CALIFORNIA INSTITUTE OF TECHNOLOGY

Division of the Humanities and Social Sciences

Separating Hyperplane Theorems

KC Border

v. 2012.08.30::11.17

1 Separation by hyperplanes

Given $p \in \boldsymbol{R}^{\mathrm{n}}$ and $\alpha \in \boldsymbol{R}$, let $[p \geqslant \alpha]$ denote the set $\{x \in \boldsymbol{R}^{\mathrm{n}} : p \cdot x \geqslant \alpha\}$ where $p \cdot x$ is the Euclidean inner product $\sum_{i=1}^{n} p_i x_i$. The sets $[p = \alpha]$, etc., are defined similarly. A **hyperplane** in $\boldsymbol{R}^{\mathrm{n}}$ is a set of the form $[p = \alpha]$, where $p \neq 0$.[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. In $\boldsymbol{R}^2$, a hyperplane is just a line, and in $\boldsymbol{R}^3$, a hyperplane is an ordinary plane. In $\boldsymbol{R}^1$, a hyperplane is just a point.

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 $[p < \alpha]$. We say that nonzero p, or the hyperplane $[p = \alpha]$, **separates** A and B if either $A \subset [p \geqslant \alpha]$ and $B \subset [p \leqslant \alpha]$, or $B \subset [p \geqslant \alpha]$ and $A \subset [p \leqslant \alpha]$. Let us agree to write $p \cdot A \geqslant p \cdot B$ to mean $p \cdot x \geqslant p \cdot y$ for all x in A and y in B. The separation is **proper** if there exists some x in A and y in B such that $p \cdot x \neq p \cdot y$. Proper separation guarantees that $A \cup B$ is not wholly included in the hyperplane $[p = \alpha]$.

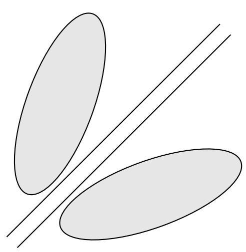

Figure 1. Strong separation.

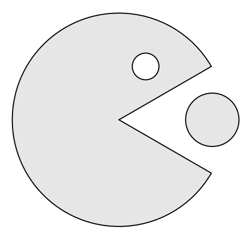

Figure 2. These sets cannot be separated by a hyperplane.

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, $A \subset [p > \alpha]$ 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 $A \subset [p \geqslant \alpha + \varepsilon]$ and $B \subset [p \leqslant \alpha]$ (or vice versa). Another 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).

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**

[1] In more general linear 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 space is a topological vector space and the linear functional is not continuous, the hyperplane is a dense subset. If the function is continuous, then the hyperplane is closed. [1, Lemma 5.55, p.198] Open and closed half spaces are topologically open and closed if and only if the functional is continuous.

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 next example shows the converse is not true.

1 Example (Examples of separation) These examples in $\boldsymbol{R}^2$ illustrate some subtleties. Let

$$A = \{(\xi, \eta) : \eta > 1/|\xi| \text{ and } \xi < 0\}, \quad B = \{(\xi, \eta) : \eta > 1/\xi \text{ and } \xi > 0\}$$

and let C be the ξ-axis,

$$C = \{(\xi, \eta) : \eta = 0\}.$$

See Figure 3.

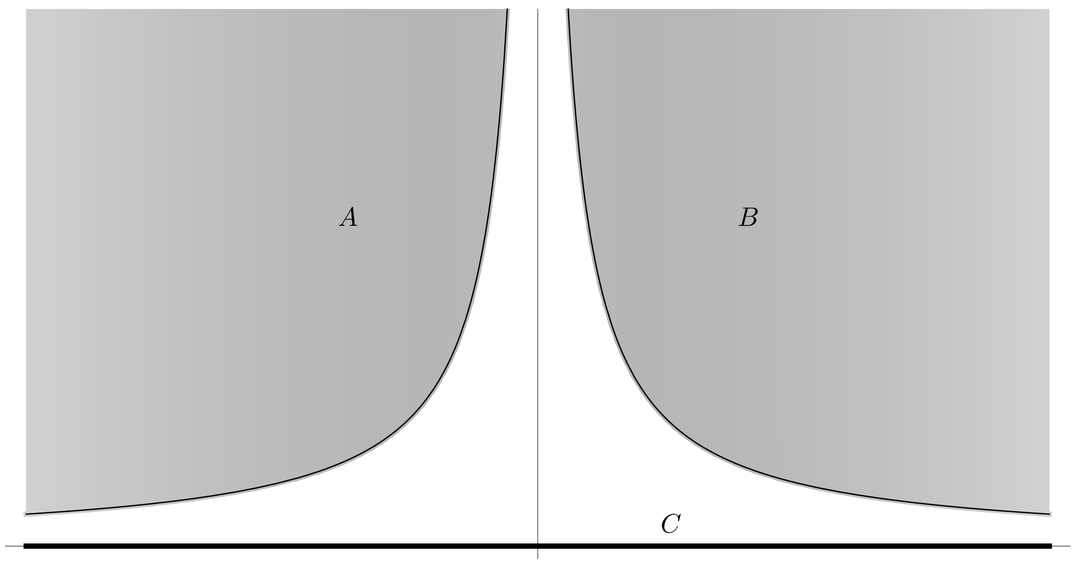

Figure 3. $A = \{(\xi, \eta) : \eta > 1/|\xi| \text{ and } \xi < 0\}$, $B = \{(\xi, \eta) : \eta > 1/\xi \text{ and } \xi > 0\}$, and $C = \{(\xi, \eta) : \eta = 0\}$.

Note that $p = (1, 0)$ strictly separates A and B, but they cannot be strongly separated.

Also, B and C can be strictly algebraically separated, but not strictly separated. Any nonzero vector p that properly separates B and C is of the form of the form $p = (0, p_2)$ where $p_2 \neq 0$. Consider $p_2 > 0$. Then $p \cdot c = 0$ for all $c \in C$ and $p \cdot b > 0$ for every $b \in B$, so p strictly algebraically separates B and C. But $0 \in C$ and $p \cdot 0 = 0$, and the points $b_n = (n, 1/n) \in B$ satisfy $p \cdot b_n \to 0$. Consequently, B and C cannot be strictly separated, that is, put into disjoint open half spaces.

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 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. □

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

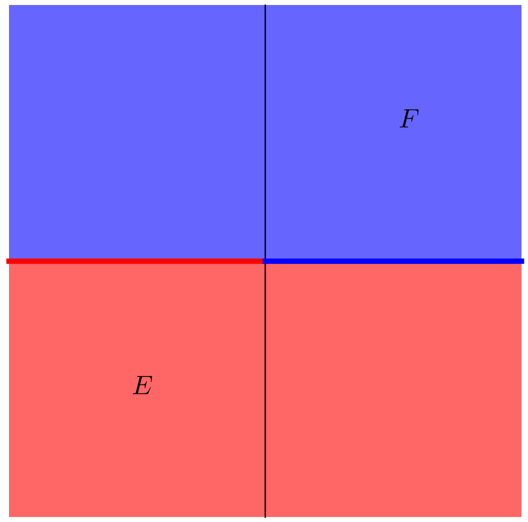

Figure 4. $E = \{(\xi, \eta) : \eta < 0 \text{ or } [\eta = 0 \text{ and } \xi < 0]\}$ and $F = \{(\xi, \eta) : \eta > 0 \text{ or } [\eta = 0 \text{ and } \xi > 0]\}$. These sets cannot be strictly separated.

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. □

3 Exercise *Prove the last assertion of the above example.*

See Appendix B for the definition and properties of the **relative interior** $\operatorname{ri} C$ of a convex set C.

4 Lemma *If p properly separates A and B, then it properly separates $\operatorname{ri} A$ and $\operatorname{ri} B$.*

Proof: Since $A \subset \overline{\operatorname{ri} A}$ and $B \subset \overline{\operatorname{ri} B}$ and $x \mapsto p \cdot x$ is continuous, if p does not properly separate the relative interiors, that is, if $p \cdot \operatorname{ri} A = p \cdot \operatorname{ri} B$, then $p \cdot A = p \cdot B$, and the separation of A and B is not proper. ■

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

5 Lemma *Let A and B be disjoint nonempty convex subsets of $\boldsymbol{R}^{\mathrm{n}}$ and suppose nonzero p in $\boldsymbol{R}^{\mathrm{n}}$ 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 $x + \boldsymbol{R}^{\mathrm{n}}_{++}$, then $p > 0$.*

5. *If B includes a set of the form $x - \boldsymbol{R}^{\mathrm{n}}_{++}$, then $p > 0$.*

Proof: The proofs of all these are more or less the same, so we shall just prove (4). Since p is nonzero by hypothesis, it suffice to show that $p \geqq 0$. Suppose by way of contradiction that $p_i < 0$ for some i. Note that $te^i + \varepsilon\mathbf{1}$ belongs to $\boldsymbol{R}^{\mathrm{m}}_{++}$ for every $t, \varepsilon > 0$. Now $p \cdot (x + te^i + \varepsilon\mathbf{1}) = p \cdot x + tp_i + \varepsilon p \cdot \mathbf{1}$. By letting $t \to \infty$ and $\varepsilon \downarrow 0$ we see that $p \cdot x + tp_i + \varepsilon p \cdot \mathbf{1} \downarrow -\infty$, which contradicts $p \cdot (x + te^i + \varepsilon\mathbf{1}) \geqslant p \cdot y$ for any y in B. Therefore $p > 0$. ∎

2 Separating Hyperplane Theorems

We now come to the main result on separation of convex sets.[3]

6 Strong Separating Hyperplane Theorem *Let K and C be disjoint nonempty convex subsets of $\boldsymbol{R}^{\mathrm{n}}$. Suppose K is compact and C is closed. Then there exists a nonzero $p \in \boldsymbol{R}^{\mathrm{n}}$ that strongly separates K and C.*

Proof: Let $d(x, y) = \|x - y\|$ denote the Euclidean distance on $\boldsymbol{R}^{\mathrm{n}}$. Define $f : K \to \boldsymbol{R}$ by $f(x) = \inf\{d(x, y) : y \in C\}$, that is $f(x)$ is the distance from x to C. The function f is continuous. To see this, observe that for any y, the distance $d(x', y) \leqslant d(x', x) + d(x, y)$ (triangle inequality), and $d(x, y) \leqslant d(x, x') + d(x', y)$. Thus $|d(x, y) - d(x', y)| \leqslant d(x, x')$, so $|f(x) - f(x')| \leqslant d(x, x')$. Thus f is actually Lipschitz continuous.

Since K is compact, f achieves a minimum on K at some point $\bar{x}$. I next claim that there is some point $\bar{y}$ in C such that $f(\bar{x}) = \|\bar{x} - \bar{y}\|$. See Figure 5. The proof of this is is

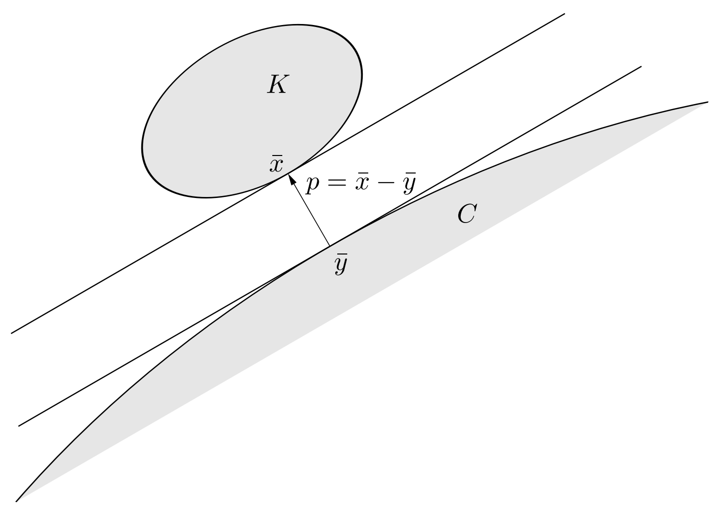

Figure 5. Minimum distance and separating hyperplanes.

straightforward in $\boldsymbol{R}^{\mathrm{n}}$, but I will give a subtler proof that works for an arbitrary Hilbert space.

[3] Theorem 6 is true in general locally convex spaces, where p is interpreted as continuous linear functional. (But remember, compact sets can be rare in such spaces.) The proof I give makes use of some of the special properties of $\boldsymbol{R}^{\mathrm{n}}$, namely that it has an inner product generating its metric. This proof thus works in any Hilbert space, with inner product denoted $p \cdot x$. 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]. A topological vector space is **locally convex** if every neighborhood of zero includes a convex neighborhood of zero. Consequently a neighborhood of any point includes a convex neighborhood of that point. Since this is always true in finite dimensional spaces, you might have trouble thinking of a space that is not locally convex. The ℓ_p-spaces for $0 < p < 1$ are not locally convex.

To see that such a $\bar{y}$ exists, for each n, let $C_n = \{y \in C : d(\bar{x}, y) \leqslant f(\bar{x}) + 1/n\}$. Then each C_n is a nonempty, closed, and convex subset of C, and $C_{n+1} \subset C_n$ for each n. Moreover $f(\bar{x}) = \inf\{d(\bar{x}, y) : y \in C_n\}$, that is, if such a $\bar{y}$ exists, it must be in C_n for every n. I now claim that $\operatorname{diam} C_n = \sup\{d(y_1, y_2) : y_1, y_2 \in C_n\} \to 0$ as $n \to \infty$. To see this, start with the **parallelogram identity**[4]

$$\|x_1 + x_2\|^2 = 2\|x_1\|^2 - \|x_1 - x_2\|^2 + 2\|x_2\|^2.$$

Now let y_1, y_2 belong to C_n. The distance from $\bar{x}$ to the midpoint of the segment joining y_1, y_2 is given by $d(\bar{x}, \frac{1}{2}y_1 + \frac{1}{2}y_2) = \|\frac{1}{2}(y_1 - \bar{x}) + \frac{1}{2}(y_2 - \bar{x})\|$. Evaluate the parallelogram identity for $x_i = \frac{1}{2}(y_i - \bar{x})$ to get

$$d(\bar{x}, \tfrac{1}{2}y_1 + \tfrac{1}{2}y_2)^2 = \tfrac{1}{2}d(y_1, \bar{x})^2 + \tfrac{1}{2}d(y_2, \bar{x})^2 - \tfrac{1}{4}d(y_1, y_2)^2.$$

so rearranging gives

$$d(y_1, y_2)^2 = 2\big[d(y_1, \bar{x})^2 - d(\bar{x}, \tfrac{1}{2}y_1 + \tfrac{1}{2}y_2)^2\big] + 2\big[d(y_2, \bar{x})^2 - d(\bar{x}, \tfrac{1}{2}y_1 + \tfrac{1}{2}y_2)^2\big]. \tag{1}$$

Now for any point $y \in C_n$, we have $f(\bar{x}) \leqslant d(\bar{x}, y) \leqslant f(\bar{x}) + 1/n$, so

$$d(\bar{x}, y)^2 - f(\bar{x})^2 \leqslant \big(f(\bar{x}) + 1/n\big)^2 - f(\bar{x})^2 = 2f(\bar{x})/n + 1/n^2.$$

Now both y_1 and $\frac{1}{2}y_1 + \frac{1}{2}y_2$ belong to C_n, so

$$\begin{aligned}
\big|d(y_1, \bar{x})^2 - d(\bar{x}, \tfrac{1}{2}y_1 + \tfrac{1}{2}y_2)^2\big| &= \big|d(y_1, \bar{x})^2 - f(\bar{x})^2 - \big(d(\bar{x}, \tfrac{1}{2}y_1 + \tfrac{1}{2}y_2)^2 - f(\bar{x})^2\big)\big| \\
&\leqslant \big|d(y_1, \bar{x})^2 - f(\bar{x})^2\big| + \big|d(\bar{x}, \tfrac{1}{2}y_1 + \tfrac{1}{2}y_2)^2 - f(\bar{x})^2\big| \\
&\leqslant 2\big(\tfrac{2}{n}f(\bar{x}) + \tfrac{1}{n^2}\big),
\end{aligned}$$

and similarly for y_2. Substituting this in (1) gives

$$d(y_1, y_2)^2 \leqslant 8(\tfrac{2}{n} + \tfrac{1}{n^2}) \to 0 \text{ as } n \to \infty,$$

so $\operatorname{diam} C_n \to 0$. Since $\boldsymbol{R}^{\mathrm{n}}$ (or any Hilbert space) is complete, the Cantor Intersection Theorem 19 asserts that $\bigcap_{n=1}^{\infty} C_n$ is a singleton $\{\bar{y}\}$. This $\bar{y}$ has the desired property. Whew! (It also follows that $\bar{y}$ is the unique point satisfying $f(\bar{x}) = d(\bar{x}, \bar{y})$, but we don't need to know that.) Maybe I should make this a separate lemma.[5]

Put $p = \bar{x} - \bar{y}$. Since K and C are disjoint, we must have $p \neq 0$. Then $0 < \|p\|^2 = p \cdot p = p \cdot (\bar{x} - \bar{y})$, so $p \cdot \bar{x} > p \cdot \bar{y}$. What remains to be shown is that $p \cdot \bar{y} \geqslant p \cdot y$ for all $y \in C$ and $p \cdot \bar{x} \leqslant p \cdot x$ for all $x \in K$:

---

[4] This says that the sum of the squares of the lengths of the diagonals of a parallelogram is equal to the sum of the squares of the lengths of the sides. (Consider the parallelogram with vertices $0, x_1, x_2, x_1 + x_2$. Its diagonals are the segments $[0, x_1 + x_2]$ and $[x_1, x_2]$, and their lengths are $\|x_1 + x_2\|$ and $\|x_1 - x_2\|$. It has two sides of length $\|x_1\|$ and two of length $\|x_2\|$.) To prove this, note that

$$(x_1 + x_2) \cdot (x_1 + x_2) = x_1 \cdot x_1 + 2x_1 \cdot x_2 + x_2 \cdot x_2$$

$$(x_1 - x_2) \cdot (x_1 - x_2) = x_1 \cdot x_1 - 2x_1 \cdot x_2 + x_2 \cdot x_2.$$

Solve the latter for $2x_1 \cdot x_2$ and substitute in into the former to get

$$(x_1 + x_2) \cdot (x_1 + x_2) = 2x_1 \cdot x_1 - (x_1 - x_2) \cdot (x_1 - x_2) + 2x_2 \cdot x_2,$$

and the desired result is restated in terms of norms.

[5] This argument can be greatly simplified in $\boldsymbol{R}^{\mathrm{n}}$ by noting that the set C_1 is compact, which need not be true in an infinite dimensional Hilbert space. Compactness implies that a sequence $y_n \in C_n$ has a subsequence converging to some point $\bar{y}$ as above.

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 $z = \bar{y} + \lambda(y - \bar{y})$ (with $0 < \lambda \leqslant 1$) on the line segment between y and $\bar{y}$ we have

$$(\bar{x} - z) \cdot (\bar{x} - z) \geqslant (\bar{x} - \bar{y}) \cdot (\bar{x} - \bar{y}).$$

Rewrite this as

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

Divide by $\lambda > 0$ to get

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

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. Hiriart-Urruty and Lemaréchal [4, pp. 41, 46] use the parallelogram identity to show that $\bar{y}$ is unique, but they stop short of computing the diameter of the sets C_n. Those authors rely on the compactness of closed balls in $\boldsymbol{R}^{\mathrm{n}}$ for the existence of $\bar{y}$, so their argument does not generalize to Hilbert spaces. A different proof appears in Rockafellar [6, Corollary 11.4.2, p. 99].

7 Corollary *Let C be a nonempty closed convex subset of $\boldsymbol{R}^{\mathrm{n}}$. Assume that the point x does not belong to C. Then there exists a nonzero $p \in \boldsymbol{R}^{\mathrm{n}}$ that strongly separates x and C.*

8 Definition *Let C be a set in $\boldsymbol{R}^{\mathrm{n}}$ and x a point belonging to C. The nonzero vector p* ***supports C at x*** *if $p \cdot y \geqslant p \cdot x$ for all $y \in C$ (or if $p \cdot y \leqslant p \cdot x$ for all $y \in C$). The hyperplane $\{y : p \cdot y = p \cdot x\}$ is a* ***supporting hyperplane for C at x****. The support is* ***proper*** *if $p \cdot y > p \cdot x$ for some y in C.*

9 Lemma *If p properly supports the 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 $y \in \operatorname{ri} C$.*

Proof: Geometrically, this says that if z is in the hyperplane, and y is on one side, the line through y and z must go through to the other side. Algebraically, let p properly support C at x, say $p \cdot C \geqslant p \cdot x$. Then there exists $y \in C$ with $p \cdot y > p \cdot x$. Let z belong to $\operatorname{ri} C$. By separation $p \cdot z \geqslant p \cdot x$, so suppose by way of contradiction that $p \cdot z = p \cdot x$. Since z is in the relative interior of C, there is some $\varepsilon > 0$ such that $z + \varepsilon(z - y)$ belongs to C. Then $p \cdot \big(z + \varepsilon(z - y)\big) = p \cdot x - \varepsilon p \cdot (x - y) < p \cdot x$, a contradiction. ∎

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

Proof of Theorem 10: ($\Longrightarrow$) This is just Lemma 9.

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

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

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

Clearly $\varnothing \neq \operatorname{ri} C \subset A$, $0 \notin A$ but $0 \in \overline{A}$, and A is a deleted cone. More importantly, A is convex (a simple exercise), and A lies in the span of $\operatorname{ri} C$ (cf. Proposition 16). (The closure of A is called the **tangent cone** to C at 0.)

Since $\boldsymbol{R}^{\mathrm{n}}$ is finite dimensional there exists a finite maximal collection of linearly independent vectors $v_1, \ldots, v_k$ that lie in $\operatorname{ri} C$. Since $\operatorname{ri} C$ contains at least one nonzero point, we have $k \geqslant 1$. Let $v = \sum_{i=1}^k v_i$, and note that $(1/k)v$ belongs to $\operatorname{ri} C$. I claim that $-v \notin \overline{A}$. See Figure 6. To

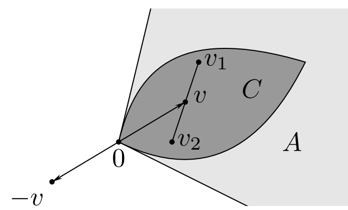

Figure 6.

see this, assume by way of contradiction that $-v$ belongs to $\overline{A}$. Thus, there exists a sequence $\{x_n\}$ in A satisfying $x_n \to -v$. Since $v_1, \ldots, v_k$ is a maximal independent set, we must be able to write $x_n = \sum_{i=1}^k \lambda_i^n v_i$. By Lemma 15 (at the end of this section), $\lambda_i^n \xrightarrow[n\to\infty]{} -1$ for each i. In particular, for some n we have $\lambda_i^n < 0$ for each i. Now if for this n we let $\lambda = \sum_{i=1}^k \lambda_i^n < 0$, then

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

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

Now by Corollary 7 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 $p \cdot (-v) < 0$ (Lemma 5). Thus p supports $\overline{A} \supset C$ at 0. Moreover, $p \cdot (1/k)v > 0$, so p properly supports C at 0. ∎

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.

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

Proof: ($\Longleftarrow$) Let A and B be nonempty convex subsets of $\boldsymbol{R}^{\mathrm{n}}$ with $\operatorname{ri} A \cap \operatorname{ri} B = \varnothing$. Put $C = A - B$. By Proposition 18 $\operatorname{ri} C = \operatorname{ri} A - \operatorname{ri} B$, so $0 \notin \operatorname{ri} C$. It suffices to show that there exists some nonzero $p \in \boldsymbol{R}^{\mathrm{n}}$ 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 7. If $0 \in \overline{C}$, it follows from Theorem 10.

($\implies$) If p properly separates A and B, then the same argument used in the proof of Theorem 10 shows that $\operatorname{ri} A \cap \operatorname{ri} B = \varnothing$. ∎

Recall Example 1, which gave examples of sets that could not be strictly separated. In each case, the boundary of one of these sets included a **half-line**, that is, a set of the form $\{(1-\lambda)x + \lambda y : \lambda \geqslant 0\}$, where $x \neq y$. In finite dimensional spaces Klee [5] has proven that absence of these half-lines is a sufficient condition for strict separation of closed convex sets. Here is the result without proof.

12 Theorem (Strict separation) *A pair of 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.*

Now I'll state without proof a general theorem that applies in infinite dimensional spaces.

13 Infinite Dimensional Separating Hyperplane Theorem *Two disjoint nonempty convex subsets of a (not necessarily locally convex) topological vector space can be properly separated by a closed hyperplane (or continuous linear functional) if one of them has a nonempty interior.*

As an application we have the following result due to Fan, Glicksberg, and Hoffman [3]. It is the basis of the saddlepoint theorem for constrained optimization.

14 Concave Alternative Theorem *Let C be a nonempty convex subset of a vector space, and let $f^1, \dots, f^m \colon C \to \boldsymbol{R}$ be concave. Letting $f = (f_1, \dots, f_m) \colon C \to \boldsymbol{R}^{\mathrm{m}}$, exactly one of the following is true.*

$$\exists \bar{x} \in C \quad f(\bar{x}) \gg 0. \tag{2}$$

Or (exclusive),

$$\exists p > 0 \; \forall x \in C \quad p \cdot f(x) \leqslant 0. \tag{3}$$

Proof: Clearly both cannot be true. Suppose (2) fails. Set

$$H = \{f(x) : x \in C\} \quad \text{and set} \quad \hat{H} = \{y \in \boldsymbol{R}^{\mathrm{m}} : \exists x \in C \quad y \leqq f(x)\}.$$

Since (2) fails, we see that H and $\boldsymbol{R}^{\mathrm{m}}_{++}$ are disjoint. Consequently $\hat{H}$ and $\boldsymbol{R}^{\mathrm{m}}_{++}$ are disjoint. Now observe that $\hat{H}$ is convex. To see this, suppose $y^1, y^2 \in \hat{H}$. Then $y^i \leqq f(x^i)$, $i = 1, 2$. Therefore, for any $\lambda \in (0, 1)$,

$$\lambda y^1 + (1-\lambda) y^2 \leqq \lambda f(x^1) + (1-\lambda) f(x^2) \leqq f\big(\lambda x^1 + (1-\lambda) x^2\big),$$

since each f^j is concave. Therefore $\lambda y^1 + (1-\lambda) y^2 \in \hat{H}$.

Thus, by the Separating Hyperplane Theorem 11, there is a nonzero vector $p \in \boldsymbol{R}^{\mathrm{m}}$ properly separating $\hat{H}$ and $\boldsymbol{R}^{\mathrm{m}}_{++}$. We may assume

$$p \cdot \hat{H} \leqslant p \cdot \boldsymbol{R}^{\mathrm{m}}_{++}. \tag{4}$$

By Lemma 5, $p > 0$. Evaluating (4) at $z = \varepsilon \mathbf{1}$ for $\varepsilon > 0$, we get $p \cdot y \leqslant \varepsilon p \cdot \mathbf{1}$. Since ε may be taken arbitrarily small, we conclude that $p \cdot y \leqslant 0$ for all y in $\hat{H}$. In particular, $p \cdot f(x) \leqslant 0$ for all x in C. ∎

A Continuity of the coordinate map

The following lemma may be obvious to you. Certainly many other authors take it for granted.

15 Lemma *Let $v_1, \dots, v_p$ be linearly independent vectors in $\boldsymbol{R}^{\mathrm{m}}$. Let $x_n = \sum_{i=1}^p \lambda_i^n v_i$. If the sequence x_n converges, say $x_n \to x$, then for each $i = 1, \dots, p$, the sequence λ_i^n converges, say $\lambda_i^n \to \lambda_i$, and $x = \sum_{i=1}^p \lambda_i v_i$.*

Proof of Lemma 15: Let V denote the $m \times p$ matrix whose columns are $v_1, \dots, v_p$, and observe that V has full rank. For each n there is a vector λ_n in $\boldsymbol{R}^{\mathrm{p}}$ satisfying

$$V \lambda_n = x_n.$$

Since V has full rank, so does the $p \times p$ matrix $V'V$. Define the p-vector λ by

$$\lambda = (V'V)^{-1}V'x.$$

That is, λ is the coefficient vector of the ordinary least squares regression of x on the v_is.[6] Or in mathematical terms $V\lambda$ is the orthogonal projection of x on the space spanned by the v_is. Since $x_n \to x$,

$$(V'V)^{-1}V'x_n \to (V'V)^{-1}V'x = \lambda,$$

but

$$(V'V)^{-1}V'y_n = (V'V)^{-1}V'V\lambda_n = \lambda_n,$$

so $\lambda_n \to \lambda$. This implies that $V\lambda_n \to V\lambda$. But $V\lambda_n \to x$, so we conclude $x = V\lambda$. ∎

B Relative interior of a convex set

A set A in a vector space is **affine** if it includes all the lines (not just line segments) generated by its points. For the algebraically inclined, A is affine if $x, y \in A$ imply $\lambda x + (1-\lambda)y \in A$ for all $\lambda \in \boldsymbol{R}$. Intersections of affine sets are affine, so every set is included in a smallest affine set, called its **affine hull**. A set A is affine if and only if for each $x \in A$, the set $A - x$ is a linear subspace (prove it as an exercise). In $\boldsymbol{R}^{\mathrm{n}}$, every linear subspace and so every affine subspace is closed. This need not be true in an infinite dimensional topological vector space.

The **relative interior** of a convex set C, denoted $\operatorname{ri} C$, is taken to be its interior relative to its affine hull. Even a one point set has a nonempty relative interior in this sense. The only convex set with an empty relative interior is the empty set. Similarly, the **relative boundary** of a convex set is the boundary relative to the affine hull. This turns out to be the closure less the relative interior.

Note that this is not the same relative interior that a topologist would mean. In particular, it is *not* true that $A \subset B$ implies $\operatorname{ri} A \subset \operatorname{ri} B$. For instance, consider a closed interval and one of its endpoints. The relative interior of the interval is the open interval and the relative interior of the singleton endpoint is itself, which is disjoint from the relative interior of the interval.

An important consequence of the definition is:

> *If $x \in \operatorname{ri} C$ and $y \in C$, the line L through x and y lies in the affine hull of C, so x is in the interior of $L \cap C$ relative to L. Thus for small enough ε, we have $x + \varepsilon(x - y) \in C$.*

We may use this fact without any special mention.

I will now state without proof some useful properties of relative interiors.

16 Proposition (Rockafellar [6, Theorem 6.3, p. 46]) *For a convex subset C of $\boldsymbol{R}^{\mathrm{n}}$,*

$$\overline{\operatorname{ri} C} = \overline{C}, \qquad \text{and} \qquad \operatorname{ri}(\operatorname{ri} C) = \operatorname{ri} C.$$

A consequence of this is that the affine hull of $\operatorname{ri} C$ and of $\overline{C}$ coincide.

17 Proposition (Rockafellar [6, Theorem 6.5, p. 47]) *Let $\{C_i\}_{i \in I}$ be a family of convex subsets of $\boldsymbol{R}^{\mathrm{n}}$, and assume $\bigcap_i \operatorname{ri} C_i \neq \varnothing$. Then*

$$\overline{\bigcap_i C_i} = \bigcap_i \overline{C_i}, \qquad \text{and for finite } I, \quad \operatorname{ri} \bigcap_i C_i = \bigcap_i \operatorname{ri} C_i.$$

[6] Econometrics can be useful to theorists too.

18 Proposition (Rockafellar [6, Corollaries 6.6.1, 6.6.2, pp. 48–49]) *For convex subsets C, C_1, C_2 of $\boldsymbol{R}^{\mathrm{n}}$, and $\lambda \in \boldsymbol{R}$,*

$$\operatorname{ri}(\lambda C) = \lambda \operatorname{ri} C, \qquad \operatorname{ri}(C_1 + C_2) = \operatorname{ri} C_1 + \operatorname{ri} C_2, \qquad \textit{and} \qquad \overline{C_1 + C_2} \supset \overline{C_1} + \overline{C_2}.$$

C Cantor Intersection Theorem

19 Cantor Intersection Theorem *In a complete metric space, if a decreasing sequence of nonempty closed subsets has vanishing diameter, then the intersection of the sequence is a singleton.*

Proof: Let $\{F_n\}$ be a decreasing sequence of nonempty closed subsets of the complete metric space (X, d), and assume $\lim_{n\to\infty} \operatorname{diameter} F_n = 0$. The intersection $F = \bigcap_{n=1}^{\infty} F_n$ cannot have more that one point, for if $a, b \in F$, then $d(a, b) \leqslant \operatorname{diameter} F_n$ for each n, so $d(a, b) = 0$, which implies $a = b$.

To see that F is nonempty, for each n pick some $x_n \in F_n$. Since $d(x_n, x_m) \leqslant \operatorname{diameter} F_n$ for $m \geqslant n$, the sequence $\{x_n\}$ is Cauchy. Since X is complete there is some $x \in X$ with $x_n \to x$. But x_n belongs to F_m for $m \geqslant n$, and each F_n is closed, so x belongs to F_n for each n. ∎

References

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

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

[3] K. Fan, I. Glicksberg, and A. J. Hoffman. 1957. Systems of inequalities involving convex functions. *Proceedings of the American Mathematical Society* 8:617–622. www.jstor.org/stable/2033529

[4] J.-B. Hiriart-Urruty and C. Lemaréchal. 2001. *Fundamentals of convex analysis*. Grundlehren Text Editions. Berlin: SpringerVerlag.

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

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

[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.