EE3001 Machine Learning **Fall 2022**

Lecture 05. Separation Theorems

Lecturer: Jie Wang Date: Oct 11, 2022

1 Introduction

Recall the example we introduced last lecture as follows.

Example 1. Suppose $p : \mathbb{R}^n \to \mathbb{R}$ satisfies $p(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in C$ and $\int_C p(\mathbf{x})d\mathbf{x} = 1$, where $C \subseteq \mathbb{R}^n$ is convex. Then

$$\int_C p(\mathbf{x})\mathbf{x}d\mathbf{x} \in C,$$

if the integral exists.

How to show the claim in Example 1 rigorously? In this lecture, we introduce a suite of powerful tools in convex analysis, called separation theorems.

2 Projection

Consider a closed convex set $C \subseteq \mathbb{R}^n$ and a point $\mathbf{x} \in \mathbb{R}^n$. If there is a point $\mathbf{z} \in C$ that is closest to $\mathbf{x}$, we call $\mathbf{z}$ the projection of $\mathbf{x}$ on C, which is denoted by $\Pi_C(\mathbf{x}) = \mathbf{z}$. That is, the point $\mathbf{z}$ solves the optimization problem as follows

$$\inf_{\mathbf{y} \in C} \|\mathbf{x} - \mathbf{y}\|^2. \tag{1}$$

Question 1.

1. Can we always find a solution $\mathbf{z}$ to the problem in (1)?
2. Is the projection unique?

Indeed, the projection is always well defined, which is confirmed by the result as follows.

Theorem 1. *Suppose that the set $C \subseteq \mathbb{R}^n$ is nonempty, convex, and closed. Then, for every $\mathbf{x} \in \mathbb{R}^n$, there exists exactly one point $\mathbf{z} \in C$ that is closest to $\mathbf{x}$.*

Proof.
We first show the existence by the Extreme Value Theorem. We denote the objective function of the problem in (1) by

$$f(\mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|^2,$$

which is clearly a continuous function of $\mathbf{y}$. Let $\mathbf{y}_0$ be an arbitrary point in C, and

$$r = \|\mathbf{x} - \mathbf{y}_0\|.$$

Then, if we denote the intersection of the ball $B(\mathbf{x}, r)$ and C by C', we can conclude that C' is nonempty, as it at least includes $\mathbf{y}_0$. We can see that the problem

$$\inf_{\mathbf{y} \in C'} f(\mathbf{y}) \tag{2}$$

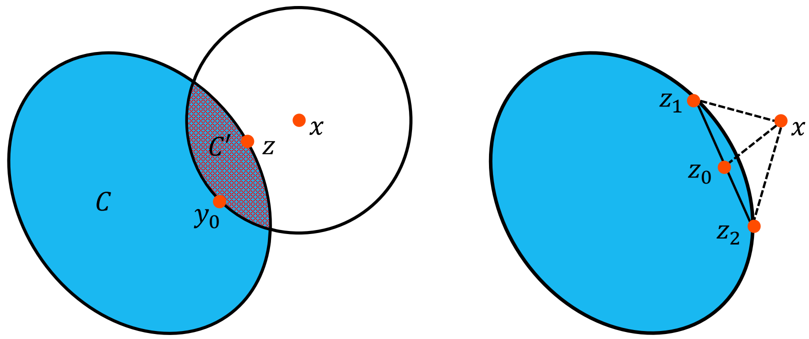

Figure 1: Illustration of the idea to prove Theorem 1. The left and right figures illustrate the idea to prove the existence and the uniqueness of the projection of a given point with respect to a nonempty closed convex set, respectively.

shares the same solution set with the problem in (1).

We next show that C' is compact. Indeed, as both the ball $B(\mathbf{x}, r)$ and C are closed, the set C' is closed as well. Moreover, the boundedness of $B(\mathbf{x}, r)$ implies that C' must be bounded. All together, we conclude that the set C' is compact.

Due to the continuity of f and the compactness of C', the Extreme Value Theorem immediately leads to the existence of $\mathbf{z}$.

We next show the uniqueness of $\mathbf{z}$. Suppose that two different point $\mathbf{z}_1$ and $\mathbf{z}_2$ solve the optimization problem in (1). Let

$$\gamma = f(\mathbf{z}_1) = f(\mathbf{z}_2) = \min_{\mathbf{y}\in C} \|\mathbf{x} - \mathbf{y}\|^2.$$

Consider the point

$$\mathbf{z}_0 = \frac{\mathbf{z}_1 + \mathbf{z}_2}{2}.$$

Then, the Pythagorean theorem leads to

$$\|\mathbf{z}_0 - \mathbf{x}\|^2 = \gamma^2 - \frac{1}{4}\|\mathbf{z}_1 - \mathbf{z}_2\|^2 < \gamma^2,$$

a contradiction. This show that $\mathbf{z}$ must be unique. □

We next give an useful result that characterizes projections.

Lemma 1. *Suppose that C is a nonempty closed convex set and let $\mathbf{x} \in \mathbb{R}^n$. Then,*

$$\mathbf{z} = \Pi_C(\mathbf{x}) \Leftrightarrow \langle \mathbf{x} - \mathbf{z}, \mathbf{y} - \mathbf{z} \rangle \leq 0, \ \forall\, \mathbf{y} \in C.$$

The above inequality is the so-called **variational inequality**.

Proof.

($\Rightarrow$) Suppose that $\mathbf{z} = \Pi_C(\mathbf{x})$. For any $\mathbf{y} \in C$, we define

$$g(t) = f(\mathbf{z} + t(\mathbf{y} - \mathbf{z})) = \|\mathbf{x} - \mathbf{z} - t(\mathbf{y} - \mathbf{z})\|^2 = \|\mathbf{x} - \mathbf{z}\|^2 - 2t\langle \mathbf{x} - \mathbf{z}, \mathbf{y} - \mathbf{z}\rangle + t^2\|\mathbf{y} - \mathbf{z}\|^2. \quad (3)$$

For any $t \in (0, 1]$, we can see that,

$$g(0) < g(t),$$

leading to

$$2\langle \mathbf{x} - \mathbf{z}, \mathbf{y} - \mathbf{z}\rangle < t\|\mathbf{y} - \mathbf{z}\|^2.$$

As the above inequality holds for any $t \in (0, 1]$, we can conclude that

$$\langle \mathbf{x} - \mathbf{z}, \mathbf{y} - \mathbf{z}\rangle \leq 0.$$

($\Leftarrow$) Suppose that

$$\langle \mathbf{x} - \mathbf{z}, \mathbf{y} - \mathbf{z}\rangle \leq 0, \forall \mathbf{y} \in C.$$

Eq. (3) implies that $g(0) < g(1)$ for any $\mathbf{y} \neq \mathbf{z}$, that is

$$f(\mathbf{z}) < f(\mathbf{y}), \forall \mathbf{y} \in C, \mathbf{y} \neq \mathbf{z}.$$

Thus, the point $\mathbf{z}$ must be the projection of $\mathbf{x}$ on C. □

Question 2. What if C is an affine set in Lemma 1?

Theorem 2 (Nonexpansiveness). *Suppose that $C \subseteq \mathbb{R}^n$ is closed and convex. Then, for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, we have*

$$\|\Pi_C(\mathbf{x}) - \Pi_C(\mathbf{y})\| \leq \|\mathbf{x} - \mathbf{y}\|.$$

Proof.
Lemma 1 leads to

$$\langle \mathbf{x} - \Pi_C(\mathbf{x}), \Pi_C(\mathbf{y}) - \Pi_C(\mathbf{x})\rangle \leq 0,$$
$$\langle \mathbf{y} - \Pi_C(\mathbf{y}), \Pi_C(\mathbf{x}) - \Pi_C(\mathbf{y})\rangle \leq 0.$$

Adding both sides we have

$$\|\Pi_C(\mathbf{x}) - \Pi_C(\mathbf{y})\|^2 + \langle \Pi_C(\mathbf{x}) - \Pi_C(\mathbf{y}), \mathbf{y} - \mathbf{x}\rangle \leq 0,$$

which implies that

$$\|\Pi_C(\mathbf{x}) - \Pi_C(\mathbf{y})\|^2 \leq \langle \Pi_C(\mathbf{x}) - \Pi_C(\mathbf{y}), \mathbf{x} - \mathbf{y}\rangle \leq \|\Pi_C(\mathbf{x}) - \Pi_C(\mathbf{y})\|\|\mathbf{x} - \mathbf{y}\|.$$

The claim follows immediately. □

***Remark* 1.** We indeed have better results than nonexpansiveness, that is, **firmly nonexpansiveness**.

3 Hyperplanes

Definition 1. [2] A hyperplane $H \subset \mathbb{R}^n$ is an $(n-1)$-dimensional affine subset of $\mathbb{R}^n$, that is,

$$H = \{\mathbf{x} \in \mathbb{R}^n : \ell(x) = \alpha\}$$

is the level set of a nontrivial linear function $\ell : \mathbb{R}^n \to \mathbb{R}$. If ℓ takes the form of

$$\ell(\mathbf{x}) = \langle \mathbf{a}, \mathbf{x} \rangle$$

for $\mathbf{a} \neq 0$ in $\mathbf{R}^n$, then

$$H = H_{(\mathbf{a},\alpha)} = \{\mathbf{x} \in \mathbb{R}^n : \langle \mathbf{a}, \mathbf{x} \rangle = \alpha\}.$$

Definition 2. Let $H = H_{(\mathbf{a},\alpha)}$ be a hyperplane in $\mathbb{R}^n$. The hyperplane H separates $\mathbb{R}^n$ into two closed half-spaces:

$$H^+_{(\mathbf{a},\alpha)} = \{\mathbf{x} \in \mathbb{R}^n : \langle \mathbf{a}, \mathbf{x} \rangle \geq \alpha\},$$
$$H^-_{(\mathbf{a},\alpha)} = \{\mathbf{x} \in \mathbb{R}^n : \langle \mathbf{a}, \mathbf{x} \rangle \leq \alpha\}.$$

We denote the corresponding open half-spaces by

$$H^{++}_{(\mathbf{a},\alpha)} = \{\mathbf{x} \in \mathbb{R}^n : \langle \mathbf{a}, \mathbf{x} \rangle > \alpha\},$$
$$H^{--}_{(\mathbf{a},\alpha)} = \{\mathbf{x} \in \mathbb{R}^n : \langle \mathbf{a}, \mathbf{x} \rangle < \alpha\}.$$

Definition 3. Let C_1 and C_2 be two nonempty sets and $H := H_{(\mathbf{a},\alpha)}$ a hyperplane in $\mathbf{R}^n$.

1. H is called a **separating hyperplane** for the sets C_1 and C_2 if they are contained in the two closed half-spaces determined by H, respectively, e.g., $C_1 \subseteq H^+_{(\mathbf{a},\alpha)}$ and $C_2 \subseteq H^-_{(\mathbf{a},\alpha)}$.
2. H is called a **strictly separating hyperplane** for the sets C_1 and C_2 if they are contained in the two open half-spaces determined by H, respectively, e.g., $C_1 \subseteq H^{++}_{(\mathbf{a},\alpha)}$ and $C_2 \subseteq H^{--}_{(\mathbf{a},\alpha)}$.
3. H is called a **strongly separating hyperplane** for the sets C_1 and C_2 if there exists β and γ with $\gamma < \alpha < \beta$, such that $C_1 \subseteq H^-_{(\mathbf{a},\gamma)}$ and $C_2 \subseteq H^+_{(\mathbf{a},\beta)}$.
4. H is called a **properly separating hyperplane** for the sets C_1 and C_2 if H separates C_1 and C_2, and C_1 and C_2 are not both contained in the hyperplane H.

If there exists a hyperplane H separating the sets C_1 and C_2 in one of the senses above, we say that C_1 and C_2 can be separated, strictly separated, strongly separated, properly separated, respectively.

Question 3. Can you find an example in which C_1 and C_2 can be strictly separated instead of being strongly separated?

4 Separation Theorems

4.1 Separation between a Point and a Convex Set

Theorem 3. *Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set, and let $\mathbf{x}_0 \notin C$. Then, the set C and the point $\mathbf{x}_0$ can be strongly separated, that is, there exists a nonzero $\mathbf{a} \in \mathbb{R}^n$ and $\alpha < \beta$ such that $C \subseteq H^-_{(\mathbf{a},\alpha)}$ and $\mathbf{x}_0 \in H^+_{(\mathbf{a},\beta)}$.*

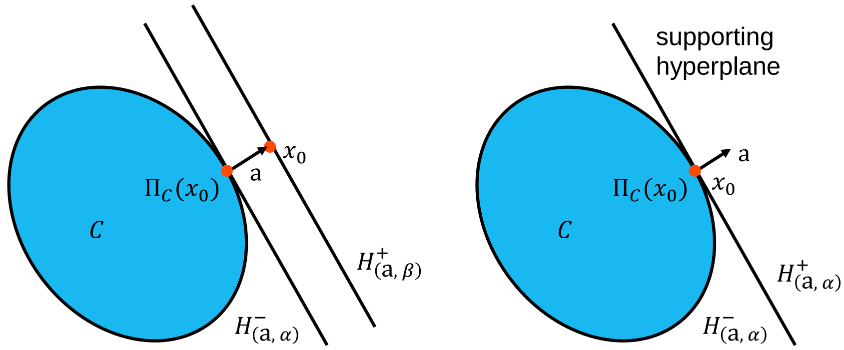

Figure 2: Separation Theorems (Theorems 3, 4 and 5).

Proof. Theorem 1 implies the existence and uniqueness of the projection of $\mathbf{x}_0$ on C. We define

$$\mathbf{a} = \mathbf{x}_0 - \Pi_C(\mathbf{x}_0).$$

As $\mathbf{x}_0 \notin C$, the vector $\mathbf{a} \neq 0$. We then define

$$\begin{aligned}\alpha =& \langle \mathbf{a}, \Pi_C(\mathbf{x}_0) \rangle, \\ \beta =& \langle \mathbf{a}, \mathbf{x}_0 \rangle.\end{aligned}$$

It is easy to see that

$$\beta = \langle \mathbf{a}, \mathbf{x}_0 - \Pi_C(\mathbf{x}_0) + \Pi_C(\mathbf{x}_0) \rangle = \|\mathbf{a}\|^2 + \alpha.$$

Thus, we have $\alpha < \beta$.

We now show that $C \subseteq H^-_{(\mathbf{a},\alpha)}$ and $\mathbf{x}_0 \in H^+_{(\mathbf{a},\beta)}$. The latter is trivial, as $\mathbf{x}_0 \in H_{(\mathbf{a},\beta)} \subset H^+_{(\mathbf{a},\beta)}$. To show the former, we note that for any $\mathbf{y} \in C$, we have

$$\langle \mathbf{a}, \mathbf{y} \rangle = \langle \mathbf{a}, \mathbf{y} - \Pi_C(\mathbf{x}_0) + \Pi_C(\mathbf{x}_0) \rangle = \langle \mathbf{a}, \mathbf{y} - \Pi_C(\mathbf{x}_0) \rangle + \alpha.$$

By Lemma 1, we have

$$\langle \mathbf{a}, \mathbf{y} - \Pi_C(\mathbf{x}_0) \rangle \leq 0.$$

Combining the above two inequalities, we have

$$\langle \mathbf{a}, \mathbf{y} \rangle \leq \alpha, \, \forall \, \mathbf{y} \in C,$$

which is equivalent to $C \subseteq H^-_{(\mathbf{a},\alpha)}$. The proof is complete. □

Theorem 3 leads to an important characterization of closed convex sets, which is stated as follows.

Corollary 1. [1] *The closure of the convex hull of a set C is the intersection of the closed half-spaces that contain C. In particular, a closed convex set is the intersection of the closed half-spaces that contain it.*

Proof. Let S be the intersection of all closed half-spaces that contain C. As every closed half-space containing C must also contain $\mathbf{cl}\,(\mathbf{conv}\ C)$ (**why?**), we must have $\mathbf{cl}\,(\mathbf{conv}\ C) \subset S$.

To show the reverse direction, we note that, for any $\mathbf{x} \notin \mathbf{cl}\,(\mathbf{conv}\ C)$—by Theorem 3—we can find a hyperplane H strongly separating $\mathbf{x}$ and $\mathbf{cl}\,(\mathbf{conv}\ C)$. Thus, the corresponding closed half-space induced by H that contains $\mathbf{cl}\,(\mathbf{conv}\ C)$ does not contain $\mathbf{x}$, so $\mathbf{x} \notin S$. This shows that $\mathbf{cl}\,(\mathbf{conv}\ C) \supset S$. □

If the set C in Theorem 3 is not closed, the set C and the point $\mathbf{x}_0 \notin C$ may not be strongly separated, as $\mathbf{x}_0$ can be a **boundary point** of C.

Definition 4. Let $C \subseteq \mathbb{R}^n$ be a nonempty set, and $\mathbf{x}_0$ a point in its boundary $\mathbf{bd}\ C$, i.e.,

$$\mathbf{x}_0 \in \mathbf{bd}\ C = \mathbf{cl}\ C \setminus \mathbf{int}\ C.$$

A hyperplane $H := H_{(\mathbf{a},\alpha)}$ is called a **supporting hyperplane** to C at the point $\mathbf{x}_0$ if $\mathbf{x}_0 \in H_{(\mathbf{a},\alpha)}$ and $C \subseteq H^-_{(\mathbf{a},\alpha)}$, that is,

$$\langle \mathbf{a}, \mathbf{x} \rangle \leq \langle \mathbf{a}, \mathbf{x}_0 \rangle = \alpha,\ \forall\, \mathbf{x} \in C.$$

Theorem 4. *Let $C \subseteq \mathbb{R}^n$ be a nonempty convex set, and let $\mathbf{x}_0 \notin C$. Then, the set C and the point $\mathbf{x}_0$ can be separated, that is, there exists a hyperplane $H_{(\mathbf{a},\alpha)}$ such that $C \subseteq H^-_{(\mathbf{a},\alpha)}$ and $\mathbf{x}_0 \in H^+_{(\mathbf{a},\alpha)}$.*

Proof. If $\mathbf{x}_0 \notin \mathbf{cl}\ C$, Theorem 3 implies that the set $\mathbf{cl}\ C$ and $\mathbf{x}_0$ can be strongly separated, and so can be the set C and $\mathbf{x}_0$. Thus, the set C and $\mathbf{x}_0$ can be separated.

We next consider the case where $\mathbf{x}_0 \in \mathbf{cl}\ C$ (to simplify notations, let $\bar{C} = \mathbf{cl}\ C$). If this is the case, we can find a sequence $(\mathbf{x}_k)$ with $\mathbf{x}_k \notin \mathbf{cl}\ C$, $k = 1, 2, \ldots$, and $\mathbf{x}_k \to \mathbf{x}_0$ (why?). Let

$$\mathbf{a}_k = \frac{\mathbf{x}_k - \Pi_{\bar{C}}(\mathbf{x}_k)}{\|\mathbf{x}_k - \Pi_{\bar{C}}(\mathbf{x}_k)\|}.$$

Theorem 3 implies that

$$\langle \mathbf{a}_k, \mathbf{y} \rangle \leq \langle \mathbf{a}_k, \mathbf{x}_k \rangle,\ \forall\, \mathbf{y} \in C.$$

As $\|\mathbf{a}_k\| = 1$ for all $k = 1, 2, \ldots,$, there exists a converging subsequence. Without loss of generality, we assume that $\mathbf{a}_k \to \mathbf{a}$. Passing to the limit on both sides of the above inequality, we have

$$\langle \mathbf{a}, \mathbf{y} \rangle \leq \langle \mathbf{a}, \mathbf{x}_0 \rangle,\ \forall\, \mathbf{y} \in C,$$

which completes the proof ($\alpha = \langle \mathbf{a}, \mathbf{x}_0 \rangle$). □

Remark *2.* To show Theorem 4, we construct convergent sequence so that we can apply the result in 3. Indeed, different convergent sequences may lead to different separating hyperplanes. Some of them are useful, that is, they can help us to distinguish different sets, while some of them are not.

Remark *3.* To show Theorem 4, we consider a special case, in which $\mathbf{x}_0 \in \mathbf{cl}\ C$. If this is the case, by Theorem 4, we can find a hyperplane such that it passes through $\mathbf{x}_0$ and separates $\mathbf{x}_0$ from C. Geometrically, this hyperplane *just touches* C and it is said to be supporting C at $\mathbf{x}_0$.

Given a nonempty set C and a point $\mathbf{x}_0 \in \mathbf{bd}\ C$, the hyperplane supporting C at $\mathbf{x}_0$ may not even exist, e.g., $C = [0,1] \cap \mathbb{Q}$ and $\mathbf{x}_0 = 0.5$. The next result shows that, if C is a nonempty convex set and $\mathbf{x}_0 \in \mathbf{bd}\ C$, we can always find a supporting hyperplane to C at $\mathbf{x}_0$.

Theorem 5 (Supporting Hyperplane Theorem). *Let $C \subseteq \mathbb{R}^n$ be a nonempty convex set, and $\mathbf{x}_0$ a point in its boundary* $\mathbf{bd}\ C$*. Then, there exists a hyperplane supporting C at $\mathbf{x}_0$.*

We omit the proof of Themore 5 as the argument is similar to that of 4.

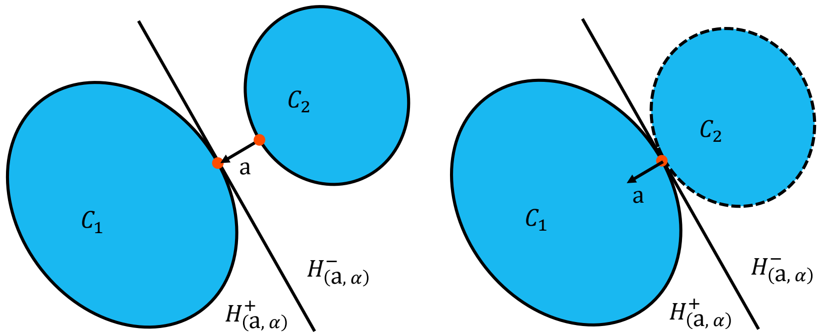

Figure 3: Separating Hyperplane Theorem.

4.2 Separation between Convex Sets

Theorem 6 (**Separating Hyperplane Theorem**). *Let C_1 and C_2 be two nonempty convex sets in $\mathbf{R}^n$. If C_1 and C_2 are disjoint, i.e., $C_1 \cap C_2 = \emptyset$, there exists a hyperplane that separates them.*

Proof. Consider the convex (why?) set:

$$C = C_1 - C_2 = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \mathbf{x}_1 - \mathbf{x}_2, \, \mathbf{x}_1 \in C_1, \mathbf{x}_2 \in C_2\}.$$

As $C_1 \cap C_2 = \emptyset$, we have $\mathbf{0} \notin C$. Then, by Theorem 4, the set C and $\mathbf{0}$ can be separated, i.e., there exists a vector $\mathbf{a} \neq 0$ such that

$$\langle \mathbf{a}, \mathbf{x} \rangle \geq 0, \, \forall \, \mathbf{x} \in C,$$

which is equivalent to

$$\langle \mathbf{a}, \mathbf{x}_1 \rangle \geq \langle \mathbf{a}, \mathbf{x}_2 \rangle, \, \mathbf{x}_1 \in C_1, \, \mathbf{x}_2 \in C_2.$$

If we let $\alpha = \inf_{\mathbf{x}_1 \in C_1} \langle \mathbf{a}, \mathbf{x}_1 \rangle$, the above inequality implies that $C_1 \subseteq H^+_{(\mathbf{a},\alpha)}$ and $C_2 \subseteq H^-_{(\mathbf{a},\alpha)}$. This completes the proof. □

Remark 4. One of the reasons why the separating theorems are important is that we want a (simple) method to distinguish sets from one another. Given two sets, if they can be separated by a hyperplane—that is, they are contained in the two closed half-spaces determined by the hyperplane—we would like the linear function associated with the hyperplane takes different values on points in these two sets. However, in some cases, even we can separate two convex sets in the sense of the first part in Definition 2, the linear function associated with the hyperplane may take the same value on the two sets, that is, we can not distinguish the two sets by the linear function. For example, consider the unit disk $C_1 = \{\mathbf{x} \in \mathbb{R}^3 : \|\mathbf{x}\| \leq 1, x_3 = 0\}$ and the x-axis $C_2 = \{\mathbf{x} \in \mathbf{R}^3 : x_2 = x_3 = 0\}$ in $\mathbb{R}^3$. These two sets are both convex and can be separated by the x, y plane. However, the corresponding linear function $\ell(\mathbf{x}) = \langle (0, 0, 1), \mathbf{x} \rangle$ takes the same value on both sets. Notice that, the aforementioned two sets are overlapping, as $C_1 \cap C_2 \neq \emptyset$.

Thus, we introduce the **proper separation theorem**, which turns out to be useful in some important optimization scenarios and is more consistent with our intuition on the geometrical meaning of separation.

We first introduce a useful lemma.

Lemma 2. *Let C be a nonempty convex set and a hyperplane H that contains C in one of its closed half-spaces in $\mathbb{R}^n$. Then,*

$$C \subset H \Leftrightarrow \textbf{relint } C \cap H \neq \emptyset.$$

Proof. Suppose that $C \subset H$. Then, we must have $\textbf{relint } C \subset H$ as $\textbf{relint } C \subseteq C$, and thus $\textbf{relint } C \cap H \neq \emptyset$.

Suppose that $\textbf{relint } C \cap H \neq \emptyset$. Let $\mathbf{x}_0 \in \textbf{relint } C \cap H$ and $H = H_{(\mathbf{a},\alpha)}$ with $\mathbf{a} \neq 0$. Without loss of generality, we assume that $C \subseteq H^{+}_{(\mathbf{a},\alpha)}$, i.e.,

$$\langle \mathbf{a}, \mathbf{x} \rangle \geq \alpha = \langle \mathbf{a}, \mathbf{x}_0 \rangle, \, \forall \, \mathbf{x} \in C,$$

which is equivalent to

$$\langle \mathbf{a}, \mathbf{x} - \mathbf{x}_0 \rangle \geq 0, \, \forall \, \mathbf{x} \in C.$$

As $\mathbf{x}_0$ is a relative interior of C, for any $\mathbf{x} \in C$ and $\mathbf{x} \neq \mathbf{x}_0$, we can find a small positive number ϵ such that (why?)

$$\mathbf{x}_\tau = \mathbf{x_0} - \tau(\mathbf{x} - \mathbf{x}_0) \in C, \, \forall \, \tau \in [0, \epsilon).$$

Thus, for any $\tau \in [0, \epsilon)$, we have

$$\langle \mathbf{a}, \mathbf{x}_\tau - \mathbf{x}_0 \rangle \geq 0 \Rightarrow -\tau \langle \mathbf{a}, \mathbf{x} - \mathbf{x}_0 \rangle \geq 0 \Rightarrow \langle \mathbf{a}, \mathbf{x} - \mathbf{x}_0 \rangle = 0 \Rightarrow \langle \mathbf{a}, \mathbf{x} \rangle = \langle \mathbf{a}, \mathbf{x}_0 \rangle = \alpha,$$

which implies that $C \subseteq H$, as $\mathbf{x}$ is an arbitrary point in C. □

Theorem 7 (Proper Separation Theorem). [1] *Let $C \subseteq \mathbb{R}^n$ be a nonempty convex set, and $\mathbf{x}_0 \in \mathbb{R}^n$ be a vector. There exists a hyperplane that properly separates C and $\mathbf{x}_0$ if and only if $\mathbf{x}_0 \notin \textbf{relint } C$.*

Proof. Suppose that there exists a hyperplane $H_{(\mathbf{a},\alpha)}$ that properly separates C and $\mathbf{x}_0$. We have two possibilities.

1. The point $\mathbf{x}_0 \notin H$. Without loss of generality, we assume that $\mathbf{x}_0 \in H^{++}_{(\mathbf{a},\alpha)}$ and $C \subseteq H^{-}_{(\mathbf{a},\alpha)}$. Then, we must have $\mathbf{x}_0 \notin C$ and thus $\mathbf{x}_0 \notin \textbf{relint } C$.
2. The set $C \not\subseteq H$. Due to Lemma 2, we have $\textbf{relint } C \cap H = \emptyset$ and thus $\mathbf{x}_0 \notin \textbf{relint } C$.

Conversely, suppose that $\mathbf{x}_0 \notin \textbf{relint } C$. To show the existence of the hyperplane that properly separates C and $\mathbf{x}_0$, we consider two cases as follows.

1. The point $\mathbf{x}_0 \notin \textbf{aff } C$. As $\textbf{aff } C$ is closed and convex, Theorem 3 implies that $\mathbf{x}_0$ and $\textbf{aff } C$ can be strongly separated, and thus $\mathbf{x}_0$ and C can be properly separated.
2. The point $\mathbf{x}_0 \in \textbf{aff } C$. We have two possibilities.
   (a) The point $\mathbf{x}_0 \notin \textbf{cl } C$. Again, by Theorem 3, there exists a hyperplane that can strongly separate $\mathbf{x}_0$ and C and thus also properly separate $\mathbf{x}_0$ and C.

(b) The point $\mathbf{x}_0 \in \mathbf{cl}\ C$. As $\mathbf{x}_0 \notin \mathbf{relint}\ C$, the point $\mathbf{x}_0$ must be a relative boundary point of C. We consider two cases as follows.

i. $\text{rank}(\mathbf{aff}\ C) = n$. Then, the point $\mathbf{x}_0$ is indeed a boundary point of C. By the Supporting Hyperplane Theorem, there exists a hyperplane $H_{(\mathbf{a},\alpha)}$ supporting C at $\mathbf{x}_0$, i.e.,

$$\langle \mathbf{a}, \mathbf{x} \rangle \leq \langle \mathbf{a}, \mathbf{x}_0 \rangle = \alpha, \ \forall \mathbf{x} \in C.$$

We can see that $\mathbf{int}\ C \cap H = \emptyset$ (why?). Then, Lemma 2 implies that $C \not\subseteq H$, that is, $H_{(\mathbf{a},\alpha)}$ properly separates C and $\mathbf{x}_0$.

ii. $\text{rank}(\mathbf{aff}\ C) < n$. Let S be the subspace that is parallel to $\mathbf{aff}\ C$, and consider the set $\hat{C} = C + S^{\perp}$. Clearly, $\text{rank}(\mathbf{aff}\ \hat{C}) = n$, and $\mathbf{x}_0 \in \mathbf{bd}\ \hat{C}$. By a similar argument with the last part, we can find a hyperplane H that properly separates $\mathbf{x}_0$ and $\hat{C}$, and thus properly separates $\mathbf{x}_0$ and C as well.

All together, the proof is complete. □

Theorem 8 (**Proper Separation of Two Convex Sets**). [1] *Let C_1 and C_2 be two nonempty convex subsets of $\mathbf{R}^n$. There exists a hyperplane that properly separates C_1 and C_2 if and only if*

$$\mathbf{relint}\ C_1 \cap \mathbf{relint}\ C_2 = \emptyset.$$

References

[1] D. Bertsekas. *Convex Optimization Theory*. Athena Scientific, 2009.

[2] O. Güler. *Foundations of optimization*. Springer, 2010.