Strict separation

For $x, y \in \mathbb{R}^n$, we write $d(x, y) = \|x - y\|$. For subsets $A, B \subseteq \mathbb{R}^n$, we define $d(A, x) = d(x, A) = \inf_{a \in A} d(x, a)$ and $d(A, B) = \inf_{a \in A, b \in B} d(a, b)$. Let $\text{diam} A = \sup_{x,y \in A} d(x, y)$ For $A \subseteq \mathbb{R}^n$ and $x \in \mathbb{R}^n$, we define $\langle A, x \rangle = \{\langle a, x \rangle : a \in A\}$.

In Euclidean space $\mathbb{R}^n$, the term compact set refers to any set that is closed and bounded.

**Theorem 1** (Separating hyperplane theorem, strict case). Let $C, K \subseteq \mathbb{R}^n$ be nonempty convex sets with $C \cap K = \emptyset$. If $C$ is closed and $K$ compact, then there exists $\psi \in \mathbb{R}^n$ with

$$\inf \langle C, \psi \rangle > \sup \langle K, \psi \rangle.$$

**Proof.** The strategy of the proof is illustrated in Figure 1. We start by proving the existence of a pair of closest points $x^*$ and $y^*$, where $x \in C$ and $y \in K$. We then show that the hyperplane with normal vector $\psi = x^* - y^*$ separates the two convex sets. Details follow.

**Claim 2.** There exist $x^* \in C$ and $y^* \in K$ such that $d(x^*, y^*) = d(C, K)$.

**Proof.** For this, pick an arbitrary point $x_0 \in C$ and define $r = 2d(x_0, K) + \text{diam} K$. By the triangle inequality, $d(x, x_0) \leq d(x, K) + \text{diam} K + d(x_0, K)$. It follows that any point $x$ with $d(x, x_0) > r$ satisfies

$$d(x, K) \geq d(x, x_0) - \text{diam} K - d(x_0, K)$$

$$> d(x_0, K).$$

As a result, the compact set $C' = C \cap \{x : d(x, x_0) \leq r\}$ obeys $d(C, K) = d(C', K)$. Since $d(\cdot, \cdot)$ is a continuous function on the compact $C' \times K$, it must attain its infimum on $C' \times K$, i.e., there must exist $(x^*, y^*) \in C' \times K$ with $d(x^*, y^*) = d(C', K) = d(C, K)$.

In the remainder of the proof, fix $x^*$ and $y^*$ as in Claim 2, and define $\psi = x^* - y^*$.

**Claim 3.** $\inf \langle C, \psi \rangle \geq \langle x^*, \psi \rangle$. 

1
Figure 1: Separating two convex sets by a hyperplane.

Proof. For the sake of contradiction, suppose that \( \langle x, \psi \rangle < \langle x^*, \psi \rangle \) for some \( x \in C \). This is equivalent to

\[
\langle x - x^*, x^* - y^* \rangle < 0.
\]

(1)

For \( \varepsilon \in (0, 1) \), the point \( x_\varepsilon = (1 - \varepsilon)x^* + \varepsilon x \) is contained in \( C \) by convexity. However,

\[
\|x_\varepsilon - y^*\|^2 = \langle x^* - y^* + \varepsilon(x - x^*), x^* - y^* + \varepsilon(x - x^*) \rangle
\]

\[
= \|x^* - y^*\|^2 + 2\varepsilon \langle x - x^*, x^* - y^* \rangle + \varepsilon^2 \|x - x^*\|^2.
\]

Hence \( d(x_\varepsilon, y^*) < d(x^*, y^*) \) for \( \varepsilon > 0 \) small enough, contradicting \( d(x^*, y^*) = d(C, K) \).

Claim 4. \( \langle x^*, \psi \rangle > \langle y^*, \psi \rangle \).

Proof. We have \( \langle x^* - y^*, \psi \rangle = \|x^* - y^*\|^2 > 0 \), where the last step uses the fact that \( x^* \neq y^* \) by the disjointness of \( C \) and \( K \).

Claim 5. \( \langle y^*, \psi \rangle \geq \sup \langle K, \psi \rangle \).

Proof. The proof is analogous to Claim 3. Specifically, suppose for the sake of contradiction that \( \langle y, \psi \rangle > \langle y^*, \psi \rangle \) for some \( y \in K \). This is equivalent to

\[
\langle y^* - y, x^* - y^* \rangle < 0.
\]

(2)

For \( \varepsilon \in (0, 1) \), the point \( y_\varepsilon = (1 - \varepsilon)y^* + \varepsilon y \) is contained in \( K \) by convexity. However,

\[
\|x^* - y_\varepsilon\|^2 = \langle x^* - y^* + \varepsilon(y^* - y), x^* - y^* + \varepsilon(y^* - y) \rangle
\]

\[
= \|x^* - y^*\|^2 + 2\varepsilon \langle y^* - y, x^* - y^* \rangle + \varepsilon^2 \|y^* - y\|^2.
\]

<0 by (2)
Hence $d(x^*, y_\varepsilon) < d(x^*, y^*)$ for $\varepsilon > 0$ small enough, contradicting $d(x^*, y^*) = d(C, K)$. \hfill \Box$

By Claims 3–5, the proof is complete. \hfill \Box$

**Nonstrict separation**

The proofs below use the following property of compact sets $K \subset \mathbb{R}^n$: given any sequence $x_1, x_2, \ldots, x_n, \ldots \in K$, there is a subsequence $x_{i_1}, x_{i_2}, \ldots, x_{i_n}, \ldots$ and some $x^* \in K$ such that $x_{i_n} \to x^*$ as $n \to \infty$. In other words, every sequence in a compact set has a convergent subsequence. The *closure* of a set $A \subseteq \mathbb{R}^n$ is a superset of $A$ defined by $\text{cl} A = \{x \in \mathbb{R}^n : d(x, A) = 0\}$. Put differently, $\text{cl} A$ is the smallest closed set that contains $A$. A point $x$ is called an *interior* point of $A$ if there exists $\varepsilon > 0$ such that $\{y \in \mathbb{R}^n : d(x, y) < \varepsilon\} \subseteq A$. The set of all interior points of $A$ is denoted $\text{int} A$.

**Lemma 6.** Let $M \in \mathbb{R}^{n \times (n+1)}$ be given by

$$M = \begin{bmatrix} 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 \\ \vdots & \vdots & \vdots & \iddots & 0 & -1 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}.$$

Let $\{M_k\}$ be a sequence with $M_k \to M$. Then for some $k$, there exists a vector $\lambda \in (0, \infty)^{n+1}$ with $M_k \lambda = 0$.

**Proof.** Since the nullspace of every $M_k$ is nonempty, we can fix a sequence $\{\lambda_k\}$ of unit vectors with $M_k \lambda_k = 0$. By passing to a subsequence if necessary, we may assume that $\lambda_k \to \lambda^*$. But then $\lambda^*$ is a unit vector with $M \lambda^* = 0$, which forces

$$\lambda^* = \frac{\pm 1}{\sqrt{n+1}} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

In particular, for all $k$ large enough, the components of $\lambda_k$ are either all positive or all negative, so that either $\lambda_k$ or $-\lambda_k$ is the desired vector. \hfill \Box

**Theorem 7** (Separating hyperplane theorem, nonstrict case). Let $X, Y \subseteq \mathbb{R}^n$ be nonempty convex subsets. If $X \cap Y = \emptyset$, then there exists a nonzero $\psi \in \mathbb{R}^n$ with

$$\inf \langle X, \psi \rangle \geq \sup \langle Y, \psi \rangle.$$
Proof. Consider the convex set $A = X - Y = \{x - y : x \in X, y \in Y\}$. Then $0 \notin A$, and our objective is to find a nonzero $\psi \in \mathbb{R}^n$ with $\inf \langle A, \psi \rangle \geq 0$. Let $B = \text{cl} A$ be the closure of $A$.

First of all, we claim that $0 \notin \text{int} B$. For the sake of contradiction, suppose otherwise. Then for $\varepsilon > 0$ small enough, $B$ contains the ball $\{v : \|v\|_\infty \leq \varepsilon\}$. In particular, $B$ contains $\varepsilon v_1, \varepsilon v_2, \ldots, \varepsilon v_{n+1}$, where $v_i$ is the $i$th column of the matrix $M$ in Lemma 6. Recall that each $v_i$ is the limit of a sequence in $A$. By Lemma 6, it follows that some $\tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_{n+1} \in A$ obey $\sum \lambda_i v_i = 0$ for some positive coefficients $\lambda_1, \lambda_2, \ldots, \lambda_{n+1}$. Since $A$ is convex, we conclude that $0 \in A$, a contradiction. Hence $0 \notin \text{int} B$, as claimed.

The remainder of the proof is illustrated in Figure 2. By the claim just settled, we can fix a sequence of points $\{z_k\}$ outside of $B$ with $z_k \to 0$. By the strict version of the separating hyperplane theorem, for each $k$ there exists a unit vector $\psi_k$ with

$$\inf \langle B, \psi_k \rangle > \langle z_k, \psi_k \rangle.$$  

(3)

Passing to a subsequence if necessary, we may assume that $\psi_k \to \psi$ for some unit vector $\psi$. We now claim that $\inf \langle B, \psi \rangle \geq 0$. Indeed, for every $v \in B$,

$$\langle v, \psi \rangle = \lim_{k \to \infty} \langle v, \psi_k \rangle$$  

since $\psi_k \to \psi$

$$\geq \lim_{k \to \infty} \langle z_k, \psi_k \rangle$$  

by (3)

$$= 0$$  

since $\|\psi_k\| = 1$ and $z_k \to 0$. \qed