THE HAHN-BANACH SEPARATION THEOREM AND OTHER SEPARATION RESULTS

ROBERT PENG

ABSTRACT. This paper will introduce and prove several theorems involving the separation of convex sets by hyperplanes, along with other interesting related results. It will begin with some basic separation results in $\mathbf{R}^n$, such as the Hyperplane Separation Theorem of Hermann Minkowski, and then it will focus on and prove the extension of this theorem into normed vector spaces, known as the Hahn-Banach Separation Theorem. This paper will also prove some supporting results as stepping stones along the way, such as the Supporting Hyperplane Theorem and the analytic Hahn-Banach Theorem.

CONTENTS

INTRODUCTION

The intent of this paper is to introduce and prove several results relating to the separation of sets by hyperplanes, terms that will be formally defined later. Intuitively, the reader can think about drawing a line between two disjoint subsets of $\mathbf{R}^2$ such that the two sets are on different sides of the line, and thinking about the conditions that will make this possible. Results that will be proved include the Hyperplane Separation Theorem in $\mathbf{R}^n$, as well as its generalization into normed vector spaces, known as the Hahn-Banach Separation Theorem. In addition to this topic being interesting in a mathematical perspective, it is also particularly useful due to its many applications, such as in optimization problems. This paper will assume that the reader is familiar with a few basic concepts (such as that of a vector space) and with some fundamental analysis results, but at least a cursory definition will be given for most of the terms and concepts used in the paper, so little background is needed.

Date: 18 August, 2014.

1. Convex Sets

Before we can proceed to results about separation by hyperplanes, we need to define what convex sets are and establish some basic results about them, since convexity is the most important property that determines whether sets are separable.

Since we will define convexity for sets in a vector space, we will first recall that a vector space is a set V equipped with two operations, vector addition and scalar multiplication, that satisfy the eight vector space axioms (the scalars are members of a field F, usually taken to be the real or complex numbers). For the purposes of this paper, the scalar field for a vector space will always be assumed to be the real numbers; in this case, we have what is called a *real vector space*.

Also, it will be useful to define a normed vector space here, which establishes a notion of length for a vector space:

Definition 1.1. A real *normed vector space* is a real vector space V, endowed with a nonnegative real function $\|\cdot\|$, called a norm, which satisfies the following properties:

(1) $\|v\| \geq 0$ for all $v \in V$, $\|v\| = 0$ if and only if $v = 0$
(2) $\|\alpha v\| = |\alpha|\|v\|$ for all $\alpha \in \mathbf{R}$, $v \in V$
(3) $\|u + v\| \leq \|u\| + \|v\|$ for all u, $v \in V$

It is important to note that the norm induces a notion of distance: the distance between two vectors u and v in a normed vector space V is given by $\|u - v\|$.

Now, we may proceed to the definition of a convex set:

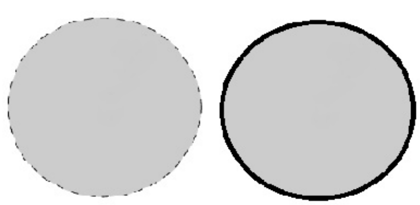
Definition 1.2. A set A in a real vector space V is said to be *convex* if for all x, $y \in A$ and all $t \in [0, 1]$, the point $tx + (1 - t)y \in A$ as well.

In other words, if x and y are in a convex set, then any point in the line segment connecting x and y will be in the set as well.

The following are several significant examples of convex sets:

Example 1.3.

- The open ball of radius r centered at x_0 in a normed vector space V, denoted $B_r(x_0)$, is a convex set. More formally, $B_r(x_0) = \{x \in V \mid \|x - x_0\| < r\}$ is a convex set.
- Similarly, the closed ball of radius r centered at x_0 in a normed vector space V, $\overline{B_r(x_0)}$, is a convex set (the bar overhead denotes the closure). More formally, $\overline{B_r(x_0)} = \{x \in V \mid \|x - x_0\| \leq r\}$ is a convex set.



Examples of open and closed balls in $\mathbf{R}^2$

It will also be useful to introduce the following definition here, as it plays a large role in convexity and separation results:

Definition 1.4. By the *Minkowski sum* (or simply *sum*) of two sets A and B in a vector space V, we mean the set $A+B$ produced by adding each vector in A to each vector in B. In other words,

$$A + B = \{a + b \mid a \in A,\ b \in B\}$$

The *Minkowski difference* (or simply *difference*) of two sets A and B is defined similarly:

$$A - B = \{a - b \mid a \in A,\ b \in B\}$$

$A+B$ and $A-B$ are clearly subsets of the same vector space as A and B. Now, we shall establish an important lemma about convex sets that will be used later on:

Lemma 1.5. *If A and B are convex subsets of a vector space V, then $A+B$ is convex.*

Proof. Let x and y be two elements in $A+B$.
Then $x = a_1 + b_1$ for $a_1 \in A$, $b_1 \in B$, and $y = a_2 + b_2$ for $a_2 \in A$, $b_2 \in B$
So then for all $t \in [0,1]$,

$$\begin{aligned} tx + (1-t)y &= t(a_1 + b_1) + (1-t)(a_2 + b_2) \\ &= ta_1 + tb_1 + (1-t)a_2 + (1-t)b_2 \\ &= \underbrace{[ta_1 + (1-t)a_2]}_{\in A} + \underbrace{[tb_1 + (1-t)b_2]}_{\in B} \end{aligned}$$

So for all $x,\ y \in A+B$ and all $t \in [0,1]$, we have that $tx + (1-t)y \in A+B$, meaning that the set is convex. □

Note that $A - B = A + (-B)$, where the set $-B$ is defined in the obvious way, so $A - B$ is convex as well if A and B are both convex.

2. Closed and Compact Sets

Now, we shall restrict our attention to the the n-dimensional real coordinate space $\mathbf{R}^n$, which is certainly a vector space. The following definitions and results will serve as a setup for the separation results we want to show in $\mathbf{R}^n$.

Definition 2.1. We say that a set A in $\mathbf{R}^n$ is *closed* if every convergent sequence $(a_n)_{n\in\mathbf{N}} \subseteq A$ has a limit a that is also in A. In other words, a *closed* set is one such that all limit points are contained in the set.

Definition 2.2. A set A in $\mathbf{R}^n$ is *bounded* if it is contained inside some ball of finite radius.

The definition of a general compact set can be rather opaque and will not be very insightful for the purposes of this paper. Instead, I offer (without proof) a famous result that clarifies what a compact set is in $\mathbf{R}^n$:

Theorem 2.3 (Heine-Borel Theorem)**.** *A subset A of $\mathbf{R}^n$ is compact if and only if it is both closed and bounded.*

Then, by the Bolzano-Weierstrass Theorem, which states that every bounded sequence in $\mathbf{R}^n$ has a convergent subsequence, we have that for a compact subset A of $\mathbf{R}^n$, every sequence has a convergent subsequence, whose limit is in A. For a subset A of $\mathbf{R}^n$, this characterization is often taken as the definition of compactness.

We shall prove the next lemma only for $\mathbf{R}^n$ (although this lemma is also true for any normed vector space).

Lemma 2.4. *Let A and B be subsets of $\mathbf{R}^n$. If A is closed and B is compact, then $A + B$ is closed.*

Proof. Let $X = A+B$. We wish to show that for any convergent sequence $(x_n) \subseteq X$, the limit point is in X as well. Assume that (x_n) converges to a point $x \in \mathbf{R}^n$. Consider $(a_n) \subseteq A$ and $(b_n) \subseteq B$ such that $(x_n) = (a_n)+(b_n)$. By the compactness of B, (b_n) has a subsequence (b_{n_k}) such that (b_{n_k}) converges to a limit point $b \in B$. Now, consider $(a_{n_k}) = (x_{n_k}) - (b_{n_k})$, which converges to $x - b$, as every convergent subsequence converges to the same limit as its sequence (meaning (x_{n_k}) converges to x). Because A is closed, $x - b$ is a point in A. Finally, consider $x = (x - b) + b$, which is a point in the set X, as it is the sum of two elements, one in A and one in B. Thus, (x_n) converges to a point $x \in X$. This implies that $X = A + B$ is closed. □

This also implies that $A - B$ is closed if A is closed and B is compact.

It will be useful to also define the closure of a set, which gives us the 'smallest' closed set that contains the given set, as well as some other related terms:

Definition 2.5.

- The *closure* of a set A in $\mathbf{R}^n$, denoted $\overline{A}$, is the set A including its limit points.
- The *boundary* of a set A in $\mathbf{R}^n$, denoted ∂A, is the set of points that are members of both the closure of A and the closure of A^c, the set complement of A. In other words, $\partial A = \overline{A} \cap \overline{A^c}$.
- The *interior* of a set A in $\mathbf{R}^n$, denoted $\mathring{A}$, is the set A excluding its boundary. In other words, $\mathring{A} = A \setminus \partial A$.

Now, the next lemma will establish that the closure of a convex set retains its convexity:

Lemma 2.6. *Let A be a subset of $\mathbf{R}^n$. If A is convex, then $\overline{A}$ is also convex.*

Proof. Let x, $y \in \overline{A}$. By the definition of closure, there exists sequences (x_n), $(y_n) \subseteq A$ with limits x and y, respectively. Now, because A is convex, we know that for each x_i and y_i, and for all $t \in [0, 1]$, we have $tx_i + (1-t)y_i \in A$. This establishes the sequence $(tx_n + (1-t)y_n) \subseteq A$, which is convergent by the convergence of (x_n) and (y_n). So we have that $tx + (1 - t)y \in \overline{A}$, as $\overline{A}$ includes all limit points of A. This establishes that $\overline{A}$ is convex. □

We need one more important result about compact sets before we can finally establish results about separation of sets in $\mathbf{R}^n$. I shall present it without proof:

Theorem 2.7 (Extreme Value Theorem)**.** *Every continuous real-valued function on a compact set attains its extreme values on that set.*

3. Hyperplane Separation in $\mathbf{R}^n$

Now, we finally have the tools to establish some separation results in $\mathbf{R}^n$ with the standard Euclidean norm, which is of course a normed vector space. Our primary means of separation here will be through hyperplanes, which I now define:

Definition 3.1. Let $p \in \mathbf{R}^n$ be nonzero, and let $c \in \mathbf{R}$ be constant. The set $\mathcal{H}(p, c)$ consisting of all vectors $x \in \mathbf{R}^n$ such that

$$p \cdot x = p_1x_1 + p_2x_2 + ... + p_nx_n = c$$

is a subset of $\mathbf{R}^{\mathrm{n}}$ called a *hyperplane*. More concisely, a *hyperplane* $\mathcal{H}(p,c)$ of $\mathbf{R}^{\mathrm{n}}$ is the set $\{x \in \mathbf{R}^{\mathrm{n}} \mid p \cdot x = c\}$.

Hyperplanes according to this definition have one dimension less than the space they reside in, so hyperplanes in $\mathbf{R}^2$ are lines, while hyperplanes in $\mathbf{R}^3$ are planes. Additionally, they divide $\mathbf{R}^{\mathrm{n}}$ into two half-spaces, which I denote

$$\mathcal{O}(p,c) = \{x \in \mathbf{R}^{\mathrm{n}} \mid p \cdot x > c\}$$

$$\mathcal{P}(p,c) = \{x \in \mathbf{R}^{\mathrm{n}} \mid p \cdot x < c\}$$

We can also consider the closure of the two half-spaces, referred to as the closed half-spaces:

$$\overline{\mathcal{O}(p,c)} = \{x \in \mathbf{R}^{\mathrm{n}} \mid p \cdot x \geq c\}$$

$$\overline{\mathcal{P}(p,c)} = \{x \in \mathbf{R}^{\mathrm{n}} \mid p \cdot x \leq c\}$$

Lastly, it is helpful to note that p can be thought of as as a vector that is orthogonal to the hyperplane at every point.

One more definition is necessary to formalize the notion of separation:

Definition 3.2.

- We say that two nonempty sets A and B in $\mathbf{R}^{\mathrm{n}}$ are *separated* by a hyperplane if there exists a $p \in \mathbf{R}^{\mathrm{n}}$ nonzero and a constant $c \in \mathbf{R}$ such that $p \cdot a \geq c \geq p \cdot b$ for all $a \in A$ and all $b \in B$.
- We say that two nonempty sets A and B in $\mathbf{R}^{\mathrm{n}}$ are *strictly separated* by a hyperplane if there exists a $p \in \mathbf{R}^{\mathrm{n}}$ nonzero and a constant $c \in \mathbf{R}$ such that $p \cdot a > c > p \cdot b$ for all $a \in A$ and all $b \in B$.

The above definitions establish that A and B either lie in different $\overline{\mathcal{O}(p,c)}$ and $\overline{\mathcal{P}(p,c)}$ (in the case of separation) or in different $\mathcal{O}(p,c)$ and $\mathcal{P}(p,c)$ (in the case of strict separation).

Now, we shall establish and prove our first separation result:

Lemma 3.3. *Let A be a nonempty, closed, and convex subset of $\boldsymbol{R}^n$. Let x_0 be a point in $\boldsymbol{R}^n$ such that $x_0 \notin A$. Then there exists an $a_0 \in A$ and $p \in \boldsymbol{R}^n$ nonzero such that $p \cdot a \geq c = p \cdot a_0 > p \cdot x_0$ for all $a \in A$.*

Note that the separation established in the lemma is stronger than separation, but not quite as strong as strict separation.

Proof. Let $a_0 \in A$ be a point that realizes the minimum distance from any point in A to x_0. That is, $a_0 \in A$ is a point such that for all $a \in A$, $\|x_0 - a_0\| \leq \|x_0 - a\|$.

We must first show that such a point exists:
First, note that the standard Euclidean norm $\|\cdot\|$ is continuous. Now, note that A is nonempty, so there is a point $z \in A$. Let $Z = A \cap \overline{B_{\|z-x_0\|}(x_0)}$, where $\overline{B_{\|z-x_0\|}(x_0)}$ is the closed ball of radius $\|z - x_0\|$ centered at x_0. Z is clearly closed and bounded, so Z is compact. Also, note that $x_0 \notin Z$. Thus, we can use the Extreme Value Theorem to find a point in Z that minimizes the distance from x_0 to any point in Z (in other words, the norm function attains a minimum in Z). Call this point a_0. Finally, note that this construction causes all points in $A \setminus Z$ to be not closer to x_0 than any point in Z (or causes $A \setminus Z$ to be empty), so a_0 is also a point that minimizes the distance from x_0 to the entire set A i.e. a_0 is a point such that $\|x_0 - a_0\| \leq \|x_0 - a\|$ for all $a \in A$.

Let $p = a_0 - x_0$ and $c = p \cdot a_0$. We first note that p is nonzero because $x_0 \notin A$, but $a_0 \in A$. We need to show that these values satisfy $p \cdot x_0 < c$ and $p \cdot a \geq c$ for all $a \in A$. The first inequality is easier to show:

$$\begin{aligned} p \cdot x_0 &= p \cdot x_0 - p \cdot a_0 + p \cdot a_0 \\ &= p \cdot (x_0 - a_0) + p \cdot a_0 \\ &= p \cdot (-p) + p \cdot a_0 \\ &= -\|p\|^2 + p \cdot a_0 \\ &< p \cdot a_0 = c \end{aligned}$$

The last line follows from the fact that because p is nonzero, $-\|p\|^2$ is necessarily negative.

Finally, we need to show that $p \cdot a \geq c$ for all $a \in A$, a harder task:
By the convexity of A, we know that $w = ta + (1-t)a_0 \in A$ for all $t \in [0, 1]$ and all $a \in A$. So now we consider

$$\begin{aligned} \|x_0 - a_0\|^2 - \|x_0 - w\|^2 &= \|x_0 - a_0\|^2 - \|x_0 - (ta + (1-t)a_0)\|^2 \\ &= \|x_0 - a_0\|^2 - \|x_0 - ta - a_0 + ta_0\|^2 \\ &= (x_0 - a_0) \cdot (x_0 - a_0) - \\ &\quad [(x_0 - a_0) + t(a_0 - a)] \cdot [(x_0 - a_0) + t(a_0 - a)] \\ &= (x_0 - a_0) \cdot (x_0 - a_0) - [(x_0 - a_0) \cdot (x_0 - a_0) + \\ &\quad 2t(a_0 - a) \cdot (x_0 - a_0) + t^2(a_0 - a) \cdot (a_0 - a)] \\ &= -2t(a_0 - a) \cdot (x_0 - a_0) - t^2(a_0 - a) \cdot (a_0 - a) \\ &= -2t(a_0 - a) \cdot (-p) - t^2\|a_0 - a\|^2 \\ &= t[2p \cdot (a_0 - a) - t\|a_0 - a\|^2] \end{aligned}$$

Now, we will show by contradiction that it cannot be true that $p \cdot a < c = p \cdot a_0$, meaning that $p \cdot a \geq c$. Assume by contradiction that $p \cdot a < c$. Then we have that

$$\begin{aligned} p \cdot (a_0 - a) &= p \cdot a_0 - p \cdot a \\ &= c - \underbrace{p \cdot a}_{< c} \\ &> 0 \end{aligned}$$

So then continuing with the manipulations above, we have that for t sufficiently small, it is certainly true that $2p \cdot (a_0 - a) > t\|a_0 - a\|^2$, which then implies that $\|x_0 - a_0\|^2 - \|x_0 - w\|^2 > 0$ for t sufficiently small. This in turn implies that $\|x_0 - a_0\| > \|x_0 - w\|$. However, this indicates that the distance from w to x_0 is less than then distance from a_0 to x_0, which cannot be true since we established that a_0 attains the minimum distance from any point in A to x_0. Thus, we have reached a contradiction, so it must be true that $p \cdot a \geq c$ for all $a \in A$.

Thus, we have established that $p \cdot a \geq c = p \cdot a_0 > p \cdot x_0$ for all $a \in A$. This completes the proof. □

Now, with this lemma, we can prove the Supporting Hyperplane Theorem, which we can use to easily prove one version of the Hyperplane Separation Theorem. First, I define what a supporting hyperplane is:

Definition 3.4. We say that a nonempty set A in $\mathbf{R}^n$ is *supported* by a hyperplane at the point $x_0 \in \mathbf{R}^n$ if there exists a $p \in \mathbf{R}^n$ nonzero such that $p \cdot a \geq p \cdot x_0$ for all $a \in A$.

This definition establishes that the set A is contained entirely in one of the closed half-spaces $\overline{\mathcal{O}(p,c)}$ or $\overline{\mathcal{P}(p,c)}$ created by the supporting hyperplane.

Theorem 3.5 (Supporting Hyperplane Theorem)**.** *Let A be a nonempty and convex subset of $\boldsymbol{R}^n$. If $x_0 \notin \mathring{A}$, then A is supported by a hyperplane at x_0.*

Proof. We distinguish two cases: $x_0 \notin \overline{A}$ and $x_0 \in \overline{A}$.

If $x_0 \notin \overline{A}$, we can apply Lemma 3.3 on $\overline{A}$ and the point x_0, as $\overline{A}$ is closed and by Lemma 2.6, $\overline{A}$ is also convex. Then there exists a $p \in \mathbf{R}^n$ nonzero such that $p \cdot a \geq c > p \cdot x_0$ for all $a \in \overline{A}$. Thus, it is in fact true that $p \cdot a > p \cdot x_0$ for all $a \in A$.

If $x_0 \in \overline{A}$, then it is necessarily true that $x_0 \in \partial A$, as $x_0 \notin \mathring{A}$. So then for all $\varepsilon > 0$, there exists some $z_\varepsilon \in B_\varepsilon(x_0) \cap \overline{A}^c$. Thus, there is a sequence (z_n) in $\overline{A}^c$ that converges to x_0. For each z_i, consider the set $\overline{A} - \{z_i\}$. Now, notice that we have the difference between a closed set $\overline{A}$ and a compact set $\{z_i\}$ (since a set containing only one point is closed and bounded), so by Lemma 2.4, $\overline{A} - \{z_i\}$ is closed. Also, we have a difference between two convex sets (a set containing only one point is trivially convex), so by Lemma 1.5, $\overline{A} - \{z_i\}$ is convex as well. Also, since $z_i \notin \overline{A}$, we have $0 \notin \overline{A} - \{z_i\}$.

Then, using Lemma 3.3 on each $\overline{A} - \{z_i\}$ and the point 0, we have that there exists $p_i \in \overline{A} - \{z_i\}$ such that $p_i \cdot (a - z_i) > p_i \cdot 0 = 0$ for all $a \in \overline{A}$. Thus, $p_i \cdot a > p_i \cdot z_i$ for all $a \in \overline{A}$. Now, let $\hat{p_i} = \frac{p_i}{\|p_i\|}$. Note that each $\hat{p_i} \in \{x \in \mathbf{R}^n \mid \|x\| = 1\}$, a compact set. Also, note that $\hat{p_i} \cdot a > \hat{p_i} \cdot z_i$ is certainly true as well. Because each $\hat{p_i}$ belongs to the same compact set, $(\hat{p_n})$ has a convergent subsequence whose limit lies in $\{x \in \mathbf{R}^n \mid \|x\| = 1\}$. Denote this limit p. Now, since (z_n) converges to x_0, by the continuity of the dot product when one argument is fixed, we have $p \cdot a \geq p \cdot x_0$ for all $a \in \overline{A}$. Thus, this inequality certainly holds for all $a \in A$ as well. □

Now, our first Hyperplane Separation Theorem can be proven easily:

Theorem 3.6 (Weak Hyperplane Separation Theorem)**.** *Let A and B be nonempty and convex subsets of $\boldsymbol{R}^n$. If A and B are disjoint, then A and B can be separated by a hyperplane.*

Proof. Consider the set $Z = A - B$. By Lemma 1.5, Z is convex since both A and B are convex. Also, note that A and B are disjoint, so $0 \notin Z$. Then, by the Supporting Hyperplane Theorem, there exists a $p \in \mathbf{R}^n$ such that $p \cdot z \geq p \cdot 0 = 0$ for all $z \in Z$. Now, note that all $z \in Z$ are of the form $a - b$ for $a \in A$ and $b \in B$. Then $p \cdot z = p \cdot (a - b) \geq 0$, so we have that $p \cdot a \geq p \cdot b$ for all $a \in A$ and $b \in B$. Thus, $\inf_{a \in A}\{p \cdot a\} \geq \sup_{b \in B}\{p \cdot b\}$, and we know that $\inf_{a \in A}\{p \cdot a\} > -\infty$ because it is bounded below by $p \cdot b_0$, where b_0 can be any point in B. Similarly, we know that $\sup_{b \in B}\{p \cdot b\} < \infty$. So we can choose c any real number between $\sup_{b \in B}\{p \cdot b\}$ and $\inf_{a \in A}\{p \cdot a\}$ to obtain $p \cdot a \geq c \geq p \cdot b$ for all $a \in A$ and all $b \in B$. □

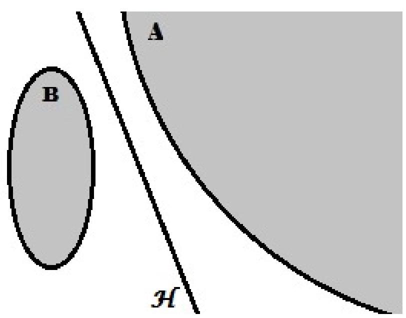
The separation in the theorem above can be strengthened into strict separation if we make more assumptions about A and B. This is the Strong Hyperplane Separation Theorem:

Theorem 3.7 (Strong Hyperplane Separation Theorem)**.** *Let A and B be nonempty and convex subsets of $\mathbf{R}^n$. Let A be closed and B be compact. If A and B are disjoint, then A and B can be strictly separated by a hyperplane.*

Proof. Again, consider the set $Z = A - B$. Z is convex by Lemma 1.5 and closed by Lemma 2.4. Also, $0 \notin Z$ because A and B are disjoint. Thus, we can apply Lemma 3.3 on the set Z and the point 0. Thus, there exists a $p \in \mathbf{R}^{\mathrm{n}}$ nonzero such that $p \cdot z \geq k > p \cdot 0 = 0$ for all $z \in Z$. Thus, in order to obtain a strict inequality, we can say that $p \cdot z > \frac{k}{2} > 0$ for all $z \in Z$. Now, note that all $z \in Z$ can be written in the form $a - b$ for $a \in A$ and $b \in B$. Thus, $p \cdot (a - b) > \frac{k}{2} > 0$. This means that $p \cdot a > \frac{k}{2} + p \cdot b > p \cdot b$ for all $a \in A$ and $b \in B$. Since this chain of inequalities is true for all $b \in B$, it remains true if we replace the middle term with its supremum. Now, because B is compact and the dot product is certainly continuous when one argument is fixed, we have by the Extreme Value Theorem that

$$\begin{aligned} \sup_{b \in B}\{\frac{k}{2} + p \cdot b\} &= \max_{b \in B}\{\frac{k}{2} + p \cdot b\} \\ &= \frac{k}{2} + p \cdot b_0 \end{aligned}$$

for some $b_0 \in B$. Now, we have $p \cdot a > \frac{k}{2} + p \cdot b_0 > p \cdot b$ for all $a \in A$ and $b \in B$. We can set $c = \frac{k}{2} + p \cdot b_0$ to get our desired inequality: $p \cdot a > c > p \cdot b$ for all $a \in A$ and $b \in B$. □



Hyperplane Separation in $\mathbf{R}^2$

There are many other interesting separation results in $\mathbf{R}^{\mathrm{n}}$ that can be proven, such as results involving cones or other types of separation. Additionally, the Hyperplane Separation Theorems have numerous applications, such as in optimization problems (See [2] for some constrained optimization results and see [6] for many other applications of hyperplanes and convex sets). However, this paper will now focus on the generalization of the Hyperplane Separation Theorem into normed vector spaces, commonly known as the Hahn-Banach Separation Theorem.

4. The Hahn-Banach Separation Theorem

The Hahn-Banach Theorem is one of the central tools and results of functional analysis. One of its formulations, usually referred to as the Hahn-Banach Separation Theorem, is a generalization of the Hyperplane Separation Theorem into normed vector spaces.

We will first establish the general Hahn-Banach Theorem for any real vector space in order to help us prove the separation theorem, but we will need several definitions first (recall that for our purposes, we assumed that the scalar field of a vector space is always the real numbers):

Definition 4.1. By a *functional*, we mean a scalar-valued function defined on a vector space. In our case, a *functional* is simply a function $p : V \to \mathbf{R}$, where V is a real vector space.

Definition 4.2. A functional p on a real vector space V is said to be *linear* if

(1) $p(u+v) = p(u) + p(v)$ for all $u, v \in V$
(2) $p(\alpha v) = \alpha p(v)$ for all $\alpha \in \mathbf{R}$, $v \in V$

Note that the dot product (with one argument fixed) in $\mathbf{R}^{\mathrm{n}}$, which we used many times in the preceding section, is an example of a linear functional.

Definition 4.3.

- A functional p on a real vector space V is called *subadditive* if $p(u+v) \leq p(u) + p(v)$ for all $u, v \in V$.
- A functional p on a real vector space V is called *positive homogeneous* if $p(\alpha v) = \alpha p(v)$ for all $\alpha > 0$ and $v \in V$.

Note that the norm is clearly a subadditive, positive homogeneous functional.

Definition 4.4. A *subspace* of a vector space V is a subset of V which is itself a vector space.

Definition 4.5. Let S be a subset of a real vector space V. The *span* of S, denoted *span*(S), is the smallest subspace of V that contains S. Equivalently, the *span* of S is the set of all finite linear combinations of elements in S, i.e. *span*$(S) = \{\lambda_1 v_1 + \lambda_2 v_2 + ... + \lambda_k v_k \mid v_i \in S,\ \lambda_i \in \mathbf{R}\}$, where $k \geq 1$.

We will also need to define some terms and symbols related to set theory. This will be nonexhaustive, as the purpose of this paper is not to delve into set theory.

Definition 4.6.

- Given a set X, a *relation* on X is a subset R of $X \times X$.
- A relation R is called *reflexive* if $(x,x) \in R$ for all $x \in X$.
- A relation R is called *transitive* if $(x,y) \in R$ and $(y,z) \in R$ implies that $(x,z) \in R$.
- A relation R is called *antisymmetric* if $(x,y) \in R$ and $(y,x) \in R$ implies that $x = y$.
- A reflexive, transitive, and antisymmetric relation R on a set X is called a *partial ordering*. In particular, we use the notation $x \preccurlyeq y$ if $(x,y) \in R$ is a partial order, and $x \prec y$ when $x \neq y$.
- Two elements x and y of a set X along with a partial ordering $\preccurlyeq$ are *comparable* if either $x \preccurlyeq y$ or $y \preccurlyeq x$
- A subset C of a set X with partial ordering $\preccurlyeq$ is called a *chain* if any two elements of S are comparable. In other words, for all x, $y \in C$, either $x \preccurlyeq y$ or $y \preccurlyeq x$.
- An element $x \in X$ is called an *upper bound* for a subset Y of X if $y \preccurlyeq x$ for all $y \in Y$.
- An element $x \in X$ is called a *maximal* element for X if $x \preccurlyeq y$ holds only when $y = x$.

The relation $\leq$ on the real numbers is a familiar example of a partial ordering.

Now, we have the definitions to state the following lemma, one of the most important propositions in set theory, which I will not offer a proof for:

Lemma 4.7 (Zorn's Lemma)**.** *If P is a nonempty partially ordered set such that every chain has an upper bound, then P contains a maximal element.*

The most important use of this lemma (for us) is that it allows us to prove that there is a 'maximal extension' in the proof of the Hahn-Banach Theorem. There are actually many interesting results about Zorn's Lemma that unfortunately do not coincide with the goals of this paper (such as its equivalence with the Axiom of Choice, etc.).

Now, we are finally in the position to state and prove the general form of the Hahn-Banach Theorem.

Theorem 4.8 (Hahn-Banach Theorem)**.** *Let V be a real vector space, and let U be a subspace of V. Let p be a subadditive, positive homogeneous functional on V. If f is a linear functional on U such that $f(v) \leq p(v)$ for all $v \in U$, then there is a linear functional F on V such that $F(v) = f(v)$ for all $v \in U$ and $F(v) \leq p(v)$ for all $v \in V$.*

In other words, this theorem allows certain linear functionals defined on some subspace to be extended to the entire vector space, making it a particularly useful tool. Additionally, several geometric/separation results follow (and are actually equivalent) to this theorem, and they are of particular interest to us.

Proof. Let P be the collection of all ordered pairs (Z, g), where Z is a subspace of V that contains U and g is a linear functional on Z that is equal to f on U and is less than or equal to p on Z. First, note that P is nonempty because it contains (U, f). Now, we establish a partial ordering on P by $(Z_1, g_1) \prec (Z_2, g_2)$ if $Z_1 \subset Z_2$ and if $g_2 = g_1$ for g_2 in Z_1. If $\{(Z_\alpha, g_\alpha) \mid \alpha \in A\}$ is a chain (where A is an arbitrary indexing set), then it would have an upper bound (Z, g), where $Z = \bigcup_{\alpha \in A} Z_\alpha$ and g is a linear functional on Z defined by $g(v) = g_\alpha(v)$ for all $v \in Z_\alpha$.

Let us first verify that Z and g are well-defined and fulfill the necessary conditions. Z is clearly a subspace of V that contains U. To see that g is well-defined, suppose that $v \in Z_\alpha$ and $v \in Z_\beta$. Because Z_α and Z_β are in a chain, we have that either $(Z_\alpha, g_\alpha) \prec (Z_\beta, g_\beta)$ or $(Z_\beta, g_\beta) \prec (Z_\alpha, g_\alpha)$. Without loss of generality, assume that the former is true. Then $Z_\alpha \subset Z_\beta$ and $g_\beta = g_\alpha$ for g_β in Z_α, meaning that $g_\beta(v) = g_\alpha(v)$ for $v \in Z_\alpha$. This implies that g is well-defined. Additionally, g is linear because for $u,\ v \in Z$, there exists some α such that $u,\ v \in Z_\alpha$, so then for $\zeta,\ \eta \in \mathbf{R}$, we have $g(\zeta u + \eta v) = g_\alpha(\zeta u + \eta v) = \zeta g_\alpha(u) + \eta g_\alpha(v) = \zeta g(u) + \eta g(v)$. So finally, we can conclude that $(Z, g) \in P$ is an upper bound of the chain $\{(Z_\alpha, g_\alpha) \mid \alpha \in A\}$ i.e. $(Z_\alpha, g_\alpha) \prec (Z, g)$ for all α.

Now, we can see that P fulfills the assumptions of Zorn's Lemma, so P has a maximal element, which I denote (Y, F). Now, we must show that Y is in fact equal to V, in which case F is our desired extension of f into V. Assume by contradiction that $Y \neq V$, so then there exists some $v_0 \in V \setminus Y$. Let $Y' = Y + span\{v_0\} = \{v + \lambda v_0 \mid v \in Y,\ \lambda \in \mathbf{R}\}$. So now, we will find some $(Y', F') \in P$ such that $(Y, F) \prec (Y', F')$, which will contradict the maximality of (Y, F).

We already have that $Y \subset Y'$ and that Y' is a subspace of V that contains U. Now, we will need to show that $F' = F$ for F' in Y, and also that the F' fulfills the conditions for (Y', F') be in the collection P (i.e. we need to show that F' is a linear functional that coincides with f on U and satisfies $F' \leq p$ on Y'). For some $\alpha \in \mathbf{R}$ fixed, we can define $F'(v + \lambda v_0) = F(v) + \lambda\alpha$ for all $v \in Y$, $\lambda \in \mathbf{R}$, with $F'(v_0) = \alpha$. Then it follows immediately that F' is linear and that $F' = F$ for F' in Y (which in turn implies that F' coincides with f on U, as (Y, F) is in P and is maximal, meaning that $(U, f) \preccurlyeq (Y, F)$). So finally, we just need to show that we can choose an α such that $F' \leq p$ on Y'.

In other words, we need an α such that when $\lambda > 0$, we have

$$F'(v + \lambda v_0) = F(v) + \lambda\alpha \leq p(v + \lambda v_0) \tag{4.9}$$

and when $\lambda < 0$, we have (by letting $\lambda = -\mu$)

$$F'(v - \mu v_0) = F(v) - \mu\alpha \leq p(v - \mu v_0) \tag{4.10}$$

for all $v \in Y$. Now, we see that (4.9) is equivalent to

$$\alpha \leq p(w + v_0) - F(w) \tag{4.11}$$

for all $w \in Y$ if we divide by λ and let $w = \frac{v}{\lambda}$, and we see that (4.10) is equivalent to

$$\alpha \geq F(x) - p(x - v_0) \tag{4.12}$$

for all $x \in Y$ if we divide by μ and let $x = \frac{v}{\mu}$. We can combine (4.11) and (4.12) to obtain

$$p(w + v_0) - F(w) \geq \alpha \geq F(x) - p(x - v_0)$$

for all w, $x \in Y$.

Thus, to see that α exists and is well-defined, we need to show that

$$\inf_{w \in Y}\{p(w + v_0) - F(w)\} \geq \sup_{x \in Y}\{F(x) - p(x - v_0)\}$$

This means that we need to show that

$$p(w + v_0) - F(w) \geq F(x) - p(x - v_0)$$

holds for all w, $x \in Y$. This is equivalent to stating that

$$F(w) + F(x) = F(w + x) \leq p(w + v_0) + p(x - v_0) \tag{4.13}$$

Now, we can show that (4.13) is true for all w, $x \in Y$, which will establish that we can find a suitable α. It is not a difficult task to do so:

$$\begin{aligned} F(w + x) &\leq p(w + x) \\ &= p(w + v_0 - v_0 + x) \\ &\leq p(w + v_0) + p(x - v_0) \end{aligned}$$

So, we are finally able to verify the existence of a proper α such that $F' \leq p$. Thus, we can conclude that (Y', F') is in the collection P and that $(Y, F) \prec (Y', F')$. This contradicts the maximality of (Y, F) established by Zorn's Lemma, so we can conclude that $Y = V$, with F being the desired extension of f into V. □

I will demonstrate the strength of the general Hahn-Banach Theorem by establishing some useful corollaries that follow. We will now work with normed vector spaces, which play a very important role in functional analysis; in particular, Banach Spaces and Hilbert Spaces are both examples of normed vector spaces. Because we are working with normed vector spaces, we may consider bounded linear functionals, which I define now:

Definition 4.14. Let V be a normed vector space. A linear functional F on V is *bounded* if there exists some $M > 0$ such that $|F(v)| \leq M\|v\|$ for all $v \in V$.

We will also need to define the norm of a bounded linear functional:

Definition 4.15. Let V be a normed vector space and F be a bounded linear functional on V. The *norm* of F is defined by $\|F\| = \sup_{\|v\| \leq 1} \{|F(v)|\}$.

First, I will prove a useful lemma that establishes the equivalence of bounded and continuous linear functionals. Thus, we can replace all the occurrences of 'bounded linear functional' with 'continuous linear functional,' and vice versa.

Lemma 4.16. *Let V be a normed vector space. A linear functional F on V is bounded if and only if it is continuous.*

Proof. Assume that F is bounded. Then we have that for all v, $h \in V$, h nonzero, $|F(v+h) - F(v)| = |F(v) + F(h) - F(v)| = |F(h)| \leq M\|h\|$ for some $M > 0$. Continuity at v follows from letting h go to zero.

Now, assume that F is continuous. Then F is obviously continuous at 0. Thus, there exists a $\delta > 0$ such that $|F(h)| = |F(h) - F(0)| \leq 1$ for all $h \in V$ with $\|h\| \leq \delta$ (we note that for any linear functional f, we have that $f(0) = 0$ due to the fact that $f(0) = f(0+0) = f(0) + f(0)$). Thus, for all $v \in V$ nonzero, we have $|F(v)| = |\frac{\|v\|}{\delta} F(\delta \frac{v}{\|v\|})| = \frac{\|v\|}{\delta}|F(\delta \frac{v}{\|v\|})| \leq \frac{\|v\|}{\delta} \cdot 1 = \frac{\|v\|}{\delta}$, so F is bounded. □

The next theorem is an easy application of the general Hahn-Banach Theorem, and it is generally called the Hahn-Banach Theorem for Normed Vector Spaces:

Theorem 4.17 (Hahn-Banach Theorem for Normed Vector Spaces)**.** *Let V be a normed vector space, and let U be a subspace of V. If f is a bounded linear functional defined on U with norm $\|f\|_U$, then there is a bounded linear extension F of f to V such that $\|F\| = \|f\|_U$.*

Proof. Let $p(v) = \|f\|_U\|v\|$ in the Hahn-Banach Theorem. It is easily verified that p is subadditive and positive homogeneous. We also see that $f \leq p$ on U. So by the Hahn-Banach Theorem, we have that there is a linear extension F of f to V such that $F \leq p$. Now, by using the linearity of F and the properties of norms, we get the following chain of inequalities: $-p(v) = -p(-v) \leq -F(-v) = F(v) \leq p(v)$ for all $v \in V$. Thus, we get that $-\|f\|_U\|v\| \leq F(v) \leq \|f\|_U\|v\|$, or in other words, $|F(v)| \leq \|f\|_U\|v\|$ for all $v \in V$ (i.e. F is bounded). Since this is true for all $v \in V$, it follows that $\|F\| \leq \|f\|_U$. Also, because F is an extension of f, it cannot have a smaller norm, so we also have $\|F\| \geq \|f\|_U$. This implies that $\|F\| = \|f\|_U$ □

Corollary 4.18. *Let V be a normed vector space, and let $v_0 \in V$ be nonzero. Then there is an bounded linear functional F on V such that $\|F\| = 1$ and $\|v_0\| = F(v_0)$.*

Proof. Let $U = span\{v_0\}$. Define f on U by $f(\alpha v_0) = \alpha\|v_0\|$, so f is clearly a bounded linear functional with $\|f\|_U = 1$. By the Hahn-Banach Theorem for Normed Vector Spaces, we see that there is a norm-preserving extension F of f to V that is also a bounded linear functional (so $\|F\| = 1$). Now, because $v_0 \in U$ and $F = f$ on U, we have that $F(v_0) = f(v_0) = \|v_0\|$. □

Now, to begin the separation results for normed vector spaces, we will need several other definitions, including a generalization of the definition of a hyperplane.

Definition 4.19. Let C be a set in a real normed vector space V. The *Minkowski functional* of C is a functional $p : V \to \mathbf{R}$, defined by $p(v) = \inf_{v \in \lambda C} \{\lambda > 0\}$.

It is clear that a Minkowski functional is always nonnegative, but it is possible that it can be equal to ∞. Note that the Minkowski functional of the unit ball $B_1(0)$ is the norm $\|\cdot\|$.

Definition 4.20. Let V be a real normed vector space, F be a continuous linear functional, and $c \in \mathbf{R}$ be constant. A *hyperplane* in V is a set of the form $\mathcal{H}(F, c) = \{x \in V \mid F(x) = c\}$.

We see now that hyperplanes in $\mathbf{R}^{\mathrm{n}}$ are just a special case of this definition, where the continuous linear functional is the dot product (with one argument fixed).

Definition 4.21. We say that two nonempty sets A and B in a real normed vector space V are *separated* by a hyperplane if there exists a continuous linear functional F on V and a constant $c \in \mathbf{R}$ such that $F(a) \geq c \geq F(b)$ for all $a \in A$ and $b \in B$.

If $\overline{\mathcal{O}(F, c)} = \{x \in V \mid F(x) \geq c\}$ and $\overline{\mathcal{P}(F, c)} = \{x \in V \mid F(x) \leq c\}$ are the two closed half-spaces of a real normed vector space V, then this definition of separation establishes that A and B lie in different $\overline{\mathcal{O}(F, c)}$ and $\overline{\mathcal{P}(F, c)}$.

We will now establish that we can ensure that the Minkowski functional be subadditive and positive homogeneous by putting some requirements on the set it is defined on:

Lemma 4.22. *Let C be a nonempty convex set in a real normed vector space V such that $0 \in \mathring{C}$. The Minkowski functional of C, $p(v) = \inf_{v \in \lambda C} \{\lambda > 0\}$, is subadditive and positive homogeneous.*

Proof. Positive homogeneity follows from

$$\begin{aligned} p(\alpha v) &= \inf_{\alpha v \in \lambda C} \{\lambda > 0\} \\ &= \alpha \inf_{v \in \frac{\lambda}{\alpha} C} \{\frac{\lambda}{\alpha} > 0\} \\ &= \alpha p(v) \end{aligned}$$

To obtain subadditivity, we can do the following: if $\varepsilon > 0$ is given, let λ and μ be positive numbers such that $u = \lambda c_1$ for some $c_1 \in C$, $v = \mu c_2$ for some $c_2 \in C$, $\lambda < p(u) + \frac{\varepsilon}{2}$, and $\mu < p(v) + \frac{\varepsilon}{2}$. So then

$$\begin{aligned} u + v &= \lambda c_1 + \mu c_2 \\ &= (\lambda + \mu)(\underbrace{\frac{\lambda}{\lambda + \mu}}_{\in [0,1]} c_1 + \underbrace{\frac{\mu}{\lambda + \mu}}_{\in [0,1]} c_2) \end{aligned}$$

which implies that $u + v \in (\lambda + \mu)C$ by the convexity of C. Thus, by the definition of the Minkowski functional, $p(u + v) \leq \lambda + \mu < p(u) + p(v) + \varepsilon$, implying that $p(u + v) \leq p(u) + p(v)$. □

Thus, the Minkowski functional is a viable functional to be used as the subadditive, positive homogeneous bounding functional required in the Hahn-Banach Theorem. Now, we will finally be working with separation results in a normed vector space. This is our first:

Lemma 4.23. *Let C be a nonempty convex subset of a real normed vector space V such that C contains an interior point, and let $x_0 \notin C$. Then there is a continuous linear functional F on V such that $F(x) \leq F(x_0)$ for all $x \in C$.*

Proof. Without loss of generality, we can assume that $0 \in \mathring{C}$ because if not, we can simply translate C. Let $X = span\{x_0\} = \{\lambda x_0 \mid \lambda \in \mathbf{R}\}$. We define the linear functional f on X by $f(\lambda x_0) = \lambda$. We have that since $x_0 \notin C$, $\lambda x_0 \notin \lambda C$ for $\lambda \neq 0$, so then if p is the Minkowski functional of C, we have $p(\lambda x_0) \geq \lambda = f(\lambda x_0)$, implying that $f(x) \leq p(x)$ for all $x \in X$. Then, by the Hahn-Banach Theorem, there is a linear extension F of f into V that satisfies $F(x) \leq p(x)$ for all $x \in V$. Now, for $x \in C$, we have that $F(x) \leq p(x) \leq 1 = F(x_0)$ (since by definition, the Minkowski functional $p(x) \leq 1$ when $x \in C$). So we have proven that $F(x) \leq F(x_0)$ for all $x \in C$. It remains to show that F is continuous.

Since we assumed without loss of generality that $0 \in \mathring{C}$, we have that for a given $\varepsilon > 0$ and for all y with $\|y\|$ sufficiently small, $\frac{y}{\varepsilon}, -\frac{y}{\varepsilon} \in C$. So because $F(\frac{y}{\varepsilon}) \leq 1$ and $F(-\frac{y}{\varepsilon}) \leq 1$, we have that $-\varepsilon \leq F(y) \leq \varepsilon$, or in other words, $|F(y)| \leq \varepsilon$ for $\|y\|$ sufficiently small i.e. $\|y\| \leq \delta$. By the linearity of F and setting $y = x - z$, we have that $|F(x) - F(z)| = |F(x - z)| \leq \varepsilon$ for $\|x - z\| \leq \delta$. In other words, F is continuous. □

Note that the above lemma is a generalization of the Supporting Hyperplane Theorem into normed vector spaces.

The following theorem is the goal of this paper. It is often called the Hahn-Banach Separation Theorem for Normed Vector Spaces, and it is the generalization of the Weak Hyperplane Separation Theorem (that we proved in the preceding section) into normed vector spaces. Notably, the proof will work in almost exactly the same manner.

Theorem 4.24 (Hahn-Banach Separation Theorem)**.** *Let A and B be nonempty and convex subsets of a real normed vector space V. Furthermore, assume that A and B are disjoint and that A has an interior point. Then there is a hyperplane that separates A and B.*

Proof. We consider the set $Z = B - A$, which is convex by Lemma 1.5. Since A and B are disjoint, $0 \notin Z$. Also, note that it is necessarily true that Z has an interior point. Then, the set Z with the point 0 fulfill the conditions of Lemma 4.23, so there exists a continuous linear functional F such that $F(z) \leq F(0) = 0$ for all $z \in Z$ (recall that for any linear functional f, we have that $f(0) = 0$). Now, note that all $z \in Z$ can be written as $b - a$ for $a \in A$ and $b \in B$. So then, we have that $F(z) = F(b - a) = F(b) - F(a) \leq 0$, or $F(a) \geq F(b)$ for all $a \in A$ and $b \in B$. Thus, $\inf_{a\in A} F(a) \geq \sup_{b\in B} F(b)$, and we know $\inf_{a\in A} F(a) > -\infty$ because it is bounded below by $F(b_0)$, where b_0 can be any point in B (and $F(b_0) \neq -\infty$ by the continuity of F). Similarly, we know that $\sup_{b\in B} F(b) < \infty$. So we can choose c any real number between $\sup_{b\in B} F(b)$ and $\inf_{a\in A} F(a)$ to obtain $F(a) \geq c \geq F(b)$ for all $a \in A$ and $b \in B$. □

We see especially now that the Weak Hyperplane Separation Theorem established in the preceding section is essentially just the Hahn-Banach Separation Theorem, with the dot product (with one argument fixed) as the continuous linear functional. In fact, we even used the same process as proving the Weak Hyperplane Separation Theorem to prove this theorem in that we first proved a generalization of the Supporting Hyperplane Theorem and then used that to prove the Hahn-Banach Separation Theorem.

There are numerous other separation results that can be established in a normed vector space with the Hahn-Banach Theorem, including a generalization of the Strong Hyperplane Separation Theorem. In fact, the Hahn-Banach Separation Theorem can be even further generalized into topological vector spaces, although this theorem takes much more work to establish and prove. The interested reader is urged to look into [1] to see the Hahn-Banach Separation Theorem for Topological Vector Spaces (although some basic knowledge of topology may be helpful before doing so).

Acknowledgments. First and foremost, I would like to thank my mentor, Clark Butler, for his extensive assistance on helping me develop the ideas behind my paper and for dealing with my numerous questions and shortcomings. This paper would not have been possible without you. I would also like to extend my gratitude to Peter May for running this fantastic REU Program and for giving me the opportunity to spend a summer studying mathematics. This program has honestly been one of my most illuminating experiences, and it certainly would not have been possible without Professor May and the Mathematics Department at the University of Chicago.

Lastly, I would like to express my deep gratitude to Adina Ciomaga, my analysis professor for giving me the motivation for this paper. I remember when she told me to prove that in $\mathbf{R}^2$, there always exists a line that separates a closed, convex set and a point not contained in the set as a challenge problem. The idea behind that proof is what led me to become so interested in the theory of separation and to pursue the generalization of this problem in my paper. Thank you so much for giving me this problem, and I dedicate this paper to you.

References

[1] Lawrence Baggett. Topological Vector Spaces and Continuous Linear Functionals. http://spot.colorado.edu/~baggett/funcchap3.pdf
[2] Kim C. Border. Separating Hyperplane Theorems. http://www.hss.caltech.edu/~kcb/Notes/SeparatingHyperplane.pdf
[3] Marián Fabian, Peter Habala, Petr Hájek, Vincente Montesinos Santalucía, Jan Pelant, and Václav Zizler. Functional Analysis and Infinite-Dimensional Geometry. Springer-Verlag New York, Inc. 2001
[4] Charles W. Groetsch. Elements of Applicable Functional Analysis. Marcel Dekker, Inc. 1980.
[5] John Nachbar. Convex Sets, Separation, and Support. http://artsci.wustl.edu/~e503jn/files/Math%20Notes/Convexity11.pdf
[6] Angelia Nedich. Convex Problems, Separation Theorems. http://www.ifp.illinois.edu/~angelia/L7_separationthms.pdf
[7] Felipe A. Ramírez. Functional Analysis Notes. http://www-users.york.ac.uk/~fr606/fa_notes/5_hahn_banach.pdf
[8] Ross M. Starr. Convex Sets, Separation Theorems, and Non-Convex Sets in R^n. http://econweb.ucsd.edu/~rstarr/113Winter2012/2010CHAP-8.pdf