From Convex Analysis to Learning, Prediction, and Elicitation*
Lecture 2: Minimax Theorems

Lunjia Hu

Consider the following basic game between two players Alice and Bob. In this game, Alice chooses an action a from a set A, and Bob chooses an action b from another set B. The result of the game is determined by a function $f : A \times B \to \mathbb{R}$: Alice gets $f(a, b)$ points whereas Bob gets $-f(a, b)$ points. Note that $f(a, b)$ could be positive, zero, or negative, depending on the choices of both players' actions a and b. Since the points of the two players sum to zero, such games are termed *zero-sum games*.

Both players hope to maximize their points. Alice's goal is to choose an action $a \in A$ that maximizes $f(a, b)$, whereas Bob's goal is to choose $b \in B$ that minimizes $f(a, b)$.

A special case of a zero-sum game is the "rock paper scissors" game. The fairness of the game depends crucially on the rule that *both players should play synchronously*. Indeed, if they play sequentially, the second player has an advantage of *knowing the first players's action*, whereas the first player has to choose their action *without knowing the second player's action*. This advantage can be quantified as follows.

Assume both players play optimally. When Alice plays first and Bob plays second, Alice will get the following points:[1]

$$\max_{a \in A} \min_{b \in B} f(a, b).$$

When Bob plays first and Alice plays second, Alice will get the following points:

$$\min_{b \in B} \max_{a \in A} f(a, b).$$

The advantage of playing second is

$$\min_{b \in B} \max_{a \in A} f(a, b) - \max_{a \in A} \min_{b \in B} f(a, b) \geq 0.$$

Roughly speaking, the minimax theorem states that under certain convexity and regularity conditions on A, B and f, playing second has no advantage:

$$\max_{a \in A} \min_{b \in B} f(a, b) = \min_{b \in B} \max_{a \in A} f(a, b).$$

The minimax theorem is a central result in game theory. It has the following interpretation that is very useful in various applications in computer science:

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

[1] We assume that the max and min can be attained in this introductory part of the notes. We will not make such implicit assumptions in the technical part later, where we use sup and inf instead.

If Alice can ensure $f(a, b) \geq t$ when she plays second, then she can ensure $f(a, b) \geq t$ even when she plays first.

This is useful in computer science because one can think of Alice's action as an *algorithm* and Bob's action as an *input*. We view $f(a, b)$ as the performance of algorithm a on input b. Usually, we'd like to design an algorithm a that performs well on *every* input b. That is, we want to ensure $f(a, b) \geq t$ for some threshold t when Alice plays the algorithm a *first*. To achieve this goal, if the minimax theorem holds, we just need to ensure $f(a, b) \geq t$ when Alice plays second: for every *fixed* input b, we aim to find an algorithm a with good performance. This often becomes much easier than directly finding a single, universal algorithm a that performs well for *every* input b.

We will prove the minimax theorem using the hyperplane separation theorem. While the connection between the two theorems may appear elusive at first, it is not surprising given that they are both instances of the following statement:

If something doesn't exist, then something else must exist.

Indeed, the minimax theorem can be viewed as the following statement:

If Bob cannot ensure $f(a, b) < t$ when he plays first, then Alice can ensure $f(a, b) \geq t$ when she plays first.

1 Dual Cone and Polar Cone

Definition 1. *We say $K \subseteq \mathbb{R}^d$ is a cone if for every $x \in K$ and every $\lambda > 0$, it holds that $\lambda x \in K$.*

Definition 2 (Dual set and polar set)**.** *For any $S \subseteq \mathbb{R}^d$, its dual S^* is defined to be*

$$S^* := \{x^* \in \mathbb{R}^d : \langle x, x^* \rangle \geq 0 \text{ for every } x \in S\}.$$

The polar S° of S is defined to be

$$S^\circ := -(S^*) = \{x^* \in \mathbb{R}^d : \langle x, x^* \rangle \leq 0 \text{ for every } x \in S\}.$$

Lemma 1. *The dual and polar of every set $S \subseteq \mathbb{R}^d$ are non-empty closed convex cones.*

Theorem 2. *Let K be a non-empty closed convex cone. Then $(K^*)^* = (K^\circ)^\circ = K$.*

Proof. We first show that $K \subseteq (K^*)^*$. Consider any $x \in K$. By the definition of K^*, every $x^* \in K^*$ satisfies $\langle x, x^* \rangle \geq 0$. This means that $x \in (K^*)^*$.

Now we show that $(K^*)^* \subseteq K$. Consider any $y \in \mathbb{R}^d \setminus K$. Our goal is to show $y \notin (K^*)^*$. Since K is a closed convex set, by the hyperplane separation theorem, there exists $h \in \mathbb{R}^d$ such that

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

Since K is a non-empty closed cone (which must contain the origin), the above inequality implies

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

This implies that $h \in K^*$ and $y \notin (K^*)^*$, as desired.

We have now proved $(K^*)^* = K$. It remains to prove $(K^\circ)^\circ = (K^*)^*$. This holds because $K^\circ = -(K^*)$ and $(K^\circ)^\circ = -(-(K^*))^* = (K^*)^*$. $\square$

2 Minimax Theorem for Inner Products

When $f(a, b)$ can be expressed as the inner product between Alice's action a and Bob's action b, we have the following fundamental minimax theorem:

Theorem 3. *Let $A, B \subseteq \mathbb{R}^d$ be convex sets. Assume that A is non-empty and compact. Then*

$$\sup_{a \in A} \inf_{b \in B} \langle a, b \rangle = \inf_{b \in B} \sup_{a \in A} \langle a, b \rangle.$$

Before we prove the theorem, we first remark that one direction of the theorem is trivial:

Lemma 4. *Let A, B be arbitrary sets and let $f : A \times B \to \mathbb{R}$ be an arbitrary function. Then*

$$\sup_{a \in A} \inf_{b \in B} f(a, b) \le \inf_{b \in B} \sup_{a \in A} f(a, b).$$

Proof. It suffices to show that for every $a_0 \in A$ and $b_0 \in B$,

$$\inf_{b \in B} f(a_0, b) \le \sup_{a \in A} f(a, b_0).$$

This is true because $\inf_{b \in B} f(a_0, b) \le f(a_0, b_0) \le \sup_{a \in A} f(a, b_0)$. $\square$

Proof of Theorem 3. By Lemma 4, it suffices to prove the reverse inequality

$$\sup_{a \in A} \inf_{b \in B} \langle a, b \rangle \ge \inf_{b \in B} \sup_{a \in A} \langle a, b \rangle.$$

Define $t := \inf_{b \in B} \sup_{a \in A} \langle a, b \rangle \in \mathbb{R} \cup \{\pm\infty\}$. For every $b \in B$, we have $\sup_{a \in A} \langle a, b \rangle \ge t$. This can be interpreted as "Bob cannot ensure $\langle a, b \rangle < t$ when he plays first". Therefore, for every $b \in B$ and every $t' < t$, there exists $a \in A$ such that

$$\langle a, b \rangle > t'.$$

Define $K := \{\lambda(a, -1) : \lambda \ge 0, a \in A\} \subseteq \mathbb{R}^{d+1}$ and $S := \{(b, t') : b \in B, t' < t\} \subseteq \mathbb{R}^{t+1}$. Our assumption that A is a non-empty compact convex set ensures that K is a non-empty closed convex cone.

Now for every $s \in S$, there exists $k \in K$ such that

$$\langle k, s \rangle > 0.$$

This means that $S \cap K^\circ = \emptyset$. By the hyperplane separation theorem, there exists a non-zero vector $h \in \mathbb{R}^{t+1}$ such that

$$\langle h, s \rangle \ge \langle h, k^\circ \rangle \quad \text{for every } s \in S \text{ and } k^\circ \in K^\circ.$$

By Lemma 1, K° is a non-empty closed convex cone (which must contain the origin), so the above inequality implies

$$\langle h, s \rangle \ge 0 \quad \text{for every } s \in S,$$
$$\langle h, k^\circ \rangle \le 0 \quad \text{for every } k^\circ \in K^\circ.$$

The second inequality implies $h \in (K^\circ)^\circ$, and by Theorem 2 we have $h \in K$. By the definition of K and the fact that $h \ne \mathbf{0}$, there exists $a \in A$ and $\lambda > 0$ such that $h = \lambda(a, -1)$. Plugging this into the first inequality above and using the definition of S, we have

$$\langle a, b \rangle \ge t \quad \text{for every } b \in B.$$

This proves that $\sup_{a \in A} \inf_{b \in B} \langle a, b \rangle \ge t$, as desired. $\square$

Remark 1. *The compactness assumption on A is necessary. Consider $A, B \subseteq \mathbb{R}^2$ such that $A = \{(x, 1)\}, B = \{(1, y)\}$. Then*

$$\sup_{a \in A} \inf_{b \in B} \langle a, b \rangle = -\infty, \quad \textit{but}$$
$$\inf_{b \in B} \sup_{a \in A} \langle a, b \rangle = +\infty.$$

There is also an example where both quantities are finite but different. Suppose A consists of points $a = (x_1, x_2, 0, x_2 + 1)$ for $x_1, x_2 \in \mathbb{R}$, and B consists of points $b = (y_1, y_2, y_3, y_4 - 1)$, where $\begin{bmatrix} y_1 & y_2 \\ y_2 & y_3 \end{bmatrix} \succeq 0$ and $y_4 \geq 0$. We have

$$\langle a, b \rangle = x_1 y_1 + x_2 y_2 + (x_2 + 1)(y_4 - 1).$$

Whenever $x_2 \neq 0$, we have

$$\inf_{b \in B} \langle a, b \rangle = -\infty.$$

Therefore,

$$\sup_{a \in A} \inf_{b \in B} \langle a, b \rangle = \sup_{a \in A : x_2 = 0} \inf_{b \in B} \langle a, b \rangle = -1.$$

Similarly, if $y_1 \neq 0$ or $y_2 + y_4 \neq 1$, we have

$$\sup_{a \in A} \langle a, b \rangle = +\infty.$$

Moreover, $y_1 = 0$ implies $y_2 = 0$. Therefore,

$$\inf_{b \in B} \sup_{a \in A} \langle a, b \rangle = \inf_{b \in B : y_1 = y_2 = 0, y_4 = 1} \sup_{a \in A} \langle a, b \rangle = 0.$$

3 Minimax Theorem for Two-Player Zero-Sum Games

Theorem 3 has many specific requirements on the action sets A, B and the payoff function f. The main requirement is that $f(a, b)$ is the inner product $\langle a, b \rangle$ between the actions, and this requires the actions a, b to be vectors of the same dimension. It also requires the action sets A and B to be convex.

We now prove a minimax theorem without these requirements. The side effect, however, is that we need to allow Alice and Bob to play *randomized* actions (termed *mixed strategies* in game theory).

Definition 3. *Let S be a finite set. We use Δ_S to denote the set of all probability distributions on S.*

Theorem 5. *Let A, B be non-empty finite sets and let $f : A \times B \to \mathbb{R}$ be an arbitrary function. Then*

$$\sup_{x \in \Delta_A} \inf_{y \in \Delta_B} \mathbb{E}_{a \sim x, b \sim y}[f(a, b)] = \inf_{y \in \Delta_B} \sup_{x \in \Delta_A} \mathbb{E}_{a \sim x, b \sim y}[f(a, b)], \tag{1}$$

where $\mathbb{E}_{a \sim x, b \sim y}[\cdot]$ takes the expectation over independently drawn $a \in A$ from distribution x and $b \in B$ from distribution y.

Before we prove the theorem, let us introduce some useful notations. For $a \in A, b \in B, x \in \Delta_A, y \in \Delta_B$, we define

$$\begin{aligned} f(x,b) &:= \mathbb{E}_{a'\sim x}[f(a',b)], \\ f(a,y) &:= \mathbb{E}_{b'\sim y}[f(a,b')], \\ f(x,y) &:= \mathbb{E}_{a'\sim x, b'\sim y}[f(a',b')]. \end{aligned}$$

Using this notation, the conclusion (1) of Theorem 5 can be equivalently written as

$$\sup_{x\in\Delta_A} \inf_{y\in\Delta_B} f(x,y) = \inf_{y\in\Delta_B} \sup_{x\in\Delta_A} f(x,y).$$

Moreover, it is easy to see that for every fixed $x \in \Delta_A$,

$$\inf_{y\in\Delta_B} f(x,y) = \min_{b\in B} f(x,b).$$

Similarly, for every fixed $y \in \Delta_B$,

$$\sup_{x\in\Delta_A} f(x,y) = \max_{a\in A} f(a,y).$$

We thus have following corollary of Theorem 5:

Corollary 6. *Let A, B be non-empty finite sets and let $f : A \times B \to \mathbb{R}$ be an arbitrary function. Then*[2]

$$\sup_{x\in\Delta_A} \min_{b\in B} f(x,b) = \sup_{x\in\Delta_A} \inf_{y\in\Delta_B} f(x,y) = \inf_{y\in\Delta_B} \sup_{x\in\Delta_A} f(x,y) = \inf_{y\in\Delta_B} \max_{a\in A} f(a,y).$$

Proof of Theorem 5. Assume without loss of generality that $A = \{1,\dots,m\}$ and $B = \{1,\dots,n\}$. Define matrix $F \in \mathbb{R}^{m\times n}$ where $F_{ij} = f(i,j)$ for every $i \in A$ and $j \in B$. Every $x \in \Delta_A$ can be viewed as a vector $x = (x_1,\dots,x_m) \in \mathbb{R}^m$ where each x_i is the probability mass $\Pr_{a\sim x}[a = i]$ on $i \in A$. Similarly, every $y \in \Delta_B$ can be viewed as a vector $y \in \mathbb{R}^n$. This allows us to view Δ_A as a subset of $\mathbb{R}^m$ and Δ_B as a subset of $\mathbb{R}^n$. Now we have

$$\mathbb{E}_{a\sim x, b\sim y}[f(a,b)] = x^{\mathsf{T}} F y = \langle x, Fy\rangle.$$

Our goal (1) becomes

$$\sup_{x\in\Delta_A} \inf_{y\in\Delta_B} \langle x, Fy\rangle = \inf_{y\in\Delta_B} \sup_{x\in\Delta_A} \langle x, Fy\rangle.$$

This follows from Theorem 3 immediately because both Δ_A and $\{Fy : y \in \Delta_B\}$ are non-empty compact convex subsets of $\mathbb{R}^m$. □

[2] One can in fact replace every "sup" with "max" and replace every "inf" with "min" in this corollary. All the min's and max's are attainable. We leave the proof of this fact to interested readers.

4 Minimax Theorem for Concave-Convex Functions

We prove a generalization of Theorem 3 to concave-convex functions in Theorem 7 below. We need the following two definitions to state Theorem 7:

Definition 4. *Let $C \subseteq \mathbb{R}^d$ be a convex set and let $f : C \to \mathbb{R}$ be a convex function. We say f is* closed *if its epigraph E_f is closed, where*

$$E_f := \{(x,t) \in C \times \mathbb{R} : f(x) \le t\}.$$

Correspondingly, we say a concave function $g : C \to \mathbb{R}$ is closed *if $-g$ is closed.*

Definition 5. *Let $A \in \mathbb{R}^m, B \subseteq \mathbb{R}^n$ be non-empty convex sets. We say $f : A \times B \to \mathbb{R}$ is a concave-convex function if*

1. *for every fixed $b \in B$, $f(a,b)$ is a concave function of $a \in A$, and*
2. *for every fixed $a \in A$, $f(a,b)$ is a convex function of $b \in B$.*

Theorem 7. *Let $A \in \mathbb{R}^m, B \subseteq \mathbb{R}^n$ be non-empty convex sets and let $f : A \times B \to \mathbb{R}$ be a concave-convex function. Assume that A is bounded, and assume that for every $b \in B$, $f(a,b)$ is a* closed *concave function of $a \in A$. Then*

$$\sup_{a \in A} \inf_{b \in B} f(a,b) = \inf_{b \in B} \sup_{a \in A} f(a,b).$$

We need the following basic lemma to prove Theorem 7. We leave the proof of the lemma to interested readers.

Lemma 8. *Let $C \subseteq \mathbb{R}^d$ be a convex set and let $f : C \to \mathbb{R}$ be a closed convex function. Then for every $t \in \mathbb{R}$, the following set is closed:*

$$\{x \in C : f(x) \le t\}.$$

Similarly, let $g : C \to \mathbb{R}$ be a closed concave function. Then for every $t \in \mathbb{R}$, the following set is closed:

$$\{x \in C : f(x) \ge t\}.$$

We also need the famous Jensen's inequality:

Definition 6. *Let S be a finite set, and let $f : S \to \mathbb{R}$ be a function. For every distribution $x \in \Delta_S$, we define $f(x) := \mathbb{E}_{s \sim x} f(s)$.*

Theorem 9 (Jensen's inequality)**.** *Let $A \subseteq \mathbb{R}^d$ be a convex set and let $f : A \to \mathbb{R}$ be a convex function. For every finite subset $S \subseteq A$ and every distribution $x \in \Delta_S$, letting $\mu_x \in A$ denote the mean $\mathbb{E}_{a \sim x}[a]$, we have*

$$f(x) \ge f(\mu_x).$$

Similarly, if f is concave (rather than convex), we have the reverse inequality

$$f(x) \le f(\mu_x).$$

Proof of Theorem 7. By Lemma 4, it suffices to prove the reverse inequality $\sup_{a\in A}\inf_{b\in B} f(a,b) \geq \inf_{b\in B}\sup_{a\in A} f(a,b)$. To prove this, it suffices to prove that for every $t_1, t_2 \in \mathbb{R}$ satisfying

$$t_1 > \sup_{a\in A}\inf_{b\in B} f(a,b), \quad \text{and} \tag{2}$$

$$t_2 < \inf_{b\in B}\sup_{a\in A} f(a,b), \tag{3}$$

it holds that

$$t_1 \geq t_2. \tag{4}$$

By (2), for every $a \in A$, there exists $b \in B$ such that $f(a,b) < t_1$. For every $b \in B$, define $S_b := \mathbb{R}^m \setminus \{a \in A : f(a,b) \geq t_1\}$. We have $\bigcup_{b\in B} S_b = \mathbb{R}^m$. By Lemma 8, each S_b is open. Since A is bounded, we can find a bounded closed ball $Q \subseteq \mathbb{R}^m$ satisfying $A \subseteq Q$. Since Q is compact, there exists a finite subset $B' \subseteq B$ such that $\bigcup_{b\in B'} S_b \supseteq Q \supseteq A$.[3] That is, for every $a \in A$, there exists $b \in B'$ such that $f(a,b) < t_1$, implying that

$$\sup_{a\in A}\min_{b\in B'} f(a,b) \leq t_1. \tag{5}$$

By (3),

$$t_2 < \inf_{b\in B}\sup_{a\in A} f(a,b) \leq \inf_{y\in\Delta_{B'}}\sup_{a\in A} f(a,y),$$

where the last inequality follows from Theorem 9 and the fact that $f(a,b)$ is a convex function of b for every fixed a. Therefore, for every $y \in \Delta_{B'}$, there exists $a \in A$ such that $f(a,y) > t_2$. As in the proof of Theorem 5, we can view $\Delta_{B'}$ as a non-empty compact convex subset of $\mathbb{R}^{|B'|}$. Now for every fixed $a \in A$, $f(a,y)$ is an affine function of $y \in \Delta_{B'}$, so in particular it is a closed convex function of $y \in \Delta_{B'}$. Similarly to our construction of B' satisfying (5), we can now construct a finite subset $A' \subseteq A$ satisfying

$$\inf_{y\in\Delta_{B'}}\max_{a\in A'} f(a,y) \geq t_2. \tag{6}$$

Our goal (4) is achieved by the following chain of inequalities:

$$\begin{aligned}
t_2 &\leq \inf_{y\in\Delta_{B'}}\max_{a\in A'} f(a,y) && \text{(by (6))}\\
&= \sup_{x\in\Delta_{A'}}\min_{b\in B'} f(x,b) && \text{(by Corollary 6)}\\
&\leq \sup_{a\in A}\min_{b\in B'} f(a,b) && \text{(by Theorem 9 and } f(a,b) \text{ being a concave function of } a \text{ for every fixed } b)\\
&\leq t_1. && \text{(by (5))}
\end{aligned}$$

□

[3] It is a standard result in mathematical analysis that every open cover of a compact set has a finite sub-cover.