Caltech Division of the Humanities and Social Sciences

Ec 181 KC Border
Convex Analysis and Economic Theory AY 2019–2020

Topic 8: Separation theorems

8.1 Hyperplanes and half spaces

Recall that a hyperplane in $\boldsymbol{R}^{\mathrm{m}}$ is a level set $\{p = \alpha\}$ of a nonzero real-valued linear function p.[1] The vector p can be thought of as a real-valued linear function on $\boldsymbol{R}^{\mathrm{m}}$, or as a vector normal (orthogonal) to the hyperplane at each point. Multiplying p and α by the same nonzero scalar does not change the hyperplane. Note that a hyperplane is an affine subspace. And all hyperplanes corresponding to p are parallel. Rewrite.

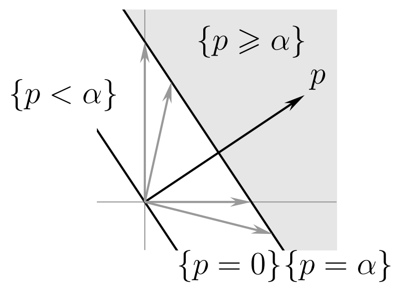


Figure 8.1.1. Parallel hyperplanes $\{p = 0\}$, $\{p = \alpha\}$, and the half space $\{p \geqslant \alpha\}$.

A **weak half space** or **closed half space** is a set of the form $\{p \geqslant \alpha\}$ or $\{p \leqslant \alpha\}$, while a **strict half space** or **open half space** is of the form $\{p > \alpha\}$ or $\{p < \alpha\}$.

8.2 Separating convex sets with hyperplanes

We say that nonzero p, or the hyperplane $\{p = \alpha\}$, **separates** A and B if either

$$A \subset \{p \geqslant \alpha\} \text{ and } B \subset \{p \leqslant \alpha\}, \quad \text{or} \quad B \subset \{p \geqslant \alpha\} \text{ and } A \subset \{p \leqslant \alpha\}.$$

Let us agree that

$$p \cdot A \geqslant p \cdot B \text{ means } p \cdot x \geqslant p \cdot y \text{ for all } x \text{ in } A \text{ and } y \text{ in } B.$$

[1] In more general vector spaces, a hyperplane is a level set $\{f = \alpha\}$ of a nonzero real-valued linear function (or *functional*, as they are more commonly called). If the linear functional is not continuous, then the hyperplane is dense. If the function is continuous, then the hyperplane is closed. See [1, Lemma 5.55, p. 198]. Open and closed half spaces are topologically open and closed if and only if the functional is continuous.

Note that if a set A is a nonempty subset of the hyperplane $\{p = \alpha\}$, then the half-spaces $\{p \leqslant \alpha\}$ and $\{p \geqslant \alpha\}$ separate A and itself, so separation by itself is not very interesting. A better notion is proper separation.

- Say that nonzero p, or the hyperplane $\{p = \alpha\}$ **properly separates** A and B if it separates them and it is not the case that $A \cup B \subset \{p = \alpha\}$, that is, if there exists some x in A and y in B such that $p \cdot x \neq p \cdot y$.

There are stronger notions of separation.

- The hyperplane $\{p = \alpha\}$ **strictly separates** A and B if A and B are in disjoint open half spaces, that is,

$$A \subset \{p > \alpha\} \text{ and } B \subset \{p < \alpha\}$$

(or vice versa).

- It **strongly separates** A and B if A and B are in disjoint closed half spaces. That is, there is some $\varepsilon > 0$ such that

$$A \subset \{p \geqslant \alpha + \varepsilon\} \text{ and } B \subset \{p \leqslant \alpha - \varepsilon\}$$

(or vice versa). An equivalent way to state strong separation is that

$$\inf_{x \in A} p \cdot x > \sup_{y \in B} p \cdot y$$

(or swap A and B).

8.2.1 Example (Kinds of separation) Let

$$A = \{(x, y) \in \boldsymbol{R}^2 : x > 0,\ y \geqslant 1/x\}, \quad B = \{(x, y) \in \boldsymbol{R}^2 : x < 0,\ y \geqslant -1/x\}$$

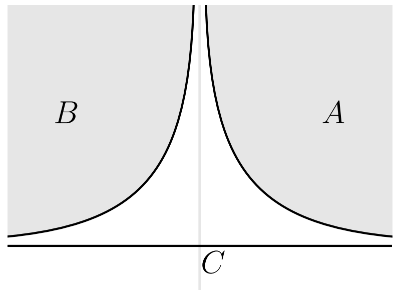
and

$$C = \{(x, y) \in \boldsymbol{R}^2 : y = 0\} \quad \text{(the } x\text{-axis)}.$$

See Figure 8.2.1. Then $p = (1, 0)$ strictly separates A and B with $A \subset \{p > 0\}$ and $B \subset \{p < 0\}$.

There is another notion of separation that is not as useful as the ones above, but I mention it here to eliminate possible confusion. Let us agree to say that p **strictly algebraically separates** A and B if $p \cdot x > p \cdot y$ for all $x \in A$ and $y \in B$ (or vice versa).[2] It should be clear that strict separation implies strict algebraic separation, but the converse is not true: For $q = (0, 1)$ the hyperplane $\{q = 0\}$, which is simply the x-axis properly separates A and C and q strictly algebraically separates A and C, but no hyperplane strictly separates A and C.

[2] This is not standard terminology, but useful to make this one particular point.



$A = \{(x,y) : x > 0,\ y \geqslant 1/x\}$, $B = \{(x,y) : x < 0,\ y \geqslant -1/x\}$,
$C = \{(x,y) : y = 0\}$.

Figure 8.2.1. A and B can be strictly separated; A and C cannot.

Let

$$E = \{(\xi,\eta) : \eta < 0 \text{ or } [\eta = 0 \text{ and } \xi < 0]\} \quad \text{and} \quad F = \{(\xi,\eta) : \eta > 0 \text{ or } [\eta = 0 \text{ and } \xi > 0]\}.$$

See Figure 8.2.4. These are disjoint and convex. Any nonzero vector p that properly separates E and F is of the form of the form $p = (0, p_2)$ where $p_2 \neq 0$. For any such p, the points $x = (-1,0) \in E$ and $b = (1,0) \in F$ satisfy $p \cdot x = p \cdot y = 0$. In particular, E and F cannot be strictly algebraically separated. □

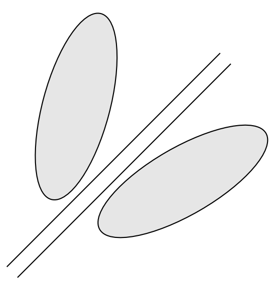


Figure 8.2.2. Strong separation.

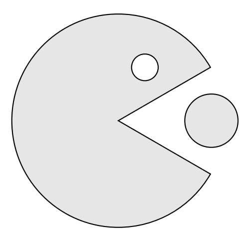


Figure 8.2.3. These sets cannot be separated with a crowbar.

8.2.2 Example In $\boldsymbol{R}^2$, consider a line L and a point x not on L. Any nonzero vector orthogonal to the line defines a linear function that strongly separates x from L. However, almost any perturbation of the function (except scalar multiplication) cannot separate them.

However, if a point in $\boldsymbol{R}^{\mathrm{m}}$ is disjoint from a compact convex set, then a whole open set of vectors defines linear functions that strongly separate them. □

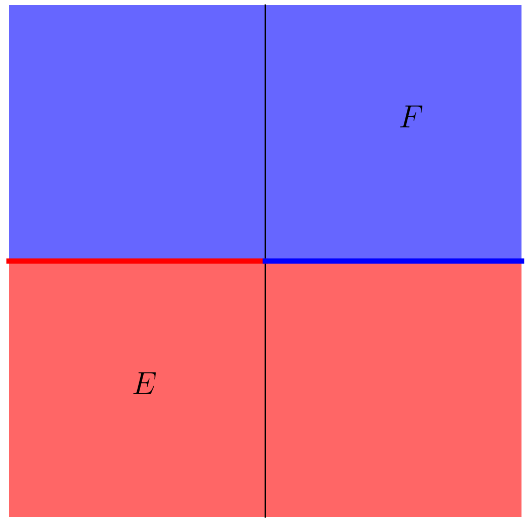


Figure 8.2.4. Two sets that can be properly separated but cannot be strictly algebraically separated are $E = \{(\xi,\eta) : \eta < 0 \text{ or } [\eta = 0 \text{ and } \xi < 0]\}$ and $F = \{(\xi,\eta) : \eta > 0 \text{ or } [\eta = 0 \text{ and } \xi > 0]\}$.

8.2.3 Exercise Prove the last assertion of the above example. □

Here are some simple results that are used so commonly that they are worth noting.

8.2.4 Exercise Prove the following.

Let A and B be disjoint nonempty convex subsets of $\boldsymbol{R}^{\mathrm{m}}$ and suppose nonzero p in $\boldsymbol{R}^{\mathrm{m}}$ properly separates A and B with

$$p \cdot A \geqslant p \cdot B.$$

1. If A is a linear subspace, then p *annihilates* A. That is, $p \cdot x = 0$ for every x in A.
2. If A is a cone, then $p \cdot x \geqslant 0$ for every x in A.
3. If B is a cone, then $p \cdot x \leqslant 0$ for every x in B.
4. If A includes a set of the form $x + \boldsymbol{R}^{\mathrm{m}}_{++}$, then $p > 0$.
5. If B includes a set of the form $x - \boldsymbol{R}^{\mathrm{m}}_{++}$, then $p > 0$.

Hint: Look at the proof of the Nonnegativity Test 0.1.1. □

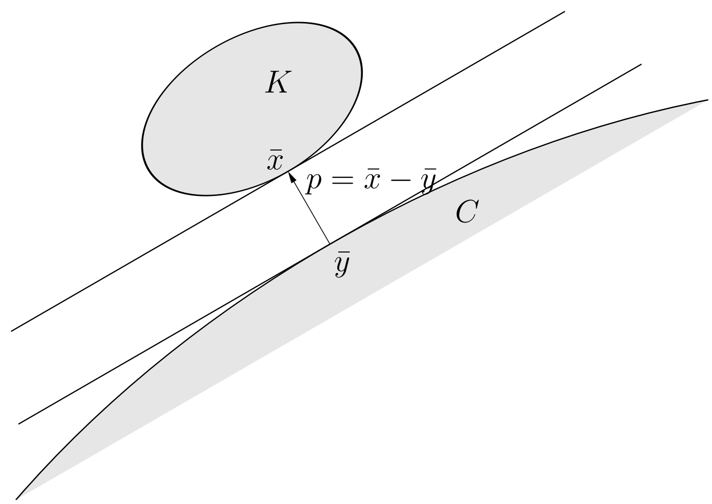


Figure 8.3.1. Minimum distance and separating hyperplanes.

8.3 Strong separating hyperplane theorem

We now come to my personal favorite result on separation of convex sets. I prove the result for Hilbert spaces of arbitrary dimension, since the proof is not much different from the proof for $\boldsymbol{R}^{\mathrm{m}}$, although the theorem is true in general locally convex spaces. Unfortunately, for general locally convex infinite-dimensional spaces, a different proof is needed.

The proof of the following theorem makes use of some facts that are presented in the Appendices.

8.3.1 Strong Separating Hyperplane Theorem *Let K and C be disjoint nonempty convex subsets of a Hilbert space. Suppose K is compact and C is closed. Then there exists a nonzero p that strongly separates K and C.*

Proof: The distance function $d(x, C)$ is a continuous function of x (Theorem A.6.1), so by the Weierstrass Theorem it achieves a minimum on K at some point $\bar{x}$. By Corollary A.15.2 on Metric Projectionthere is some point $\bar{y}$ in C such that

$$d(\bar{x}, \bar{y}) = d(\bar{x}, C) = \min\{d(\bar{x}, y) : y \in C\}.$$

Put

$$p = \bar{x} - \bar{y}. \tag{1}$$

See Figure 8.3.1. Since K and C are disjoint, we must have $p \neq 0$, so its norm satisfies $0 < \|p\|^2 = p \cdot p = p \cdot (\bar{x} - \bar{y})$, so

$$p \cdot \bar{x} > p \cdot \bar{y}.$$

What remains to be shown is that $p \cdot x \geqslant p \cdot \bar{x}$ for all $x \in K$ and $p \cdot \bar{y} \geqslant p \cdot y$ for all $y \in C$.

So let y belong to C. Since $\bar{y}$ minimizes the distance (and hence the square of the distance) to $\bar{x}$ over C, for any point

$$z = \bar{y} + \lambda(y - \bar{y}) \tag{2}$$

with $0 < \lambda \leqslant 1$ on the line segment between y and $\bar{y}$ we have

$$\|\bar{x} - z\|^2 \geqslant \|\bar{x} - \bar{y}\|^2 = \|p\|^2. \tag{3}$$

By (2) we have $\bar{x} - z = \bar{x} - \bar{y} - \lambda(y - \bar{y}) = p - \lambda(y - \bar{y})$, so we may rewrite (3) as

$$\begin{aligned} 0 &\geqslant \|p\|^2 - \big(p - \lambda(y - \bar{y})\big) \cdot \big(p - \lambda(y - \bar{y})\big) \\ &= \|p\|^2 - p \cdot p + 2\lambda p \cdot (y - \bar{y}) - \lambda^2 (y - \bar{y}) \cdot (y - \bar{y}) \\ &= 2\lambda p \cdot (y - \bar{y}) - \lambda^2 (y - \bar{y}) \cdot (y - \bar{y}). \end{aligned}$$

Divide by $2\lambda > 0$ to get

$$0 \geqslant p \cdot (y - \bar{y}) - \tfrac{\lambda}{2}(y - \bar{y}) \cdot (y - \bar{y}).$$

Letting $\lambda \downarrow 0$, we conclude $p \cdot \bar{y} \geqslant p \cdot y$.

A similar argument for $x \in K$ completes the proof. ∎

This proof is a hybrid of several others. The manipulation in the last series of inequalities appears in von Neumann and Morgenstern [7, Theorem 16.3, pp. 134–38], and is probably older. The role of the parallelogram identity (used in the proof of Corollary A.15.2) is well known, see for instance, Hiriart-Urruty and Lemaréchal [3, pp. 41, 46] or Rudin [6, Theorem 12.3, p. 293]. You can replace the last step of dividing by λ with the Kuhn–Tucker conditions for a minimum at $\lambda = 0$ to deduce $p \cdot \bar{y} \geqslant p \cdot y$. A different proof for $\boldsymbol{R}^{\mathrm{m}}$ appears in Rockafellar [5, Corollary 11.4.2, p. 99].

Theorem 8.3.1 is true in general locally convex spaces, where p is interpreted as a continuous linear functional and $p \cdot x$ is replaced by $p(x)$. (But remember, compact sets can be rare in such spaces.) Roko and I give a proof of the general case in [1, Theorem 5.79, p. 207], or see Dunford and Schwartz [2, Theorem V.2.10, p. 417].

8.3.2 Corollary *Let C be a nonempty closed convex subset of a Hilbert space. Assume that the point x does not belong to C. Then there exists a nonzero p that strongly separates x and C.*

8.3.3 Corollary (Bipolar theorem) *If A is a nonempty subset of $\boldsymbol{R}^{\mathrm{m}}$, its double dual cone A^{**} is the closed convex conical hull of A.*

Proof: Let C be the closed convex conical hull of A. Exercise 3.2.2 implies that $C^* = A^*$, and $C \subset A^{**}$. Pick $x \notin C$ and let p be a nonzero vector strongly separating x from C. Since C is a cone, by multiplying by -1 if necessary, we can rig it so that

$$p \cdot x > 0 \geqslant p \cdot C.$$

The second inequality shows that $p \in C^*$, so the first shows that $x \notin C^{**} = A^{**}$. Thus $x \notin C$ implies $x \notin A^{**}$, so $C = A^{**}$. ∎

8.3.4 Theorem *Let A be a subset of $\boldsymbol{R}^{\mathrm{m}}$. Then*

$$\overline{\mathrm{co}}\, A = \bigcap \{H : A \subset H \text{ and } H \text{ is a closed half space}\}.$$

In particular, a closed convex set is the intersection of all the closed half spaces that include it.

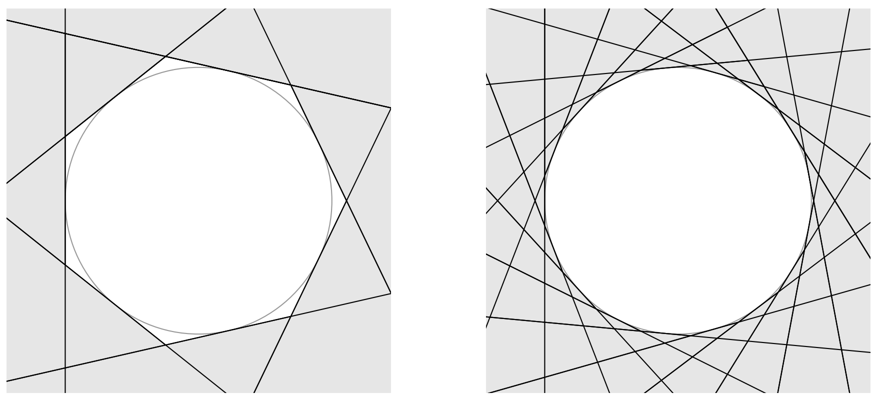


Figure 8.3.2. A disk is an intersection of (infinitely many) supporting half spaces.

Proof: Clearly $\overline{\mathrm{co}}\, A$ is included in the intersection since every closed half space is also a closed convex set. It is also clear that the result is true for $A = \varnothing$. So assume A, and hence $\overline{\mathrm{co}}\, A$, is nonempty.

It suffices to show that if $x \notin \overline{\mathrm{co}}\, A$, then there is a closed half space that includes $\overline{\mathrm{co}}\, A$ but does not contain x. By the Strong Separating Hyperplane Theorem 8.3.1 there is a nonzero p that strongly separates the closed convex set $\overline{\mathrm{co}}\, A$ from the compact convex set $\{x\}$. But this just means there is closed half space $\{p \geqslant \alpha\}$ includes $\overline{\mathrm{co}}\, A$, but doesn't contain x. ∎

8.4 Supporting hyperplanes

8.4.1 Definition *Let C be a nonempty set in a topological vector space and let x be a point belonging to C. The nonzero real-valued linear function p* ***supports***

***C* at *x* (as a minimizer)** *if*

$$\text{for all } y \in C, \quad p \cdot y \geqslant p \cdot x.$$

In this case, we may write $p \cdot C \geqslant p \cdot x$.

We say that p **supports *C* at *x* (as a maximizer)** *if*

$$\text{for all } y \in C, \quad p \cdot x \geqslant p \cdot y.$$

In this case, we may write $p \cdot x \geqslant p \cdot C$.

The hyperplane $\{y : p \cdot y = p \cdot x\}$ *is called a* **supporting hyperplane for *C* at *x***. *We may also say that the half-space* $\{z : p \cdot z \geqslant p \cdot x\}$ *supports* C *at* x *if* p *supports* C *at* x *as a minimizer, etc.*

The support is **proper** *if* $p \cdot y \neq p \cdot x$ *for some* y *in* C.

8.4.2 Lemma *If* p *properly supports the nonempty convex set* C *at* x, *then the relative interior of* C *does not meet the supporting hyperplane. That is, if* $p \cdot C \geqslant p \cdot x$, *then* $p \cdot y > p \cdot x$ *for all* $y \in \operatorname{ri} C$.

Picture?

Proof: Let p properly support C at x, say $p \cdot C \geqslant p \cdot x$, and let y belong to $\operatorname{ri} C$. By the definition of proper support, there exists $z \in C$ with $p \cdot z > p \cdot x$. Since $y \in \operatorname{ri} C$ and $z \in C$, there is some $\varepsilon > 0$ such that

$$w = y + \varepsilon(y - z) \text{ belongs to } C.$$

Then

$$y = (1 - \lambda)z + \lambda w, \qquad \text{where } \lambda = 1/(1 + \varepsilon).$$

Thus

$$p \cdot y = (1 - \lambda)p \cdot z + \lambda p \cdot w > p \cdot x,$$

since $p \cdot w \geqslant p \cdot x$ (as $w \in C$) and $p \cdot z > p \cdot x$. ∎

8.4.3 Remark Note that if C is a singleton, then C can never be properly supported because C will alway lie entirely within the supporting hyperplane. But if $C = \{x\}$, we have $C = \operatorname{ri} C$, so the following theorem still works.

8.4.4 Finite Dimensional Supporting Hyperplane Theorem *Let* C *be a nonempty convex subset of* $\boldsymbol{R}^{\mathrm{m}}$ *and let* $\bar{x}$ *belong to* C. *Then there is a hyperplane properly supporting* C *at* $\bar{x}$ *if and only if* $\bar{x} \notin \operatorname{ri} C$.

Proof: ($\implies$) This is just Lemma 8.4.2.

($\impliedby$) Without loss of generality, we can translate C by $-\bar{x}$, and thus assume $\bar{x} = 0$. (See Exercises 4.2.1 and 4.2.3.)

Assume $0 \notin \operatorname{ri} C$. (This implies that C is not a singleton, and also that $C \neq \boldsymbol{R}^{\mathrm{m}}$.) Define

$$A = \bigcup_{\lambda > 0} \lambda \operatorname{ri} C.$$

Clearly $\operatorname{ri} C \subset A$, A lies in the span of $\operatorname{ri} C$, and $0 \notin A$ but $0 \in \overline{A}$.

More importantly, A is convex. To see this observe that it is enough to show that $\kappa\Big(A \times A \times (0,1)\Big) \subset A$. We shall show the stronger inclusion that if $\alpha, \beta > 0$, then $\alpha A + \beta A \subset A$ (cf. Exercise 1.1.7.6). So let $x, y \in A$ where $x = \lambda x'$, $y = \mu y'$, where $x', y' \in \operatorname{ri} C$ and $\lambda, \mu > 0$. Then

$$\alpha x + \beta y = \alpha\lambda x' + \beta\mu y' = (\alpha\lambda + \beta\mu)\frac{\alpha\lambda x' + \beta\mu y'}{\alpha\lambda + \beta\mu} \in (\alpha\lambda + \beta\mu) \operatorname{ri} C \subset A,$$

where the containment follows from the convexity of $\operatorname{ri} C$.

Believe it or not, the crux of the proof is showing that there is some point somewhere that does not belong to the closure of A. That is what the next paragraph is about.

Since $\boldsymbol{R}^{\mathrm{m}}$ is finite dimensional there exists a finite maximal collection of linearly independent vectors $v_1, \dots, v_k$ that lie in $\operatorname{ri} C$. Since $\operatorname{ri} C$ contains at least one nonzero point, we have $k \geqslant 1$. Let $v = \frac{1}{k}\sum_{i=1}^k v_i$, which belongs to $\operatorname{ri} C$. I claim that $-v \notin \overline{A}$. See Figure 8.4.1. To see this, assume by way of contra-

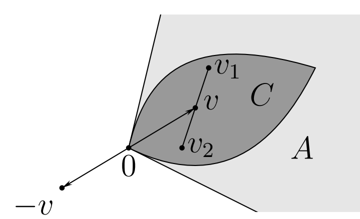


Figure 8.4.1

diction that $-v$ belongs to $\overline{A}$. Thus, there exists a sequence $\{x_n\}$ in A satisfying $x_n \to -v$. Since $v_1, \dots, v_k$ is a maximal independent set, we must be able to write $x_n = \sum_{i=1}^k \lambda_i^n v_i$. By Lemma A.12.1 on continuity of coordinates, $\lambda_i^n \xrightarrow[n\to\infty]{} -1/k$ for each i. In particular, for some n we have $\lambda_i^n < 0$ for each i. For this n let $\lambda = \sum_{i=1}^k \lambda_i^n < 0$, then

$$0 = \tfrac{1}{1-\lambda} x_n + \sum_{i=1}^k \Big(\tfrac{-\lambda_i^n}{1-\lambda}\Big) v_i \in A, \quad \text{as } A \text{ is convex,}$$

which is a contradiction. Hence $-v \notin \overline{A}$.

Now by Corollary 8.3.2 there exists some nonzero p strongly separating $-v$ from $\overline{A}$. That is, $p \cdot (-v) < p \cdot y$ for all $y \in \overline{A}$. Moreover, since $\overline{A}$ is a cone, $p \cdot y \geqslant 0 = p \cdot 0$ for all $y \in \overline{A}$, and $p \cdot (-v) < 0$ (Exercise 8.2.4). Thus p supports $\overline{A} \supset C$ at 0. Moreover, $p \cdot v > 0$, so p properly supports C at 0. ∎

Now I'll state without proof some general theorems that apply in infinite dimensional spaces.

8.4.5 Infinite Dimensional Supporting Hyperplane Theorem *If C is a convex set with nonempty interior in a topological vector space, and x is a boundary point of C, then there is a nonzero continuous linear functional properly supporting C at x.*

For a proof see [1, Lemma 7.7, p. 259]. If a convex set has an empty interior, then it may fail to have supporting closed hyperplanes at boundary points. For example, in ℓ_1, the positive cone P has an empty interior, so every point of P is a boundary point, but P cannot be supported by a continuous linear functional at any strictly positive sequence, see [1, Example 7.8, p. 259]. However, in Banach spaces we have the following result, a proof of which is in [1, Theorem 7.43, p. 284].

Where do we show that $\operatorname{int} \ell_{1+} = \varnothing$?

8.4.6 Bishop–Phelps Theorem *Let C be a nonempty closed convex subset of a Banach space. Then the set of points at which C is supported by a nonzero continuous linear functional is dense in the boundary of C.*

8.5 More separating hyperplane theorems

The next theorem yields only proper separation but requires only that the sets in question have disjoint relative interiors. In particular it applies whenever the sets themselves are disjoint. It is a strictly finite-dimensional result.

8.5.1 Finite Dimensional Separating Hyperplane Theorem *Two nonempty convex subsets of $\boldsymbol{R}^{\mathrm{m}}$ can be properly separated by a hyperplane if and only if their relative interiors are disjoint.*

Proof: ($\Longleftarrow$) Let A and B be nonempty convex subsets of $\boldsymbol{R}^{\mathrm{m}}$ with $\operatorname{ri} A \cap \operatorname{ri} B = \varnothing$. Put $C = A - B$. By Proposition 5.2.9 $\operatorname{ri} C = \operatorname{ri} A - \operatorname{ri} B$, so $0 \notin \operatorname{ri} C$. It suffices to show that there exists some nonzero $p \in \boldsymbol{R}^{\mathrm{m}}$ satisfying $p \cdot x \geqslant 0$ for all $x \in C$, and $p \cdot y > 0$ for some $y \in C$. If $0 \notin \overline{C}$, this follows from Corollary 8.3.2. If $0 \in \overline{C}$, it follows from Theorem 8.4.4.

($\Longrightarrow$) If p properly separates A and B, then the same argument used in the proof of Theorem 8.4.4 shows that $\operatorname{ri} A \cap \operatorname{ri} B = \varnothing$. ∎

Finally, Theorem 8.4.5 can be used to prove the following.

8.5.2 Infinite Dimensional Separating Hyperplane Theorem *Two disjoint nonempty convex subsets of a topological vector space can be properly separated by a closed hyperplane (or continuous linear functional) if one of them has a nonempty interior.*

For the sake of completeness, I mention without proof an additional result, due to Klee [4], which deals with strict separation in finite dimensional spaces.

8.5.3 Theorem (Strict separation) *Two disjoint nonempty closed convex subsets of a finite dimensional vector space can be strictly separated by a hyperplane if neither includes a half-line in its boundary.*

References

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

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

[3] J.-B. Hiriart-Urruty and C. Lemaréchal. 2001. *Fundamentals of convex analysis*. Grundlehren Text Editions. Berlin: Springer–Verlag.

[4] V. L. Klee, Jr. 1956. Strict separation of convex sets. *Proceedings of the American Mathematical Society* 7:735–737.
http://www.jstor.org/stable/2033382

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

[6] W. Rudin. 1973. *Functional analysis*. New York: McGraw Hill.

[7] J. von Neumann and O. Morgenstern. 1944. *The theory of games and economic behavior*. Princeton: Princeton University Press. Reprint. New York: John Wiley & Sons, 1967.