorem 3.1. *Let $\mathcal{X}$ be a Banach space. Then $\mathcal{X}$ cannot be represented as a countable union of nowhere dense sets.*

Recall that a set $E \subseteq \mathcal{X}$ is said to be nowhere dense if $(\overline{E})^\circ = \emptyset$. In other words, $\overline{E}$ doesn't contain any open balls.

Proof. Assume for a contradiction that $\mathcal{X} = \bigcup_{n=1}^\infty E_n$ with every E_n being nowhere dense. In particular, we have $\mathcal{X} = \bigcup_{n=1}^\infty \overline{E}_n$. Let $\mathrm{B}_{r_1}(x_1) \subseteq \mathcal{X}$ be any open ball. Since E_1 is nowhere dense, it follows that $\mathrm{B}_{r_1}(x_1) \cap \overline{E}_1^c$ is a nonempty open set. Thus there exists an open ball, say $\mathrm{B}_{r_2}(x_2)$, such that $\mathrm{B}_{r_2}[x_2] \subseteq \mathrm{B}_{r_1}(x_1) \cap \overline{E}_1^c$ and $r_2 < 2^{-2}$. Since E_2 is nowhere dense, it follows that $\mathrm{B}_{r_2}(x_2) \cap \overline{E}_2^c$ is a nonempty open set. So by the same reason as before, there exists an open ball, say $\mathrm{B}_{r_3}(x_3)$, such that $\mathrm{B}_{r_3}[x_3] \subseteq \mathrm{B}_{r_2}(x_2) \cap \overline{E}_2^c$ and $r_3 < 2^{-3}$. Continuing this process, we obtain a descending sequence of open balls $(\mathrm{B}_{r_n}(x_n))$ such that

$$\mathrm{B}_{r_n}[x_n] \subseteq \mathrm{B}_{r_{n-1}}(x_{n-1}) \cap \overline{E}_{n-1}^c \quad \text{and} \quad r_n < 2^{-n}$$

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

Now let $\varepsilon > 0$ and choose $N \in \mathbb{N}$ such that $2^{-N} < \varepsilon$. Then $n > m \geq N$ implies

$$\begin{aligned}
\|x_m - x_n\| &\leq r_m \\
&< 2^{-m} \\
&\leq 2^{-N} \\
&< \varepsilon.
\end{aligned}$$

Thus (x_n) is a Cauchy sequence. Being a Cauchy sequence in a Banach space, we see that (x_n) is convergent, say $x_n \to x$. Since $x_n \in \mathrm{B}_{r_k}(x_k)$ for any $n \geq k$, we have $x \in \mathrm{B}_{r_k}[x_k]$. In particular, this implies

$$\begin{aligned}
x &\in \bigcap_{n=1}^\infty \mathrm{B}_{r_n}[x_n] \\
&\subseteq \bigcap_{n=1}^\infty \overline{E}_n^c \\
&= \left(\bigcup_{n=1}^\infty \overline{E}_n \right)^c \\
&= \mathcal{X}^c \\
&= \emptyset,
\end{aligned}$$

which is a contradiction. □

3.1 Uniform Boundedness Principle

Theorem 3.2. *(Uniform Boundedness Principle) Let $\mathcal{X}$ and $\mathcal{Y}$ be two Banach spaces. Denote by $\mathcal{L}(\mathcal{X},\mathcal{Y})$ the set of all bounded linear operators from $\mathcal{X}$ to $\mathcal{Y}$. Suppose $\mathcal{A} \subseteq \mathcal{L}(X,\mathcal{Y})$ such that for any $x \in \mathcal{X}$ the set $\{\|Tx\| \mid T \in \mathcal{A}\}$ is bounded above. Then the set $\{\|T\| \mid T \in \mathcal{A}\}$ is bounded above.*

Proof. For each $n \in \mathbb{N}$, let

$$E_n = \{x \in \mathcal{X} \mid \|Tx\| \leq n \text{ for all } T \in \mathcal{A}\}.$$

Observe that (E_n) is an ascending sequence of closed sets. Indeed, it is clearly ascending. To see that each E_n is closed, view it as an infinite intersection of closed sets, namely

$$E_n = \bigcap_{T\in\mathcal{A}} \{x \in \mathcal{X} \mid \|Tx\| \leq n\}.$$

Moreover, for any $x \in \mathcal{X}$ the set $\{\|Tx\| \mid T \in \mathcal{A}\}$ is bounded above, say $\{\|Tx\| \mid T \in \mathcal{A}\} \leq N$ for some $N \in \mathbb{N}$. It follows that $x \in E_N$ and since $x \in \mathcal{X}$ was arbitrary, we see that

$$\mathcal{X} = \bigcup_{n=1}^{\infty} E_n.$$

By the Baire Category Theorem, there must exist some $M \in \mathbb{N}$ such that E_M is not nowhere dense. In other words, E_M contains a nonempty contains a nonempty open ball, say $\mathrm{B}_r(x_0)$. By choosing r small enough, we can assume $\mathrm{B}_r[x_0] \subseteq E_M$. Then for any $x \in \mathrm{B}_1[0]$, we have

$$\begin{aligned}
\|T(rx)\| &\leq \|T(rx + x_0) - Tx_0\| \\
&\leq \|T(rx + x_0)\| + \|Tx_0\| \\
&\leq M + M \\
&= 2M
\end{aligned}$$

for all $T \in \mathcal{A}$. It follows that $\|T\| \leq 2M/r$ for all $T \in \mathcal{A}$. Thus the set $\{\|T\| \mid T \in \mathcal{A}\}$ is bounded above. □

Here is a simple application of the uniform boundedness principle.

Proposition 3.1. *Let (T_n) be a sequence of bounded linear operators $T_n\colon \mathcal{X} \to \mathcal{Y}$ between Banach spaces $\mathcal{X}$ and $\mathcal{Y}$. Assume for each $x \in \mathcal{X}$ the sequence $(T_n x)$ converges in $\mathcal{Y}$. Then the map $T\colon \mathcal{X} \to \mathcal{Y}$ defined by*

$$Tx := \lim_{n\to\infty} T_n x$$

for all $x \in \mathcal{X}$ is a bounded linear operator.

Proof. Since for each $x \in \mathcal{X}$ the sequence $(T_n x)$ is convergent we see that it must be bounded. Let $M_x = \sup_{n\in\mathbb{N}} \|T_n x\| < \infty$. By the uniform boundedness principle, there exists $M > 0$ such that $\sup_{n\in\mathbb{N}} \|T_n\| \leq M < \infty$. Therefore

$$\begin{aligned}
\|Tx\| &= \|\lim_{n\to\infty} T_n x\| \\
&= \lim_{n\to\infty} \|T_n x\| \\
&\leq \sup_{n\in\mathbb{N}} \|T_n x\| \\
&\leq \sup_{n\in\mathbb{N}} \|T_n\| \|x\| \\
&\leq M\|x\|.
\end{aligned}$$

It follows that T is bounded. □

4 Open Mapping Theorem and Closed Graph Theorem

4.1 Main Theorem

Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces. Consider the space $\mathcal{X} \times \mathcal{Y}$ with addition and scalar-multiplication defined pointwise. We endow $\mathcal{X} \times \mathcal{Y}$ with a norm defined by

$$\|(x,y)\|_{\mathcal{X}\times\mathcal{Y}} = \|x\|_{\mathcal{X}} + \|y\|_{\mathcal{Y}}. \tag{8}$$

for all $(x,y) \in \mathcal{X}$. It's easy to prove that $(\mathcal{X} \times \mathcal{Y}, \|\cdot\|_{\mathcal{X}\times\mathcal{Y}})$ is a Banach space. If context is clear, then we drop $\mathcal{X} \times \mathcal{Y}$ from the subscript in $\|\cdot\|_{\mathcal{X}\times\mathcal{Y}}$ in order to clean notation. We'll use the usual projection maps $\pi_1 \colon \mathcal{X} \times \mathcal{Y} \to \mathcal{X}$ and $\pi_2 \colon \mathcal{X} \times \mathcal{Y} \to \mathcal{Y}$ defined by $\pi_1(x,y) = x$ and $\pi_2(x,y) = y$. Clearly both π_1 and π_2 are bounded linear operators.

Theorem 4.1. *(Main result) Let $\mathcal{Z} \subseteq \mathcal{X} \times \mathcal{Y}$ be a closed subspace such that $\pi_2(\mathcal{Z}) = \mathcal{Y}$. If $U \subseteq \mathcal{X}$ is open, then $\pi_2(\pi_1^{-1}(U) \cap \mathcal{Z}))$ is an open subset of $\mathcal{Y}$.*

Remark 5. Note that by symmetry if instead of assuming $\pi_2(\mathcal{Z}) = \mathcal{Y}$ we assume $\pi_1(\mathcal{Z}) = \mathcal{X}$, then we have for any open set $V \subseteq \mathcal{Y}$ we have $\pi_1(\pi_2^{-1}(V) \cap \mathcal{Z}))$ is an open subset of $\mathcal{X}$.

4.2 Applications of the Main Theorem

Before we prove Theorem (4.1), let us show how to use it to prove both the open mapping theorem and the closed graph theorem.

4.2.1 Open Mapping Theorem

Theorem 4.2. *(Open mapping theorem) Let $T \colon \mathcal{X} \to \mathcal{Y}$ be a surjective bounded linear operator. Then T is an open map, meaning that for any open subset U of X, the set $T(U)$ is an open subset of $\mathcal{Y}$.*

Proof. Let $\mathcal{Z} = \{(x, Tx) \mid x \in \mathcal{X}\} \subseteq \mathcal{X} \times \mathcal{Y}$ and let U be an open subset of $\mathcal{X}$. Observe that $\mathcal{Z}$ is a closed subspace precisely because T is a bounded linear operator. Furthemore we have $\pi_2(\mathcal{Z}) = \mathcal{Y}$ since T is surjective. Finally, note that $T(U) = \pi_2(\pi_1^{-1}(U) \cap \mathcal{Z}))$. It follows from Theorem (4.1) that $T(U)$ is an open subset of $\mathcal{Y}$. □

4.2.2 Inverse Mapping Theorem

Theorem 4.3. *Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces and let $T \colon \mathcal{X} \to \mathcal{Y}$ be a bounded linear map which is bijective. Then $T^{-1} \colon \mathcal{Y} \to \mathcal{X}$ is also a bounded linear map.*

Proof. That T^{-1} is linear follows from basic linear algebra. The nontrivial part is that T^{-1} is also bounded. To see why, it suffices to show that T^{-1} is continuous. Let $U \subseteq \mathcal{X}$ be open. Then its preimage under T^{-1} is $T(U)$ since T is bijective. Since T is onto, it follows from the open mapping theorem, that $T(U)$ is open. Thus T^{-1} is continuous. □

4.2.3 Closed Graph Theorem

Theorem 4.4. *Let $T \colon \mathcal{X} \to \mathcal{Y}$ be a linear map such that $x_n \to x$ and $Tx_n \to y$ implies $y = Tx$, or in other words, if the graph of T given by $\{(x, Tx) \mid x \in \mathcal{X}\} \subseteq \mathcal{X} \times \mathcal{Y}$ is a closed set, then T is bounded.*

Proof. Again take $\mathcal{Z} = \{(x, Tx) \mid x \in \mathcal{X}\} \subseteq \mathcal{X} \times \mathcal{Y}$ and let V be an open subset of $\mathcal{Y}$. Since T is linear, $\mathcal{Z}$ is a subspace of $\mathcal{X} \times \mathcal{Y}$. Furthermore, $\mathcal{Z}$ is closed by assumption. Also we clearly have $\pi_1(\mathcal{Z}) = \mathcal{X}$. Finally, note that $T^{-1}(V) = \pi_1(\pi_2^{-1}(V) \cap \mathcal{Z}))$. It follows from Theorem (4.1) that $T^{-1}(V)$ is an open subset of $\mathcal{X}$. Thus T is continuous, and hence bounded. □

4.3 Zabreiko's Lemma

The proof of Theorem (4.1) will depend on the following lemma:

Lemma 4.5. *(Zabreiko) Let $\mathcal{X}$ be a Banach space and let $p \colon \mathcal{X} \to [0, \infty)$ be a seminorm on $\mathcal{X}$. Suppose p is* ***countably subadditive****, that is, suppose for every absolutely convergent series $\sum_{n=1}^{\infty} x_n$ in $\mathcal{X}$, we have*

$$p\left(\sum_{n=1}^{\infty} x_n\right) \leq \sum_{n=1}^{\infty} p(x_n).$$

Then there exists $C > 0$ such that $p(x) \leq C\|x\|$ for every $x \in \mathcal{X}$.

Proof. For each $n \in \mathbb{N}$, let $E_n = \{p \leq n\}$.

Step 1: We will find an $N \in \mathbb{N}$ and $r > 0$ such that $\mathrm{B}_r(0) \subseteq \overline{E}_N$. Observe that E_n is convex and symmetric (here symmetric means $x \in E_n$ implies $-x \in E_n$). From here it is easy to show that $\overline{E}_n$ is convex, symmetric, and closed. Clearly

$$\mathcal{X} = \bigcup_{n=1}^{\infty} \overline{E}_n.$$

So by the Baire category theorem, there exists an $N \in \mathbb{N}$ and an open ball $\mathrm{B}_r(x_0)$ such that $\mathrm{B}_r(x_0) \subseteq \overline{E}_N$. Since $\overline{E}_N$ is symmetric, we have $\mathrm{B}_r(-x_0) \subseteq \overline{E}_N$. Then for each $x \in \mathrm{B}_r(0)$, we have

$$x = \frac{1}{2}(x - x_0) + \frac{1}{2}(x + x_0)$$

where $x - x_0 \in \mathrm{B}_r(-x_0) \subseteq \overline{E}_N$ and $x + x_0 \in \mathrm{B}_r(x_0) \subseteq \overline{E}_N$. Since $\overline{E}_N$ is convex, it follows that $x \in \overline{E}_N$. Therefore $\mathrm{B}_r(0) \subseteq \overline{E}_N$.

Step 2: We will show $\mathrm{B}_r(0) \subseteq E_N$. Let $x \in \mathrm{B}_r(0)$, let $\rho > 0$ such that $\|x\| < \rho < r$, let $q > 0$ such that $q < 1 - \rho/r$, and let $y = (r/\rho)x$. Then observe that $y \in \mathrm{B}_r(0) \subseteq \overline{E}_N$. In particular, this implies $\mathrm{B}_{qr}(y) \cap E_N \neq \emptyset$, so we can choose $y_0 \in \mathrm{B}_{qr}(y) \cap E_N$. Since $y_0 \in \mathrm{B}_{qr}(y)$, we have

$$\|y_0 - y\| < qr$$

In other words, dividing both sides by q give us $(y - y_0)/q \in \mathrm{B}_r(0) \subseteq \overline{E}_N$. In particular, this implies $\mathrm{B}_{qr}((y - y_0)/q) \cap E_N \neq \emptyset$, so we can choose $y_1 \in \mathrm{B}_{qr}((y - y_0)/q) \cap E_N$. Again since $y_1 \in \mathrm{B}_{qr}((y - y_0)/q)$, we have

$$\left\| \frac{y - y_0 - qy_1}{q} \right\| < qr$$

In other words, dividing both sides by q gives us $(y - y_0 - qy_1)/q^2 \in \mathrm{B}_r(0) \subseteq \overline{E}_N$. In particular, this implies $\mathrm{B}_{qr}((y - y_0 - qy_1)/q^2) \cap E_N \neq \emptyset$, so we can choose $y_2 \in \mathrm{B}_{qr}((y - y_0 - qy_1)/q^2) \cap E_N$. More generally, for each $n \geq 2$, we choose

$$y_n \in \mathrm{B}_{qr}\left(\frac{y - y_0 - qy_1 - \cdots - q^{n-1}y_{n-1}}{q^n} \right).$$

In this case, we obtain a sequence $(y_n) \subseteq E_N$ such that

$$\|y - y_0 - qy_1 - q^2y_2 - \cdots - q^n y_n\| < q^n r \tag{9}$$

for all $n \in \mathbb{N}$. Since $\|y_n\| \leq r + qr$ for all $n \in \mathbb{N}$ and $0 < q < 1$, we have $\sum_{n=0}^{\infty} q^n y_n$ is absolutely convergent. Therefore by (9) we have $y = \sum_{n=1}^{\infty} q^n y_n$. Thus\

$$\begin{aligned}
p(x) &= p\left(\frac{\rho}{r} y\right) \\
&= \frac{\rho}{r} p(y) \\
&= \frac{\rho}{r} p\left(\sum_{n=1}^{\infty} q^n y_n\right) \\
&\leq \frac{\rho}{r} q^n \sum_{n=1}^{\infty} p(y_n) \\
&\leq \frac{\rho}{r} q^n N \\
&= \frac{\rho}{r} \frac{N}{1-q} \\
&\leq N.
\end{aligned}$$

It follows that $\mathrm{B}_r(0) \subseteq E_N$.

Step 3: Let $x \in \mathcal{X}$ be arbitrary nonzero. Then $(r/2)x/\|x\| \in \mathrm{B}_r(0)$ and hence $p((r/2)x/\|x\|) \leq N$. This implies $p(x) \leq (2N/r)\|x\|$. $\square$

Remark 6. We make two remarks.

1. Zabreiko's lemma implies p is continuous. Indeed, suppose $x_n \to x$. Then
$$\begin{aligned}|p(x_n) - p(x)| &\leq p(x_n - x)\\ &\leq C\|x_n - x\|\\ &\to 0.\end{aligned}$$

2. Zabreiko's lemma can be used to prove the uniform boundedness principle. Indeed, take $p(x) = \sup_{T\in\mathcal{A}} \|Tx\|$. Then it can be shown that p satisfies the properties from Zabreiko's lemma. Therefore there exist $C > 0$ such that
$$\sup_{T\in\mathcal{A}} \|Tx\| \leq C\|x\|.$$
Thus for any $T \in \mathcal{A}$ we have $\|Tx\| \leq C\|x\|$ which implies $\|T\| \leq C$ for all $T \in \mathcal{A}$.

4.4 Proof of Main Theorem

We now wish to prove Theorem (4.1).

Proof. Let $p \colon \mathcal{Y} \to [0, \infty)$ be defined by
$$p(y) := \inf\{\|x\| \mid (x, y) \in \mathcal{Z}\}.$$

It's easy to show that p is a seminorm. We claim that it is also countably subadditive. Indeed, let $\sum_{n=1}^{\infty} y_n$ be an absolutely convergent series such that $\sum_{n=1}^{\infty} p(y_n) < \infty$. Let $\varepsilon > 0$ and for each $n \in \mathbb{N}$ choose $x_n \in \mathcal{X}$ such that $\|x_n\| < p(y_n) + \varepsilon/2^n$ and $(x_n, y_n) \in \mathcal{Z}$. Then
$$\sum_{n=1}^{\infty} \|x_n\| \leq \sum_{n=1}^{\infty} p(y_n) + \varepsilon < \infty.$$

Hence $\sum_{n=1}^{\infty} x_n$ is absolutely convergent. Since $\mathcal{Z}$ is a subspace, we have $(\sum_{n=1}^{N} x_n, \sum_{n=1}^{N} y_n) \in \mathcal{Z}$ for all $N \in \mathbb{N}$. Since $\mathcal{Z}$ is closed, we have $(\sum_{n=1}^{\infty} x_n, \sum_{n=1}^{\infty} y_n) \in \mathcal{Z}$. Then
$$\begin{aligned}p\left(\sum_{n=1}^{\infty} y_n\right) &= \inf\left\{\|x\| \mid \left(x, \sum_{n=1}^{\infty} y_n\right) \in \mathcal{Z}\right\}\\ &\leq \left\|\sum_{n=1}^{\infty} x_n\right\|\\ &\leq \sum_{n=1}^{\infty} ||x_n||\\ &\leq \sum_{n=1}^{\infty} p(y_n) + \varepsilon.\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, it follows that p is countably subadditive. So we can apply Zabreiko's lemma to obtain that p is continuous.

Now let $U = \mathrm{B}_1(0)$ be the open unit ball in $\mathcal{Y}$. Then
$$\begin{aligned}\pi_2(\pi^{-1}(\mathrm{B}_1(0) \cap \mathcal{Z})) &= \pi_2\{(x, y) \mid x \in \mathrm{B}_1(0) \text{ and } (x, y) \in \mathcal{Z})\\ &= \{y \in \mathcal{Y} \mid p(y) < 1\}\\ &= \{p < 1\}.\end{aligned}$$

Implies $\pi_2(\pi^{-1}(\mathrm{B}_1(0) \cap \mathcal{Z}))$ is open since p is continuous. The general case open sets U can be easily be obtained using linearity and homogeneity. □

5 Hilbert Space Applications

Let $\mathcal{H}$ be a Hilbert space and let $\mathcal{K}$ and $\mathcal{L}$ be closed subspaces of $\mathcal{H}$. We ask, is $\mathcal{K} + \mathcal{L}$ a closed subspace?

Proposition 5.1. *The following are equivalent:*

1. $\mathcal{K} \cap \mathcal{L} = (\mathcal{K}^\perp + \mathcal{L}^\perp)^\perp$
2. $\mathcal{K}^\perp \cap \mathcal{L}^\perp = (\mathcal{K} + \mathcal{L})^\perp$
3. $(\mathcal{K} \cap \mathcal{L})^\perp = \overline{\mathcal{K}^\perp + \mathcal{L}^\perp}$
4. $(\mathcal{K}^\perp \cap \mathcal{L}^\perp)^\perp = \overline{\mathcal{K} + \mathcal{L}}$

Proof. 1 implies 2, 1 implies 3, and 2 implies 4 are easy. It suffices to show 1. Let $x \in \mathcal{K} \cap \mathcal{L}$ and let $y \in \mathcal{K}^\perp + \mathcal{L}^\perp$. Then $y = z + w$ where $z \in \mathcal{K}^\perp$ and $w \in \mathcal{L}^\perp$. So $\langle x, y\rangle = \langle x, z + w\rangle = \langle x, z\rangle + \langle x, w\rangle$. Therefore $x \perp \mathcal{K}^\perp + \mathcal{L}^\perp$ and hence $x \in (\mathcal{K}^\perp + \mathcal{L}^\perp)^\perp$. Thus $\mathcal{K} \cap \mathcal{L} \subseteq (\mathcal{K}^\perp + \mathcal{L}^\perp)^\perp$. Conversely, we have $(\mathcal{K}^\perp + \mathcal{L}^\perp)^\perp \subseteq (\mathcal{K}^\perp)^\perp = \mathcal{K}$ and $(\mathcal{K}^\perp + \mathcal{L}^\perp)^\perp \subseteq (\mathcal{L}^\perp)^\perp = \mathcal{L}$. Thus $\mathcal{K} \cap \mathcal{L} \supseteq (\mathcal{K}^\perp + \mathcal{L}^\perp)^\perp$. □

Lemma 5.1. *Assume $\mathcal{K}$ and $\mathcal{L}$ are closed subspaces of a Hilbert space $\mathcal{H}$ and assume $\mathcal{K} + \mathcal{L}$ is closed. Then there exists a constant $C > 0$ such that every $z \in \mathcal{K} + \mathcal{L}$ there exists $x \in \mathcal{K}$ and $y \in \mathcal{L}$ such that $z = x + y$ and $\|x\| \leq C\|z\|$ and $\|y\| \leq C\|z\|$.*

Proof. Consider $\mathcal{K} \times \mathcal{L} \subseteq \mathcal{H} \times \mathcal{H}$. Then $\mathcal{K} \times \mathcal{L}$ is a closed subspace of the Banach space $\mathcal{H} \times \mathcal{H}$. Hence it is a Banach space itself. Consider the map $T \colon \mathcal{K} \times \mathcal{L} \to \mathcal{K} + \mathcal{L}$ given by

$$T((x, y)) = x + y.$$

Then T is a bounded linear operator. Furthermore, it is easy to see that T is surjective. By the open mapping theorem the image $T(\mathrm{B}_1(0))$ must be open. Since $0 \in T(\mathrm{B}_1(0))$ there exists $c > 0$ such that $\mathrm{B}_c(0) \subseteq T(\mathrm{B}_1(0))$. This means for all $z \in \mathcal{K} + \mathcal{L}$ with $\|z\| < c$ we have $z \in T(\mathrm{B}_1(0))$, that is, there exists $x \in \mathcal{K}$ and $y \in \mathcal{L}$ such that $z = x + y$ and $\|(x, y)\| = \|x\| + \|y\| < 1$. Now by scaling, for any $z \in \mathcal{K} + \mathcal{L}$ we have $\|(c/2)z/\|z\|\| < c$. Therefore there exists $x' \in \mathcal{K}$ and $y' \in \mathcal{L}$ such that $(c/2)z/\|z\| = x' + y'$ with $\|x'\| + \|y'\| < 1$. Then set $x = (2/c)\|z\|x'$ and $y = (2/x)\|z\|y'$. We have

$$\|x\| + \|y\| < \frac{2}{c}\|z\|.$$

□

Proposition 5.2. *$\mathcal{K} + \mathcal{L}$ is closed if and only if $\mathcal{K}^\perp + \mathcal{L}^\perp$ is closed.*

Proof. It's enough to show ($\Longrightarrow$) with the other being a simple consequence of this one. Assume $\mathcal{K} + \mathcal{L}$ is closed. Using the previous proposition, we have $\overline{\mathcal{K}^\perp + \mathcal{L}^\perp} = (\mathcal{K} \cap \mathcal{L})^\perp$ so it is enough to show that $(\mathcal{K} \cap \mathcal{L})^\perp \subseteq \mathcal{K}^\perp + \mathcal{L}^\perp$. Let $y \in (\mathcal{K} \cap \mathcal{L})^\perp$. Consider $\ell \colon \mathcal{K} + \mathcal{L} \to \mathbb{R}$ defined by $\ell(x) = \langle a, y\rangle$ where $a \in \mathcal{K}$ is such that $x = a + b$ and $b \in \mathcal{L}$. To see that ℓ is well-defined, suppose $x = a' + b'$ where $a' \in \mathcal{K}$ and $b' \in \mathcal{L}$. Then $a - a' = b - b'$. It follows that $b - b' \in \mathcal{K} \cap \mathcal{L}$. Hence

$$\begin{aligned} \langle a, y\rangle &= \langle b - b' + a', y\rangle \\ &= \langle a', y\rangle + \langle b - b', y\rangle \\ &= \langle a', y\rangle. \end{aligned}$$

Thus ℓ is well-defined. It is easy to see that ℓ is linear. Furthermore, we claim ℓ is bounded. By the previous lemma, there exists $C > 0$ such that for any $x \in \mathcal{K} + \mathcal{L}$ there exists a decomposition $x = a + b$ where $a \in \mathcal{K}$ and $b \in \mathcal{L}$ such that $\|a\| \leq C\|x\|$ and $\|b\| \leq C\|x\|$. Then

$$\begin{aligned} |\ell(x)| &= |\langle a, y\rangle| \\ &\leq \|a\|\|y\| \\ &\leq C\|y\|\|x\|. \end{aligned}$$

Thus ℓ is a bounded linear functional.

We extend ℓ to the whole $\mathcal{H}$ by setting

$$\widetilde{\ell}(x) = \begin{cases} \ell(x) & \text{if } x \in \mathcal{K} + \mathcal{L} \\ 0 & \text{if } x \in (\mathcal{K} + \mathcal{L})^\perp. \end{cases}$$

This is still a bounded linear functional. So by the Riesz representation theorem for Hilbert spaces, there exists some $z \in \mathcal{H}$ such that $\widetilde{\ell}(x) = \langle x, z\rangle$ for all $x \in \mathcal{H}$. Then $y = (y - z) + z$. For any $k \in \mathcal{K}$ we have $\ell(k) = \widetilde{\ell}(k)$. In particular, $y - z \in \mathcal{K}^\perp$. Furthermore, note that $\ell|_{\mathcal{L}} = 0$. Indeed, if $x \in \mathcal{L}$ then we use the decomposition $0 + x = x$ to get $\ell(x) = \langle 0, y\rangle = 0$. Thus $\widetilde{\ell}|_{\mathcal{L}} = \widetilde{\ell}|_{\mathcal{L}} = 0$ and hence $z \in \mathcal{L}^\perp$. Therefore we see that $(\mathcal{K} \cap \mathcal{L})^\perp \subseteq \mathcal{K}^\perp + \mathcal{L}^\perp$. □

Remark 7. The same results holds for all reflexive Banach spaces.

5.0.1 Ker T Star Equals Im T Perp

Proposition 5.3. *Let $T\colon \mathcal{H} \to \mathcal{H}$ be a bounded operator. Then the following are true.*

1. $\ker T = (\operatorname{im} T^*)^\perp$.
2. $\ker T^* = (\operatorname{im} T)^\perp$.
3. $(\ker T)^\perp = \overline{\operatorname{im} T^*}$.
4. $(\ker T^*)^\perp = \overline{\operatorname{im} T}$.

Proof. Since identities 2-4 are simple consequences of 1, we will just prove 1 and leave the rest as an exercise. Consider the Hilbert space $\mathcal{H} \times \mathcal{H}$ with an inner products defined by

$$\langle (x_1, y_1), (x_2, y_2) \rangle = \langle x_1, x_2 \rangle + \langle y_1, y_2 \rangle.$$

Let $\mathcal{K} = \{(x, Tx) \mid x \in \mathcal{H}\}$ and $\mathcal{L} = \mathcal{H} \times 0$. Then $\mathcal{K}$ and $\mathcal{L}$ are both closed subspaces of $\mathcal{H} \times \mathcal{H}$. Observe that $\mathcal{K} \cap \mathcal{L} = \ker T \times 0$ and $\mathcal{K} + \mathcal{L} = \mathcal{H} \times \operatorname{im} T$. Also note that $\mathcal{L}^\perp = 0 \times \mathcal{H}$ and

$$\begin{aligned} \mathcal{K}^\perp &= \{(x_1, y_1) \mid \langle x_1, x \rangle + \langle y_1, Tx \rangle = 0 \text{ for all } x \in \mathcal{H}\} \\ &= \{(x_1, y_1) \mid \langle x_1 + T^* y_1, x \rangle = 0 \text{ for all } x \in \mathcal{H}\} \\ &= \{(x_1, y_1) \mid x_1 + T^* y_1 = 0\} \\ &= \{(-T^* y_1, y_1) \mid y_1 \in \mathcal{H}\}. \end{aligned}$$

Thus $\mathcal{K}^\perp \cap \mathcal{L}^\perp = 0 \times \ker T^*$ and $\mathcal{K}^\perp + \mathcal{L}^\perp = \operatorname{im} T^* \times \mathcal{H}$. Thus

$$\begin{aligned} \ker T \times 0 &= \mathcal{K} \cap \mathcal{L} \\ &= (\mathcal{K}^\perp + \mathcal{L}^\perp)^\perp \\ &= (\operatorname{im} T^* \times \mathcal{H})^\perp \\ &= (\operatorname{im} T^*)^\perp \times 0. \end{aligned}$$

It follows that $\ker T = (\operatorname{im} T^*)^\perp$. □

Proposition 5.4. *Let $T\colon \mathcal{H} \to \mathcal{H}$ be a bounded linear operator. Then* $\operatorname{im} T$ *is a closed subspace if and only if* $\operatorname{im} T^*$ *is a closed subspace.*

Proof. Uses open mapping theorem. □

5.1 Characterizing Surjectivity of a Bounded Operator

From Proposition (5.3), we see that T is injective if and only if T^* has dense image. Also T is surjective if and only if T^* is injective and $\operatorname{im} T^*$ is closed. Let's state this as a theorem.

Theorem 5.2. *Let $T\colon \mathcal{H} \to \mathcal{H}$ be a bounded operator. Then the following are equivalent.*

1. *T is surjective.*
2. *There exists $c > 0$ such that $\|T^* x\| \geq c\|x\|$ for all $x \in \mathcal{H}$.*
3. *T^* is injective and* $\operatorname{im} T^*$ *is closed.*

Proof. (1 implies 2) Suppose T is surjective. Let $E = \{x \in \mathcal{H} \mid \|T^* x\| \leq 1\}$. For any $x \in E$ and $z \in \mathcal{H}$ such that $Tz = y$, we have

$$\begin{aligned} |\langle x, y \rangle| &= |\langle x, Tz \rangle| \\ &= |\langle T^* x, z \rangle| \\ &\leq \|T^* x\| \|z\| \\ &\leq \|z\|. \end{aligned}$$

So the set E is weakly bounded. In fact, by uniform boundedness principle, E is bounded. Therefore there exists $C > 0$ such that $\|z\| \leq C$ for all $x \in E$. In other words, if $\|T^* x\| \leq 1$, then $\|x\| \leq C$.

Now let $x \in \mathcal{H}$ be any arbitrary nonzero vector. Since T^* is injective, we have $T^*x \neq 0$. Consider $y = x/\|T^*x\|$. Then

$$\begin{aligned} \|T^*y\| &= \left\| T^* \left(\frac{x}{\|T^*x\|} \right) \right\| \\ &= \frac{1}{\|T^*x\|} \|T^*x\| \\ &= 1, \end{aligned}$$

and hence $y \in E$. In particular, $\|y\| \leq C$. It follows that $\|x\| \leq C\|T^*x\|$. In other words

$$c\|x\| \leq \|T^*x\|$$

for all $x \neq 0$ where $c = 1/C > 0$.

(2 implies 3) Suppose there exists $c > 0$ such that $\|T^*x\| \geq c\|x\|$ for all $x \in \mathcal{H}$. Now let $x \in \ker T^*$. Then $\|T^*x\| = 0$ which implies $\|x\| = 0$ which implies $x = 0$. Thus T^* is injective. Now let (y_n) be a convergent sequence in $\operatorname{im} T^*$ which converges to $y \in \mathcal{H}$. For each $n \in \mathbb{N}$ choose $x_n \in \mathcal{H}$ such that $y_n = T^*x_n$. Then observe for all $n \in \mathbb{N}$ we have

$$\begin{aligned} \|x_m - x_n\| &\leq \frac{1}{c} \|T^*(x_m - x_n)\| \\ &= \frac{1}{c} \|y_m - y_n\|. \end{aligned}$$

Thus since (y_n) is a Cauchy sequence, it follows that (x_n) is a Cauchy sequence. Since $\mathcal{H}$ is a Hilbert space, we see that (x_n) is convergent, say $x_n \to x$. Then since T^* is bounded/continuous, we see that $T^*x = y$. Thus $\operatorname{im} T^*$ is closed.

(3 implies 1) Suppose T^* is injective and $\operatorname{im} T^*$ is closed. Since T^* is injective, we see that $\operatorname{im} T$ is dense in $\mathcal{H}$. Since $\operatorname{im} T^*$ is closed, it follows that $\operatorname{im} T$ is closed (this depends on the open mapping theorem). Therefore $\operatorname{im} T = \mathcal{H}$, and so T is surjective. □

5.1.1 Quasi Inner-Product

Let $\mathcal{H}$ be a Hilbert space. Suppose $B \colon \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ satisfies

1. B is linear in the first coordinate and conjugate linear in the second coordinate.
2. There exists $C > 0$ such that $|B(x, y)| \leq C\|x\|\|y\|$.
3. There exists $c > 0$ such that $|B(x, x)| \geq c\|x\|^2$.

Then there exists $T \colon \mathcal{H} \to \mathcal{H}$ invertible such that $B(x, y) = \langle Tx, y \rangle$ for all $x, y \in \mathcal{H}$. Equivalently for any bounded linear functional $\ell \colon \mathcal{H} \to \mathbb{C}$ there exists a unique $z \in \mathcal{H}$ such that $\ell(x) = B(x, z)$ for all $x \in \mathcal{H}$.