The Hahn-Banach Theorem

Rein, Jasper, Lukas

March 2020

In this standalone article our aim is to give a proof of the Hahn-Banach Theorem for real linear spaces. The proof is aimed at students following an introductory course in functional analysis.

1 An introduction to POSets

The proof of the Hahn-Banach Theorem makes use of Zorn's Lemma, a statement equivalent to the Axiom of Choice. Before proving this theorem, we thus first introduce some concepts used by Zorn's Lemma. The structure of the following section is taken from [2].

Definition 1.1. A *poset*, short for partially ordered set, is a set P equiped with a relation " $\leq$ " between elements of P, which satisfies the following conditions:
(i) $a \leq a$ for all $a \in P$ (reflexivity)
(ii) $(a \leq b) \wedge (b \leq c) \implies (a \leq c)$ for all $a, b, c \in P$ (transitivity)
(iii) $(a \leq b) \wedge (b \leq a) \implies (a = b)$ for all $a, b \in P$ (antisymmetry)

Note that a poset need not to be a *linear order*, which is an order in which every two elements are comparable.

Examples of posets are given by the powerset $\mathcal{P}(X)$ of a set X, partially ordered by inclusion, and by the reals $\mathbb{R}$ with the usual ordering. A counterexample is given by the complex numbers ordered by the usual ordering on the modulus, because condition (iii) is not satisfied.

We also introduce the following definitions:

Definition 1.2.

- A subset $C \subset P$ is called a *chain* if C with the restricted order of P is a linear order.
- An *upper bound* $b \in P$ of a subset $Q \subset P$ is an element of P such that $q \leq b$ for all $q \in Q$.
- A *maximal element* $m \in P$ is an element of P such that $(m \leq p) \implies (m = p)$ for all $p \in P$.

Note that this maximal element does not have to be unique.

We are now able to state Zorn's Lemma. Note that we state it as a definition, since we treat it as an axiom.

Definition 1.3 (Zorn's Lemma)**.** If every chain $C \subset P$ of a poset $(P, \leq)$ has an upper bound in P, then P has a maximal element.

We illustrate this axiom by using it to prove two examples.

Example 1. Let R be a commutative ring with $1 \neq 0$. We will show that R contains a maximal ideal; a proper ideal, so not equal to R, that is not contained in any other ideal. Let P be the poset of all proper ideals of R, ordered by inclusion. Note that $P \neq \emptyset$, because $\{0\} \in P$. We now let C be any a chain in P. If $C = \emptyset$, $\{0\}$ is an upper bound of C. If $C \neq \emptyset$, we look at $\bigcup C$. We claim that $\bigcup C$ is again a proper ideal.

For all $a, b \in \bigcup C$ there are $C_1, C_2 \in C$ such that $a \in C_1, b \in C_2$. Because C is a chain, we assume without loss of generality that $C_2 \subseteq C_1$, and thus that $a, b \in C_1$. Because C_1 is an ideal, this means that $a + b \in C_1 \subseteq \bigcup C$. Furthermore, for all $c \in \bigcup C$ there is a $C_3 \in C$ such that $c \in C_3$. Because C_3 is an ideal, this means that $rc, cr \in C_3 \subseteq \bigcup C$ for all $r \in R$. We conclude that $\bigcup C$ is an ideal of R. Finally, we see that $1 \notin C'$ for all $C' \in C$, because otherwise $r = r \cdot 1 \in C'$ for all $r \in R$, which would mean that $C' = R$. This implies that $1 \notin \bigcup C$, and thus that $\bigcup C$ is a proper ideal of R. Because $\bigcup C \in P$ and $C' \subseteq \bigcup C$ for all $C' \in C$, we conclude that every chain in P has an upper bound in P. By Zorn's Lemma, P has a maximal element M, which is a proper ideal not contained in any other proper ideal of P, and thus is a maximal ideal of P.

Example 2. Zorn's Lemma can also be used to show, and is even equivalent to the fact that every vector space V has a basis. To prove this, we look at the poset P of all linearly independent subsets of V, again ordered by inclusion. Note that $P \neq \emptyset$, because $\{0\} \in P$. We now let C be any chain in P. As in the previous example, $\{0\}$ is an upper bound for the empty set, so this case is settled. For $C \neq \emptyset$, we again look at $\bigcup C$. We claim that $\bigcup C$ also is a linearly independent subset of V. Indeed, if $v_1 \in \bigcup C$ could be written as a finite linear combination of vectors in $\bigcup C$, C being a chain would imply that there exists a $C' \in C$ such that v_1, together with all these vectors, also lie in C'. However, this would contradict our assumption that C' is linearly independent. We conclude that $\bigcup C$ is an upper bound for C in P.

Having accomplished that every chain in P has an upper bound in P, we conclude, using Zorn's Lemma, that P has a maximal element M, and we assert that this M forms a basis of V. On the contrary, assume that there exists a $v_0 \in V$ that is linearly independent from M. Then $M \cup \{v_0\} \in P$, while $M \subsetneq M \cup \{v_0\}$, which contradicts the maximality of M.

2 The Hahn-Banach Theorem

We first prove a general version of the Hahn-Banach theorem on real linear spaces. The following proof is from a mix of [1] and [4].

Theorem 1. *Let X be a linear space over $\mathbb{R}$ with a subadditive, positive homogeneous functional p. This means p satisfies*

- $p(x + y) \leq p(x) + p(y)$ *for all* $x, y \in X$ *(subadditivity)*
- $p(\alpha x) = \alpha p(x)$ *for all* $x \in X$, $\alpha \in \mathbb{R}_{>0}$ *(positive homogeneity).*

Let M be a subspace of X and $f : M \longrightarrow \mathbb{R}$ be a linear functional defined on M, bounded by p, thus $f(z) \leq p(z)$ for all $z \in M$.

Then there exists an extension $F : X \longrightarrow \mathbb{R}$ of f such that $F(x) \leq p(x)$ for all $x \in X$.

Proof. The proof is an application of Zorn's Lemma on the poset of extensions.

Define Σ to be the set of tuples (W, g) where W is a subspace of X containing M and $g : W \longrightarrow \mathbb{R}$ is a linear functional (also denoted as $g \in W^*$), such that $g(x) \leq p(x)$ for all $x \in W$ and $g \circ i_M = f$, where $i_M : M \longrightarrow W$ is the inclusion map, or in other words, g is an extension of f. In a more compact form:

$$\Sigma := \{(W, g) \mid W \text{ subspace of } V, g \in W^*, \forall_{x \in W} g(x) \leq p(x), g \circ i_M = f\}. \tag{1}$$

Because Σ is a subset of $\bigcup_{M\subset W\subset X}\{W\}\times W^*$, we can say that Σ is a set.

Next we define a partial order on Σ, where $(U_1,g_1)\leq(U_2,g_2)$ precisely when $U_1\subset U_2$ and $g_2\circ i_{U_1}=g_1$ where $i_{U_1}:U_1\longrightarrow U_2$ is the inclusion map.

We check that Σ satisfies the prerequisites of Zorn's lemma, namely that Σ is nonempty and every ascending chain has an upper bound. We have $(M,f)\in\Sigma$ thus it is nonempty. Now let $(U_i,g_i)_{i\in I}$ be a chain in Σ, then define $g:\bigcup_{i\in I}U_i\longrightarrow\mathbb{R}$ as $g(x)=g_i(x)$ if $x\in U_i$. This is well-defined on the intersections because if $x\in U_i\cap U_j$, then by the ascending chain condition we can assume without loss of generality that $g_j\circ i_{U_i}=g_i$. This means $g_i(x)=g_j(i_{U_i}(x))=g_j(x)$ so $g:\bigcup_{i\in I}U_i\longrightarrow\mathbb{R}$ is a well-defined linear functional that is bounded by p and is an extension of f. Thus $(\bigcup_{i\in I}U_i,g)\in\Sigma$ is an upper bound of $(U_i,g_i)_i$. Then Zorn's lemma gives us a maximal element (N,g).

To prove the main theorem, it is sufficient to show that $N=X$. We prove this by contradiction, so suppose a $x\in X\setminus N$ exists. Then $N\subsetneq\mathbb{R}x\oplus N\subset X$. We define an extension h of g by defining

$$h(x):=\inf_{z\in N}\{p(x+z)-g(z)\}. \tag{2}$$

And by extending linearly as $h(\alpha x+z):=\alpha h(x)+g(z)$ for $\alpha\in\mathbb{R}$ and $z\in N$, we have defined a linear functional $h:\mathbb{R}x\oplus N\longrightarrow\mathbb{R}$.

Now our goal is to prove that $h(\alpha x+z)\leq p(\alpha x+z)$ for all $\alpha x+z\in\mathbb{R}x\oplus N$, because this would imply that $(\mathbb{R}x\oplus N,h)\in\Sigma$, contradicting maximality of (N,g). We make the following two remarks.

(A) for all $z\in N$ we have $h(x)\leq p(x+z)-g(z)$ thus $h(x+z)=h(x)+g(z)\leq p(x+z)$.

(B) for all $y,z\in N$ we have

$$g(z)-g(y)=g(z-y)\leq p(z-y)=p((x+z)-(x+y))\leq p(x+z)+p(-(x+y)) \tag{3}$$

which means $-p(-(x+y))-g(y)\leq p(x+z)-g(z)$. Now if we fix the y and minimize over z, we conclude that

$$-p(-(x+y))-g(y)\leq\inf_{z}\{p(x+z)-g(z)\}=h(x) \tag{4}$$

which means $-h(x+y)\leq p(-(x+y))$ thus $h(-x-y)\leq p(-x-y)$ for all $y\in N$.

Now let $\alpha x+z\in\mathbb{R}x\oplus N$ be arbitrary. We separate three cases:

- $\alpha>0$: We have $h(\alpha x+z)=\alpha h(x+\alpha^{-1}z)\leq\alpha p(x+\alpha^{-1}z)$ by remark (A) and because $\alpha>0$. Then $\alpha p(x+\alpha^{-1}z)=p(\alpha x+z)$ because $\alpha>0$. Thus $h(\alpha x+z)\leq p(\alpha x+z)$.
- $\alpha=0$: Then $h(\alpha x+z)=h(z)=g(z)\leq p(z)=p(\alpha x+z)$.
- $\alpha<0$: Then $h(\alpha x+z)=(-\alpha)h(-x-\alpha^{-1}z)\leq(-\alpha)p(-x-\alpha^{-1}z)$ by remark (B) and because $-\alpha>0$. Thus $(-\alpha)p(-x+(-\alpha^{-1})z)=p(\alpha x+z)$ because $-\alpha>0$, so $h(\alpha x+z)\leq p(\alpha x+z)$.

Thus in all cases $h(\alpha x+z)\leq p(\alpha x+z)$.

This proves that $(\mathbb{R}x\oplus N,h)\in\Sigma$ which contradicts maximality of (N,g) so $X=N$. This means $g:X\longrightarrow\mathbb{R}$ is an extension of f with $g(z)\leq p(z)$ for all $z\in X$. □

When $X=H$ is a Hilbert space and M is a closed subspace, the statement is proved much easier. In that case, f has a Riesz representation

$$f(x)=\langle x,z\rangle \tag{5}$$

for some $z\in M$. Since the inner product is defined on all of H, this formula already is defined on all of H and does not have to be further extended.

3 Corollaries and Applications

The following section discusses some corollaries and applications of the Hahn-Banach Theorem found in [2].

3.1 Corollaries of the Hahn-Banach Theorem

Corollary 1 (Hahn-Banach Theorem extended to norms)**.** *Let X be a normed linear space over $\mathbb{R}$, M a subspace of X and $f : M \longrightarrow \mathbb{R}$ a bounded linear functional on M. Then there exists a bounded extension $F : X \longrightarrow \mathbb{R}$ of f with $\|F\| = \|f\|_M$.*

Proof. Define a subadditive, positive homogeneous functional on X by $p(x) := \|f\|_M\|x\|$. We have $f(x) \leq \|f(x)\| \leq \|f\|_M\|x\| = p(x)$ for $x \in M$. Then by the Hahn-Banach theorem (Theorem 1) there exists an extension F of f such that $F(x) \leq p(x)$ for all $x \in X$.

Now we note that p is also absolutely homogeneous, so we get $-p(x) = -p(-x) \leq -F(-x) = F(x) \leq p(x)$ for all $x \in X$ which means $-F(x) \leq p(x)$, so $|F(x)| \leq p(x) = \|f\|_M\|x\|$ for $x \in X$. Thus F is bounded with $\|F\| \leq \|f\|_M$. We also have $\|F\| \geq \|Fx\|\|x\|^{-1} = \|fx\|\|x\|^{-1}$ for $x \in M \setminus \{0\}$, and thus $\|F\| \geq \|f\|_M$. We conclude $\|F\| = \|f\|_M$. □

Corollary 2. *Let X be a normed linear space, $0 \neq x_0 \in X$. Then there exists a bounded linear functional F such that $F(x_0) = \|x_0\|$ and $\|F\| = 1$.*

Proof. Define $M := \{ax_0 \mid a \in \mathbb{R}\}$ and consider the functional $f : M \longrightarrow \mathbb{R}, ax_0 \mapsto a\|x_0\|$. We see f is a linear functional on a subspace of X with the desired properties and $\|f\| = 1$. By the previous corollary there exists an extension $F : X \longrightarrow \mathbb{R}$ such that $\|F\| = \|f\| = 1$. □

Corollary 3. *For every $x \in X$ we have $\|x\| = \sup\left\{\frac{|f(x)|}{\|f\|} | f \in X^*, f \neq 0\right\}$.*

Proof. From the previous corollary we know there exists a linear functional such that $\frac{|F(x)|}{\|F\|} = \|x\|$, so $\sup \frac{|f(x)|}{\|x\|} \geq \|x\|$. We also know that $|f(x)| \leq \|f\|\|x\|$, or equivalently $\frac{|f(x)|}{\|f\|} \leq \|x\|$, so we have equality. □

3.2 Application of the Hahn-Banach Theorem: linear extension of distance function

For a subset $A \subset X$ of a normed linear space X and $x \in X$ we can define the distance function $d(x, A) := \inf_{z\in A}\|x - z\|$. The Hahn-Banach theorem gives the following theorem.

Theorem 2. *Let M be a subspace of a normed linear space X. For $x_0 \in X\backslash M$ with $d(x_0, M) > 0$, there exists an $F \in X^*$ such that $\|F\| = 1$, $F|_M = 0$ and $F(x_0) = d(x_0, M)$.*

Proof. We have $x_0 \notin M$. Define the linear map $f : \mathbb{R}x_0 \oplus M \longrightarrow \mathbb{R} : \alpha x_0 + z \mapsto \alpha d(x_0, M)$. This gives us

$$\|\alpha x_0 + z\| = |\alpha|\|x_0 + \alpha^{-1}z\| \geq |\alpha| \inf_{y\in M}\|x_0 - y\| = |\alpha|d(x_0, M) = |f(\alpha x_0 + z)|. \quad (6)$$

Thus $\|f\| \leq 1$.

Furthermore, by definition of the infimum, there exists a sequence $(z_n)_n$ in M such that $\|x_0 - z_n\| \to d(x_0, M)$ as $n \to \infty$. Thus

$$d(x_0, M) = f(x_0 - z_n) \leq \|f\|\|x_0 - z_n\| \to \|f\|d(x_0, M). \quad (7)$$

This means $\|f\| \geq 1$ because $d(x_0, M) > 0$ and we conclude $\|f\| = 1$. Now we apply the Hahn-Banach Theorem to get an extension $F : X \longrightarrow \mathbb{R}$ of f with $\|F\| = \|f\| = 1$. □

4 Geometric form of the Hahn-Banach theorem

The Hahn-Banach Theorem is also known in its geometric form. Before we state this form, we first recall the definition of the Minkowski functional (also see Exercise 1.18 of [3]).

4.1 The Minkowski functional

The Minkowski functional will be used as the subadditive, positively homogeneous functional p where we can apply the Hahn-Banach theorem.

Definition 4.1. Let W be a convex subset of a normed vector space X such that 0 is an interior point of W. For each $x \in X$, we define the *Minkowski functional* of x as

$$p(x) := \inf\{\alpha^{-1} \mid \alpha > 0,\ \alpha x \in W\} = \inf\{\lambda \geq 0 \mid x \in \lambda W\}. \tag{8}$$

We make the following observations:

- $p(x)$ is always finite because 0 is an interior point,
- $p(x) \geq 0$ because $\alpha^{-1} \geq 0$ for all $\alpha > 0$,
- p is not definite if W is not bounded, for example the whole space, because we can then enlarge α or shrink λ as much as we want and get a zero $p(x)$ without x being zero.
- p might not be absolutely homogeneous if W is not symmetric around the origin ie. $x \in W \Leftrightarrow -x \in W$.

Lemma 3. The Minkowski functional satisfies the following properties:

- $p(x + y) \leq p(x) + p(y)\ \forall x, y \in X$ (subadditivity)
- $p(\lambda x) = \lambda p(x)\ \forall x \in X, \lambda \geq 0$ (positive homogeneity)
- $p(x) < 1 \implies x \in W$
- $x \in W \implies p(x) \leq 1$

Proof.

- Let $\alpha, \beta > 0$, and let $\alpha x, \beta y \in W$. Since W is convex,

$$\frac{\alpha^{-1}\alpha x + \beta^{-1}\beta y}{\alpha^{-1} + \beta^{-1}} = \frac{x + y}{\alpha^{-1} + \beta^{-1}} \in W. \tag{9}$$

  This means that $p(x + y) \leq \alpha^{-1} + \beta^{-1} \leq p(x) + p(y)$.

- We note that $p(0) = 0$. For $\alpha > 0$, we see that

$$p(\alpha x) = \inf\{\beta^{-1} \mid \beta > 0,\ \alpha\beta x \in W\} = \alpha \inf\{\gamma^{-1} \mid \gamma > 0,\ \gamma x \in W\} = \alpha p(x). \tag{10}$$

- When $p(x) < 1$, there exists an $\alpha > 1$ such that $\alpha x \in W$. Because $0 \in W$ and W is convex, this means that $x = \alpha x \cdot \alpha^{-1} + (1 - \alpha^{-1})0 \in W$.

  In terms of the λ definition we have a $\lambda < 1$ such that $x \in \lambda W$, if $\lambda = 0$ then we are done, while if $\lambda > 0$ then $\lambda^{-1} > 1$ and $\lambda^{-1}x \in W$ so $x = \lambda^{-1}x \cdot \lambda + (1 - \lambda)0 \in W$.

- If $x \in W$, $1x \in W$, and $x \in 1W$ thus in both definitions $p(x) \leq 1$.

$\square$

4.2 The Supporting Hyperplane Theorem

The Supporting Hyperplane Theorem follows as an application of the Hahn-Banach Theorem on the Minkowski functional.

We start with a point x_0 outside of a convex subset C and we trace a line from x_0 to a point y in the interior of C. On this line we can define a linear functional f that outputs for each point P on the line, the unique ratio $\lambda \in \mathbb{R}$ such that $\lambda(x_0 - y) = P - y$. Thus $f(x_0 - y) = 1$.

Then this f is bounded above by the Minkowski functional and we can extend it to a functional α defined on the whole space. The hyperplane $\alpha^{-1}(\alpha(x_0)) = y + \alpha^{-1}(1)$, called the supporting hyperplane, then *separates* the point x_0 and C, in the sense that $C \subset y + \alpha^{-1}((-\infty, 1])$ and $x_0 \in y + \alpha^{-1}(1)$.

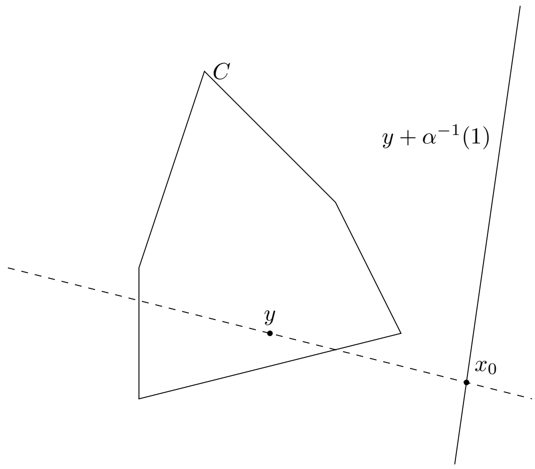

Figure 1: Illustration of the Supporting Hyperplane Theorem. The hyperplane $\alpha^{-1}(\alpha(x_0)) = y + \alpha^{-1}(1)$ is a supporting hyperplane.

The following proof is from [4].

Theorem 4 (Supporting Hyperplane Theorem)**.** *Let C be a convex subset of a real normed vector space X with nonempty interior $y \in \operatorname{int}(C)$, and let $x_0 \in X \setminus C$.*
Then there exists a continuous linear functional $\alpha \in X^*$ such that $1 = \alpha(x_0 - y) \geq \alpha(x - y)$ for all $x \in C$.

Proof. Take a $y \in \operatorname{int}(C)$. Consider the translations $C' := C - y$ and $x_0' := x_0 - y \neq 0$, then we have that $0 \in \operatorname{int}(C')$. Define a linear functional $f : \mathbb{R}x_0' \longrightarrow \mathbb{R}$ as $f(\lambda x_0') := \lambda$, this is well defined because $x_0' \neq 0$. Define the Minkowski functional p on C' as discussed before in equation (8).

- We have $p(x_0') \geq 1$ by the contraposition of the third point of Lemma 3, which implies for $\lambda > 0$ that $p(\lambda x_0') = \lambda p(x_0') \geq \lambda = f(\lambda x_0')$ by positive homogeneity.
- We also know $f(-x_0') = -1 \leq 0 \leq p(-x_0')$ by non-negativity of p, which means for $\lambda \geq 0$ that $f(-\lambda x_0') = -\lambda \leq 0 \leq p(-\lambda x_0')$.

Thus $f(x) \leq p(x)$ for all $x \in \mathbb{R}x_0'$. Then by the Hahn-Banach Theorem, there is an extension $\alpha \in X^*$ such that $\alpha(x) \leq p(x)$ for all $x \in X$.

We remark that $\alpha(x_0') = f(x_0') = 1 \geq p(x)$ for all $x \in C'$ by point 4 of Lemma 3, which means $\alpha(x_0') \geq p(x) \geq \alpha(x)$ for all $x \in C'$. Thus if we translate back, for all $x \in C$ we have $\alpha(x_0 - y) \geq \alpha(x - y)$.

Finally we show continuity of α. Because $0 \in \mathrm{int}(C')$, there exists a $\delta > 0$ such that $B(0, \delta) \subset \mathrm{int}(C')$. Let $\epsilon > 0$, then for $\|y\| < \delta\epsilon$, we have $\pm\frac{y}{\epsilon} \in \mathrm{int}(C') \subset C$ which means $p(\pm\frac{y}{\epsilon}) \leq 1$ by point 4 of Lemma 3. Thus $\pm\alpha(\frac{y}{\epsilon}) = \alpha(\pm\frac{y}{\epsilon}) \leq p(\pm\frac{y}{\epsilon}) \leq 1$, so $|\alpha(y)| \leq \epsilon$, which proves continuity of α.

This completes the proof. □

4.3 The Hyperplane Separation Theorem

The hyperplane separation theorem says that two convex sets where one of them has an interior point can be separated by a hyperplane. This is a corollary of the supporting hyperplane theorem, because we can reduce one of the convex sets to the origin and enlarge the other convex set by taking the Minkowski sum.

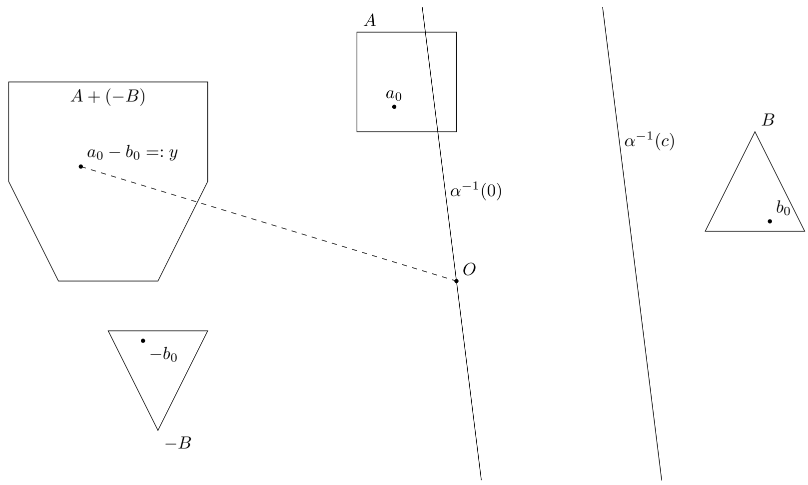

Figure 2: Illustration of the Hyperplane Separation Theorem. One of the possible separating hyperplanes is illustrated by $\alpha^{-1}(c)$.

The proof is from [4]. Before we prove this, we show a lemma about convex sets and Minkowski sums.

Lemma 5. Let A, B be convex subsets of a real vector space X. Then $A + B := \{a + b \mid a \in A, b \in B\}$ is convex. Here $A + B$ is called the *Minkowski sum* of A and B.

Proof. Let $(1 - t)(a_1 + b_1) + t(a_2 + b_2)$ be a convex combination in $A + B$. Then this equals $(1 - t)a_1 + ta_2 + (1 - t)b_1 + tb_2$, and by convexity of A and B, the summands are respectively in A and B, thus their sum is in $A + B$. Thus $A + B$ is convex. □

Theorem 6 (Hyperplane Separation Theorem)**.** *Let A, B be nonempty disjoint subsets of a normed real vector space X, and* $\mathrm{int}(A) \neq \emptyset$. *Then there is a continuous linear functional $\alpha \in X^*$ and a $c \in \mathbb{R}$ such that $\alpha(a) \leq c \leq \alpha(b)$ for all $a \in A$ and $b \in B$.*

Proof. Consider $Z := A - B = A + (-B) = \{a - b \mid a \in A, b \in B\}$, then if B is convex then $-B$ also, and $A + (-B)$ as well by Lemma 5. Note that $0 \notin Z$ because A and B are disjoint. Furthermore if $a_0 \in \mathrm{int}(A)$, then there exists a neighbourhood $B(a_0, \delta) \subset A$ of a_0. Fix some $b_0 \in B \neq \emptyset$ and then $B(a_0 - b_0, \delta) = B(a_0, \delta) - b_0$ is a neighbourhood of $a_0 - b_0$ contained in $A - B = Z$, which means $a_0 - b_0 \in \mathrm{int}(Z) \neq \emptyset$.

Thus Z satisfies the conditions of Theorem 4 by taking $x_0 = 0 \in X \setminus Z$ and $y := a_0 - b_0 \in \text{int}(Z)$. This means we get a continuous linear functional $\alpha \in X^*$ with $1 = \alpha(0-y) \geq \alpha((a-b)-y)$ for all $a - b \in Z = A - B$. Then by linearity of α we get $\alpha(a-b) \leq \alpha(0) = 0$.

Thus $\alpha(a) \leq \alpha(b)$ for any $a \in A, b \in B$, and it follows that

$$\sup_{a\in A} \alpha(a) \leq \inf_{b\in B} \alpha(b). \tag{11}$$

Then because $a_0 \in A, b_0 \in B$, we have

$$\alpha(a_0) \leq \sup_{a\in A} \alpha(a) \leq \inf_{b\in B} \alpha(b) \leq \alpha(b_0) \tag{12}$$

so the infimum and the supremum exist.

Now choose a value $c \in \mathbb{R}$ such that $\sup_{a\in A} \alpha(a) \leq c \leq \inf_{b\in B} \alpha(b)$. For this choice of c we have $\alpha(a) \leq c \leq \alpha(b)$ for all $a \in A$, $b \in B$ which means the hyperplane $\alpha^{-1}(c)$ separates A and B. □

References

[1] Israel Gohberg, Seymour Goldberg and Marinus A. Kaashoek. *Basic Classes of Linear Operators*. Birkhäuser Basel, 2003. DOI: 10.1007/978-3-0348-7980-4. URL: https://doi.org/10.1007/978-3-0348-7980-4.

[2] Alexander Gorodnik. *Lecture notes on Functional Analysis*. URL: https://www.math.uzh.ch/gorodnik/functional_analysis/lecture6.pdf.

[3] Heinz Hanßmann. *Functionaalanalyse*. Epsilon Uitgaven, Amsterdam, 2015.

[4] Robert Peng. *The Hahn-Banach separation theorem and other separation results*. URL: http://math.uchicago.edu/ may/REU2014/REUPapers/Peng.pdf.