From Convex Analysis to Learning, Prediction, and Elicitation*
Lecture 1: Hyperplane Separation Theorems

Lunjia Hu

A central result in convex analysis is the hyperplane separation theorem. The theorem is geometrically very intuitive, yet it has broad and powerful applications in various areas of computer science, statistics, and economics. It states that every pair of disjoint convex sets can be separated by a hyperplane:

Theorem 1 (Hyperplane separation theorem)**.** *Let $C_1, C_2 \subseteq \mathbb{R}^d$ be two disjoint convex sets. There exists a non-zero vector $h \in \mathbb{R}^d$ such that*

$$\langle x_1, h\rangle \ge \langle x_2, h\rangle \quad \text{for every } x_1 \in C_1, x_2 \in C_2. \tag{1}$$

We will discuss its proof shortly. What makes this theorem really powerful is its *constructive* nature. It shows:

If something does not exist, then something else must exist.

In the context of Theorem 1, "something" is a point x that belongs to both C_1 and C_2 (which does not exist), and "something else" is the separating hyperplane h. We will see many more examples of similar nature later in the course.

1 Conditions for Strict Separation

An important question about this theorem is the conditions for strict separation: under what conditions can we make the non-strict inequality "$\ge$" in (1) become strict "$>$"? The answer to this question will be very useful for our future discussions about *minimax theorems* and *Lagrange duality*, especially for understanding *Slater's condition*.

Here are two fairly illustrative examples where strict separation fails.

1. $C_1 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \le 1\}, C_2 = \{(x, y) \in \mathbb{R}^2 : x = 1, y > 0\}$;
2. $C_1 = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 \le 1\}, C_2 = \{(x, y, z) \in \mathbb{R}^3 : x = 1, y \ge 0, yz \ge 1\}$.

In the following theorem, we summarize three sufficient conditions for strict separation:

Theorem 2 (Strict hyperplane separation)**.** *Let $C_1, C_2 \subseteq \mathbb{R}^d$ be two disjoint convex sets. Assume at least one of the following conditions is satisfied:*

1. either C_1 or C_2 is open;

*https://lunjiahu.com/convex-analysis/

2. C_1 is closed and C_2 is compact;

3. C_1, C_2 can both be written as the intersection of a convex polyhedron[1] and a convex open set (this is crucial for understanding Slater's condition in Lagrange duality, which we will discuss later).

Then there exists a vector $h \in \mathbb{R}^d$ such that

$$\langle x_1, h\rangle > \langle x_2, h\rangle \quad \text{for every } x_1 \in C_1, x_2 \in C_2. \tag{2}$$

2 Proving Hyperplane Separation Theorems

We need the following lemma to prove Theorem 1. We use $\mathbf{0}$ to denote the zero vector (a.k.a. the origin).

Lemma 3. *Let $C \subseteq \mathbb{R}^d$ be an open convex set. Assume $\mathbf{0} \notin C$. Then there exists $h \in \mathbb{R}^d$ such that*

$$\langle x, h\rangle > 0 \quad \text{for every } x \in C.$$

We present two proofs of the lemma. Both proofs use induction.

Proof 1. We prove this lemma by induction on d. We first need to prove the lemma for the special cases $d = 1, 2$. We leave this part as an exercise for interested readers. We focus on the $d \geq 3$ case from now on.

Define $X' \subseteq \mathbb{R}^d$ to be the two-dimensional linear subspace consisting of vectors with all but the first two coordinates being zero: $X' := \{(x_1, x_2, 0, \ldots, 0) \in \mathbb{R}^d : x_1, x_2 \in \mathbb{R}\}$. Now $C \cap X'$ is a (possibly empty) open convex subset of the two-dimensional subspace X'. By the induction hypothesis for $d = 2$, there exists $h_0 \in X'$ such that $\langle x, h_0\rangle > 0$ for every $x \in C \cap X'$. We assume without loss of generality that $h_0 = (0, 1, 0, 0, \ldots, 0)$.

For every $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$, define $\mathsf{proj}(x) := (x_2, x_3, \ldots, x_d) \in \mathbb{R}^{d-1}$. We show that $\mathsf{proj}(x) \neq \mathbf{0}$ for every $x \in C$. We prove this by contradiction. If $\mathsf{proj}(x) = \mathbf{0}$ for some $x \in C$, then x must belong to $C \cap X'$. However, this implies that $x_2 = \langle x, h_0\rangle > 0$, contradicting the assumption that $\mathsf{proj}(x) = \mathbf{0}$.

We have now shown that $C' := \{\mathsf{proj}(x) : x \in C\}$ is a subset of $\mathbb{R}^{d-1}$ that does not contain the origin $\mathbf{0}$. Since C is convex and open, it is clear that C' is also convex and open. By the induction hypothesis for $d-1$, there exists $h_1 \in \mathbb{R}^{d-1}$ such that $\langle \mathsf{proj}(x), h_1\rangle > 0$ for every $x \in C$. Let $h \in \mathbb{R}^d$ be the vector whose first coordinate is zero, and the remaining $d-1$ coordinates are the coordinates of h_1. For every $x \in C$, we have $\langle x, h\rangle = \langle \mathsf{proj}(x), h_1\rangle > 0$, as desired. □

Proof 2. We prove this lemma by induction on d. This inductive idea comes from the proof of the Hahn-Banach Theorem, which is a fundamental result in functional analysis.[2]

We first need to prove the lemma for the special cases $d = 1, 2$. We leave this part as an exercise for interested readers. We focus on the $d \geq 3$ case from now on and assume without loss of generality that C is non-empty.

[1] A convex polyhedron is the intersection of finitely many closed halfspaces. A closed halfspace is a set of the form $\{x \in \mathbb{R}^d : \langle x, h\rangle \leq b\}$ for some $h \in \mathbb{R}^d$ and $b \in \mathbb{R}$.

[2] See `https://www.math.ksu.edu/~nagy/real-an/ap-e-h-b.pdf`.

Let $S \subseteq \mathbb{R}^d$ be an arbitrary linear subspace with dimension $d-1$ such that $C \cap S \neq \emptyset$. By the inductive hypothesis, there exists a linear function $f_S : S \to \mathbb{R}$ such that

$$f_S(x) > 0 \quad \text{for every } x \in C \cap S. \tag{3}$$

Since $C \cap S \neq \emptyset$, we know that f_S is not the constant zero function.

Pick $v \in \mathbb{R}^d \setminus S$ arbitrarily. Every $x \in \mathbb{R}^d$ can be uniquely decomposed as $x = x_S + \alpha_x v$, where $x_S \in S$ and $\alpha_x \in \mathbb{R}$. We construct the following convex set $C' \subseteq \mathbb{R}^2$:

$$C' := \{(f_S(x_S), \alpha_x) : x \in C\}. \tag{4}$$

The fact that f_S is not the constant zero function ensures that C' is open. By the inductive hypothesis (3), we have $(0,0) \notin C'$. Therefore, there exists a linear function $g : \mathbb{R}^2 \to \mathbb{R}$ such that

$$g(x') > 0 \quad \text{for every } x' \in C'. \tag{5}$$

Define $f : \mathbb{R}^d \to \mathbb{R}$ such that $f(x) := g((f_S(x_S), \alpha_x))$ for every $x \in \mathbb{R}^d$. Now f is a linear function, and by (4) and (5), it satisifies $f(x) > 0$ for every $x \in C$, completing the proof. □

We are now ready to prove Theorem 1. We say a set $A \subseteq \mathbb{R}^d$ is *affine* if it can be written as $A = S + v$ for a linear subspace $S \subseteq \mathbb{R}^d$ and an offset $v \in \mathbb{R}^d$. The *dimension* of A is defined as the dimension of S. The *affine hull* of a set $C \subseteq \mathbb{R}^d$ is the smallest affine set containing C.

Proof of Theorem 1. Define $C := C_1 - C_2$. That is,

$$C = \{x_1 - x_2 : x_1 \in C_1, x_2 \in C_2\}.$$

By the assumption that C_1 and C_2 are disjoint, we have $\mathbf{0} \notin C$. Our hope is to apply Lemma 3 to C, but this cannot be done directly because C is not necessarily open. We thus consider the following two cases:

Case 1: the affine hull of C has dimension d. In this case, the interior of C (denoted by $\mathsf{int}\, C$) is a non-empty open convex set which does not contain 0. We can thus apply Lemma 3 to $\mathsf{int}\, C$ to get $h \in \mathbb{R}^d$ such that $\langle x, h\rangle > 0$ for every $x \in \mathsf{int}\, C$. Moreover, every $x \in C$ is the limit of a convergent sequence of points in $\mathsf{int}\, C$, so we have $\langle x, h\rangle \geq 0$ for every $x \in C$. This proves (1).

Case 2: the affine hull of C (denoted by A) has dimension below d. If A does not contain the origin, then the theorem trivially holds. If A contains the origin, then it is a subspace of $\mathbb{R}^d$ with dimension $d' < d$. Focusing on this subspace (instead of the entire $\mathbb{R}^d$) reduces to Case 1 (with d replaced by d'). □

2.1 Strict Separation from Closedness and Compactness

Theorem 4. *Let $C_1, C_2 \subseteq \mathbb{R}^d$ be two disjoint non-empty convex sets. Assume that C_1 is closed and C_2 is compact. Then there exists $h \in \mathbb{R}^d$ such that*

$$\min_{x_1 \in X_1} \langle x_1, h\rangle > \max_{x_2 \in X_2} \langle x_2, h\rangle,$$

where both the min and the max can be attained.

Proof. We claim that there exist $x_1^* \in X_1$ and $x_2^* \in X_2$ such that

$$0 < \|x_1^* - x_2^*\|_2 \le \|x_1 - x_2\|_2 \quad \text{for every } x_1 \in X_1 \text{ and } x_2 \in X_2.$$

This claim holds by a basic mathematical analysis argument. We omit the proof here.

Define $\ell := \|x_1^* - x_2^*\|_2 > 0$. We enlarge C_2 to C_2' as follows:

$$C_2' := C_2 + B(0, \ell) = \{x_2 + z : x_2 \in C_2, \|z\|_2 < \ell\}.$$

It is clear that C_2' is a convex set disjoint from C_1. By Theorem 1, there exists $h \in \mathbb{R}^d$ with $\|h\|_2 = 1$ such that

$$\langle x_1, h\rangle \ge \langle x_2 + z, h\rangle \quad \text{for every } x_1 \in C_1, x_2 \in C_2 \text{ and } z \in \mathbb{R}^d \text{ with } \|z\|_2 < \ell.$$

Choosing $z \to \ell h$, we have $\langle z, h\rangle \to \ell$, so

$$\langle x_1, h\rangle \ge \langle x_2, h\rangle + \ell \quad \text{for every } x_1 \in C_1, x_2 \in C_2.$$

However, since $\|x_1^* - x_2^*\|_2 = \ell$, we have

$$\langle x_1^*, h\rangle \le \langle x_2^*, h\rangle + \ell.$$

Combining the two inequalities above, we have

$$\langle x_1^*, h\rangle = \min_{x_1 \in C_1} \langle x_1, h\rangle,$$
$$\langle x_2^*, h\rangle = \max_{x_2 \in C_2} \langle x_2, h\rangle,$$

and

$$\min_{x_1 \in C_1} \langle x_1, h\rangle = \max_{x_2 \in C_2} \langle x_2, h\rangle + \ell > \max_{x_2 \in C_2} \langle x_2, h\rangle.$$

□

2.2 Strict Separation from Intersections of Convex Polyhedra and Convex Open sets

Theorem 5 (Strict separation)**.** *Let $O_1, O_2 \subseteq \mathbb{R}^d$ be convex open sets and let $P_1, P_2 \subseteq \mathbb{R}^d$ be convex polyhedra. Define $C_1 = O_1 \cap P_1$ and $C_2 = O_2 \cap P_2$. Assume C_1 and C_2 are disjoint. Then there exists a vector $h \in \mathbb{R}^d$ such that*

$$\langle x_1 - x_2, h\rangle > 0 \quad \textit{for every } x_1 \in C_1, x_2 \in C_2.$$

We need the following helper lemmas:

Lemma 6. *Let $C \subseteq \mathbb{R}^d$ be a non-empty convex set. Assume $\mathbf{0} \notin C$. Then there exists a vector $h \in \mathbb{R}^d$ such that*

$$\langle x, h\rangle \ge 0 \quad \textit{for every } x \in C, \textit{ and} \tag{6}$$
$$\langle x, h\rangle > 0 \quad \textit{for some } x \in C. \tag{7}$$

Proof. Similarly to the proof of Theorem 1, we can assume without loss of generality that the affine hull of C has dimension d. Now by Theorem 1, there exists $h \in \mathbb{R}^d$ such that (6) holds. We show that (7) must also hold. If not, then $\langle x, h\rangle = 0$ for every $x \in C$. This means that C is contained in a $(d-1)$-dimensional subspace, contradicting with the assumption that the affine hull of C has dimension d. $\square$

Lemma 7. *Let $O \subseteq \mathbb{R}^d$ be an open convex set. Let $p_1, \ldots, p_m, q_1, \ldots, q_n \in \mathbb{R}^d$ and $a_1, \ldots, a_m$, $b_1, \ldots, b_n \in \mathbb{R}$ be vectors and real numbers. Suppose there does* not *exist $o \in O$ that satisfy all of the following inequalities and equalities:*

$$\langle o, p_1\rangle \le a_1, \ldots, \langle o, p_m\rangle \le a_m, \langle o, q_1\rangle = b_1, \ldots, \langle o, q_n\rangle = b_n.$$

Then there exist $g_1, \ldots, g_m \ge 0$ and $h_1, \ldots, h_n \in \mathbb{R}$ such that

$$\sum_{i=1}^m g_i(\langle o, p_i\rangle - a_i) + \sum_{j=1}^n h_j(\langle o, q_j\rangle - b_j) > 0 \quad \textit{for every } o \in O.$$

Proof. We assume without loss of generality that O is non-empty. Let C be the set of vectors $(x_1, \ldots, x_m, y_1, \ldots, y_n) \in \mathbb{R}^{m+n}$ such that there exists $o \in O$ satisfying

$$\langle o, p_1\rangle - a_1 \le x_1, \ldots, \langle o, p_m\rangle - a_m \le x_m, \langle o, q_1\rangle - b_1 = y_1, \ldots, \langle o, q_n\rangle - b_n = y_n.$$

Clearly, C is a non-empty convex set. The assumptions of the lemma ensure that $\mathbf{0} \notin C$. By Lemma 6, there exist $g_1, \ldots, g_m, h_1, \ldots, h_n \in \mathbb{R}$ satisfying

$$\sum_{i=1}^m g_i x_i + \sum_{j=1}^n h_j y_j \ge 0 \quad \text{for every } (x_1, \ldots, x_m, y_1, \ldots, y_n) \in C, \tag{8}$$

$$\sum_{i=1}^m g_i x_i + \sum_{j=1}^n h_j y_j > 0 \quad \text{for some } (x_1, \ldots, x_m, y_1, \ldots, y_n) \in C. \tag{9}$$

It is clear that $g_1, \ldots, g_m \ge 0$: if some g_i is negative, we can always increase x_i so that (8) is violated. For every $o \in O$, we define

$$f(o) := \sum_{i=1}^m g_i(\langle o, p_i\rangle - a_i) + \sum_{j=1}^n h_j(\langle o, q_j\rangle - b_j).$$

We prove the lemma by induction on m. We start from the base case $m = 0$. Inequality (8) implies that $f(o) \ge 0$ for every $o \in O$. We show that the inequality is strict:

$$f(o) > 0 \quad \text{for every } o \in O. \tag{10}$$

Suppose (10) does not hold. Then f is an affine function that is nonnegative on an open set O with $f(o) = 0$ for some $o \in O$.[3] This is only possible when f is the constant zero function, violating (9).

Now we consider the general case $m > 0$. Again, by (8) we have $f(o) \ge 0$ for every $o \in O$, and the proof is completed if the strict inequality (10) holds. We thus focus on the case where (10) does

[3] We say a function $f : \mathbb{R}^d \to \mathbb{R}$ is affine if there exists a linear function $f_0 : \mathbb{R}^d \to \mathbb{R}$ and an offset $r \in \mathbb{R}$ such that $f(x) = f_0(x) + r$ for every $x \in \mathbb{R}^d$.

not hold. As we argued earlier, f must be the constant zero function on O. If $g_1 = \cdots = g_m = 0$, then we get the same contradiction with (9) as the $m = 0$ case. Therefore, some g_i must be positive. Assume without loss of generality that $g_1 > 0$.

Note that the *inequality* $\langle o, p_1 \rangle \le a_1$ is a necessary condition for the *equality* $\langle o, p_1 \rangle = a_1$. Thus by the inductive hypothesis, there exist $g'_1 \in \mathbb{R}, g'_2, \ldots, g'_m \ge 0$, and $h'_1, \ldots, h'_n \in \mathbb{R}$ such that

$$\sum_{i=1}^{m} g'_i(\langle o, p_i \rangle - a_i) + \sum_{j=1}^{n} h'_j(\langle o, q_j \rangle - b_j) > 0 \quad \text{for every } o \in O. \tag{11}$$

We have shown $f(o) = 0$ for every $o \in O$, so

$$\sum_{i=1}^{m} g_i(\langle o, p_i \rangle - a_i) + \sum_{j=1}^{n} h_j(\langle o, q_j \rangle - b_j) = 0 \quad \text{for every } o \in O. \tag{12}$$

For a constant $\alpha \ge 0$, defining $g''_i = g'_i + \alpha g_i$ and $h''_j = h'_j + \alpha h_j$, by (11) and (12) we have

$$\sum_{i=1}^{m} g''_i(\langle o, p_i \rangle - a_i) + \sum_{j=1}^{n} h''_j(\langle o, q_j \rangle - b_j) > 0 \quad \text{for every } o \in O.$$

Since $g_1 > 0$, when α is sufficiently large, we have $g''_1 \ge 0$. Moreover, we always have $g''_i = g'_i + \alpha g_i \ge 0$ for $i = 2, \ldots, m$. This completes the proof. □

Proof of Theorem 5. Since P_1 is a convex polyhedron, it can be written as $P_1 = \{x \in \mathbb{R}^d : K_1 x \le a_1\}$ for a matrix $K_1 \in \mathbb{R}^{m_1 \times d}$ and a vector $a_1 \in \mathbb{R}^{m_1}$. Here, for two vectors $v, v' \in \mathbb{R}^{m_1}$, we say $v \le v'$ if every coordinate of v does not exceed the corresponding coordinate of v'. Similarly, P_2 can be written as $P_2 = \{x \in \mathbb{R}^d : K_2 x \le a_2\}$ for a matrix $K_2 \in \mathbb{R}^{m_2 \times d}$ and a vector $a_2 \in \mathbb{R}^{m_2}$.

The assumption that C_1 and C_2 are disjoint means that there does *not* exist $(o_1, o_2) \in O_1 \times O_2$ satisfying

$$K_1 o_1 \le a_1, K_2 o_2 \le a_2, o_1 - o_2 = \mathbf{0}.$$

By Lemma 7, there exist $g_1 \in \mathbb{R}^{m_1}_{\ge 0}, g_2 \in \mathbb{R}^{m_2}_{\ge 0}, h \in \mathbb{R}^d$ such that

$$f(o_1, o_2) := \langle K_1 o_1 - a_1, g_1 \rangle + \langle K_2 o_2 - a_2, g_2 \rangle + \langle o_1 - o_2, h \rangle > 0 \quad \text{for every } (o_1, o_2) \in O_1 \times O_2.$$

Now for every $x_1 \in C_1$ and $x_2 \in C_2$, we have $(x_1, x_2) \in O_1 \times O_2$, $K_1 x_1 - a_1 \le \mathbf{0}, K_2 x_2 - a_2 \le \mathbf{0}$. Therefore,

$$0 < f(x_1, x_2) \le \langle x_1 - x_2, h \rangle,$$

completing the proof. □