DSC 211 Introduction to Optimization Winter 2024 Instructor: Jun-Kun Wang
Scribe: Merlin Chang, Marialena Sfyraki Feb 27, 2024
Editors/TAs: Merlin Chang, Marialena Sfyraki

Lecture 15: Min-Max Optimization

1 Introduction

This lecture and next lecture concern solving

$$\inf_{x \in X} \sup_{y \in Y} g(x, y).$$

In general, we want to use inf and sup rather than min and max because the minimum or maximum might not exist. Every time we know for sure the $\inf, \sup$ are attained, we can substitute them with $\min, \max$.

1.1 Applications

There are a few interpretations of the min-max optimization problems.

1.1.1 Two Player Game

One immediate intuitive understanding of this problem would be in the form of a two-player game, in which y wants to maximize $g(\cdot, \cdot)$, while x would like to minimize $g(\cdot, \cdot)$.

1.1.2 Empirical Risk Minimization

Recall that Empirical risk minimization is the following,

$$\min_x \sum_{i=1}^{n} f_i(x)$$

One way we can more robustly formulate ERM is to consider it as a 2 player zero-sum game. Our goal is try to minimize our worst loss. Which is

$$\min_x \max_{i \in [n]} f_i(x) \iff \min_x \max_{y \in \Delta_n} y_i f_i(x)$$

This is an example of "distributionally robust optimization" problems.

1.1.3 Bi-linear Game

Let's consider the following setting:

$$\min_{x \in \Delta_n} \max_{y \in \Delta_m} y^\top A x$$

where $\Delta_n := \{x \in \mathbb{R}^n : \sum_{i=1}^n x[i] = 1, x[i] \geq 0\}$. We can see that

$$\begin{aligned} y^\top A x &= \sum_{i,j} y[i]x[j]A_{i,j} \\ &= \sum_{i,j} Pr(i_y = i)Pr(j_x = j)A_{i_y,j_x} \\ &= \mathbb{E}_{i_y \sim P_y, j_x \sim P_x}[A_{i_y,j_x}] \end{aligned}$$

Since x is trying to minimize and y is trying to maximize, we can view the above optimization $y^\top Ax$ is the loss for x and reward for y when they choose their action from their respective distributions.
Furthermore, if we let A be the mis-classification matrix as

$$A_{i,j} = \begin{cases} 1 & , \text{ if } h_i(z_j) \neq l_j \\ 0 & , otherwise \end{cases}$$

where we let z be a point in the training set $\{z_j \in \mathbb{R}^d, l_j = \{+1, -1\}\}_{j=1}^m$ and $H := \{h_i(\cdot)\}_{i=1}^n$ be a set of prediction functions, i.e., $h_i(\cdot) : \mathbb{R}^d \to \{+1, -1\}$, the min max problem becomes

$$\min_{x \in \Delta_n} \max_{y \in \Delta_m} y^\top Ax \equiv \min_{x \in \Delta_n} \max_{y \in \Delta_m} \sum_{i,j} y[i]x[j]1[h_i(z_j) \neq l_j].$$

In this case, x is trying to minimize loss, which can interpreted as searching for the best prediction function. y is trying to maximize reward, which is same as finding the hardest point to to predict adversarially.

2 Theory

We need to introduce a few notions about what does it mean to solve a min-max problem.

Definition 1 (Saddle points/ Nash Equilibrium)**.** *Let $x \in X$ and $y \in Y$ and $g(\cdot,\cdot) : X \times Y \to \mathbb{R}$. A pair of points $(x_*, y_*) \in X \times Y$ is a saddle point of $g(\cdot,\cdot)$ if*

$$g(x_*, y) \leq g(x_*, y_*) \leq g(x, y_*)$$

for any $x \in X$ and $y \in Y$.

Remark 1: When x and y are at equilibrium (x_*, y_*), they have no incentive to change. Given x_*, if y switches from y_* to any point $y \in \mathcal{Y}$, the value of $g(\cdot,\cdot)$ will not be greater than that of $g(x_*, y_*)$. Similarly, given y_*, if x switches from x_* to any point $x \in \mathcal{X}$, the value of $g(\cdot,\cdot)$ will not be smaller than that of $g(x_*, y_*)$.
Remark 2: The goal of solving a min-max optimization problem is to find the saddle point.
The following lemma is always true.

Lemma 1. *Let $g(\cdot,\cdot) : X \times Y \to \mathbb{R}$, where X and Y are not empty. Then,*

$$\inf_{x \in X} \sup_{y \in Y} g(x, y) \geq \sup_{y \in Y} \inf_{x \in X} g(x, y).$$

Proof. By the definition of infimum, we know that for any $x' \in X$

$$g(x', y) \geq \inf_{x \in X} g(x, y)$$

This also implies that

$$\sup_{y \in Y} g(x', y) \geq \sup_{y \in Y} \inf_{x \in X} g(x, y)$$

for any $x' \in X$, thus we know that the x that achieves the infimum also satisfy the above inequality, thus

$$\inf_{x \in X} \sup_{y \in Y} g(x, y) \geq \sup_{y \in Y} \inf_{x \in X} g(x, y).$$

□

Theorem 1. *Let $g(x, y) : X \times Y \to \mathbb{R}$, where X and Y are not empty. A point (x_*, y_*) is a saddle point if and only if*

- *The supremum in $\sup_{y \in Y} \inf_{x \in X} g(x, y)$ is attained at y_* & the infimum in $\inf_{x \in X} \sup_{y \in Y} g(x, y)$ is attained at x_*.*

- *Also, $\sup_{y \in Y} \inf_{x \in X} g(x, y) = \inf_{x \in X} \sup_{y \in Y} g(x, y)$, lemma 1 becomes equality.*

Proof. We will show both directions.

- ($\Longrightarrow$): Let (x_*, y_*) be a saddle point for g.
  We know by definition of the saddle point

$$g(x_*, y) \leq g(x_*, y_*) \leq g(x, y_*)$$

for any $x \in X$ and $y \in Y$, we have

$$\inf_{x\in X}\sup_{y\in Y} g(x,y) \le \sup_{y\in Y} g(x_*,y) = g(x_*,y_*) = \inf_{x\in X} g(x,y_*) \le \sup_{y\in Y}\inf_{x\in X} g(x,y) \quad (1)$$

Thus

$$\inf_{x\in X}\sup_{y\in Y} g(x,y) \le \sup_{y\in Y}\inf_{x\in X} g(x,y).$$

By lemma 1, we know that the above inequality is strict equality, where

$$\inf_{x\in X}\sup_{y\in Y} g(x,y) = \sup_{y\in Y}\inf_{x\in X} g(x,y).$$

This equality also implies that the following in (1) are true,

$$\inf_{x\in X}\sup_{y\in Y} g(x,y) = \sup_{y\in Y} g(x_*,y), \inf_{x\in X} g(x,y_*) = \sup_{y\in Y}\inf_{x\in X} g(x,y)$$

which implies that infimum and supremum for $\inf_{x\in X}\sup_{y\in Y} g(x,y)$ and $\sup_{y\in Y}\inf_{x\in X} g(x,y)$ are attained at x_* and y_* respectively. This shows that if (x_*,y_*) is a saddle point, then the two items are satisfied.

- ($\Longleftarrow$): Assume that we have the above two items in the theorem satisfied. We know by definition of sup and inf that

$$\inf_{x\in X} g(x,y_*) \le g(x_*,y_*) \le \sup_{y\in Y} g(x_*,y) \quad (2)$$

By the assumed conditions, we have

$$\inf_{x\in X} g(x,y_*) = \sup_{y\in Y}\inf_{x\in X} g(x,y) = \inf_{x\in X}\sup_{y\in Y} g(x,y) = \sup_{y\in Y} g(x_*,y). \quad (3)$$

Hence, by (2) and (3), we have

$$\inf_{x\in X} g(x,y_*) = g(x_*,y_*) = \sup_{y\in Y} g(x_*,y)$$

which implies that if $g(x_*,y_*)$ is the supremum of g in terms of y, then

$$g(x_*,y_*) \ge g(x_*,y), \forall y \in Y$$

The same argument applies for x,

$$g(x_*,y_*) \le g(x,y_*), \forall x \in X$$

Thus, we achieved the condition for (x_*,y_*) to be a saddle point, where

$$g(x_*,y) \le g(x_*,y_*) \le g(x,y_*), \forall x \in X, \forall y \in Y$$

$\square$

There are a few important implications of the above theorem.

Remark. If the inf sup problem and sup inf problem have different values, then there is no saddle point.

Remark. If a saddle point exists, then there might be multiple ones, and all of them must have the same minimax value, i.e., $\sup_{y\in Y}\inf_{x\in X} g(x,y) = \inf_{x\in X}\sup_{y\in Y} g(x,y)$

Remark. The theorem also implies that, if saddle points exists, the set of saddle points is the Cartesian product of $X^* \times Y^*$ when nonempty. The set X^*, Y^* are the set of solutions for the infimum and supremum of the following problems,

$$\inf_{\boldsymbol{x}\in\boldsymbol{X}} \sup_{y\in Y} g(x,y), \quad \sup_{\boldsymbol{y}\in\boldsymbol{Y}} \inf_{x\in X} g(x,y)$$

This implies that the way we choose x_*, y_* is completely independent from each other.[1]

Let's look at a few examples to illustrate the theorem.

Example 1. Let $g(x,y) = (x-y)^2, X = [-1,1], Y = [-1,1]$.

From the x's perspective,

$$\inf_{x\in X}\sup_{y\in Y}(x-y)^2 = \inf_{x\in X}(1+|x|)^2 = 1$$

and from the y's perspective,

$$\sup_{y\in Y}\inf_{x\in X}(x-y)^2 = \sup_{y\in Y} 0 = 0$$

thus, there is no saddle point.

Example 2. Let $g(x,y) = xy, X = (0,1], Y = (0,1]$.

In this case, although the min-max optimization from both player's perspective is the same,

$$\inf_{x\in X}\sup_{y\in Y} xy = \inf_{x\in X} x = 0 = \sup_{y\in Y}\inf_{x\in X} xy = \sup_{y\in Y} 0$$

but in condition $\inf_{x\in X} x = 0$, the infimum is not attained within X, thus saddle point still does not exist.

Theorem 1 tells us the general direction of finding a saddle point is the approach from both directions. Namely, we need to find a minimizer of x in $\sup_{y\in Y} g(x,y)$ and a maximizer of y in $\inf_{x\in X} g(x,y)$.

There are a few immediate ideas to measure progress to wards the saddle point, but they are have issues.

- Difference: $g(x, y) - g(x_*, y_*)$.
  This value can be negative or equal to zero for an infinite number of points that are not saddle points.

- Norm distance: $\|x - x_*\|_2^2 + \|y - y_*\|_2^2$.
  This quantity can go to zero at an arbitrarily slow rate

Let's consider splitting the problem into two according to theorem 1, where from the x's perspective, let's try to solve

$$\inf_{x \in X} \ell(x)$$

where $\ell(x) := \sup_{y \in Y} g(x, y)$. In this view, we can use the standard measure of suboptimality gap, which is

$$\ell(x) - \inf_{x \in X} \ell(x) = \sup_{y \in Y} g(x, y) - \inf_{x \in X} \sup_{y \in Y} g(x, y)$$

The same idea can be applied to the y's perspective and arrive at a similar measure of convergence when we only focus on the variable y,

$$\sup_{y \in Y} h(y) - h(y) = \sup_{y \in Y} \inf_{x \in X} g(x, y) - \inf_{x \in X} g(x, y)$$

where $h(y) := \inf_{x \in X} g(x, y)$. Finally, we can simply combine the above two measures to form a joint measure for saddle point progress,

$$\sup_{y \in Y} g(x, y) - \inf_{x \in X} \sup_{y \in Y} g(x, y) + \sup_{y \in Y} \inf_{x \in X} g(x, y) - \inf_{x \in X} g(x, y)$$

if we assume the existence of a saddle point, then the above measure reduces to

$$\sup_{y \in Y} g(x, y) - \inf_{x \in X} g(x, y)$$

since $\inf_{x \in X} \sup_{y \in Y} g(x, y) = \sup_{y \in Y} \inf_{x \in X} g(x, y)$ when saddle points exists.

Definition 2 (Duality Gap). *For a function $g : X \times Y \to \mathbb{R}$, the duality gap on a pair of points $(x', y') \in X \times Y$ is defined as*

$$\sup_{y \in Y} g(x', y) - \inf_{x \in X} g(x, y')$$

Remark. The duality gap is always non-negative even if saddle points don't exist. As $\sup_{y \in Y} g(x', y) \geq \inf_{x \in X} g(x, y') \forall x' \in X, \forall y' \in Y$. We will prove this property in the next lecture.

Bibliographic notes

The materials of this lecture is based on Chapter 11 of [1]. Please refer to Chapter 11 of [1] for more details about saddle points and its related theory.

References

[1] Francesco Orabona A Modern Introduction to Online Learning arXiv:1912.13213. 2023