17-740	Spring 2025
Algorithmic Foundations of Interactive Learning	

Lecture 6: Minimax via No Regret I

Lecturer: Steven Wu
Scribe: Benji Li, Steven Man

6.1 2-Player Zero-Sum Game

We will start with the most basic two-player normal-form zero-sum game. It has the following elements:

- 2 Players: **Row**, **Col**
- Actions: R, C
- Payoff Matrix: $\mathbf{M} \in \mathbb{R}^{|R| \cdot |C|}$
  - $\mathbf{M}_{ij}$ = Amount of money **Row** wins from **Col**, if **Row** plays $i \in R$ and **Col** plays $j \in C$
- Who goes first?
  1. **Row** goes first, plays i:
     - **Col** plays "best response": $\arg\min_j \mathbf{M}_{ij}$
     - **Row** plays: $\max_i(\min_j \mathbf{M}_{ij})$
  2. **Col** goes first, plays j:
     - **Row** plays "best response": $\arg\min_i \mathbf{M}_{ij}$
     - **Col** plays: $\max_i(\min_j \mathbf{M}_{ij})$

A classical example is the infamous "Rock-Paper-Scissors," which can be described by a 3×3-payoff matrix below:

$$\mathbf{M} = \left(\begin{array}{c|ccc} & R & P & S \\ \hline R & 0 & -1 & +1 \\ P & +1 & 0 & -1 \\ S & -1 & +1 & 0 \end{array}\right)$$

Now consider a thought experiment where one of the two players has to commit to playing some action first, and then the other player can choose their action accordingly. In this case, there is a clear advantage to play second. In the Rock-Paper-Scissor game, this can be written as:

$$\max_i \min_j \mathbf{M}_{ij} = -1$$

$$\min_j \max_i \mathbf{M}_{ij} = +1$$

To read the expressions above, you can go from left to right. For example, $\max_i \min_j$ means the row player chooses to play a row indexed by i first, and then the column play gets to choose a column j later. If both players are optimizing their objective, the resulting payoff (received by the row player) is then $\max_i \min_j M_{ij}$.

One could show that everyone wants to go second in any zero-sum game.

Theorem 1 (Everybody Wants to Go Second)

$$\max_i \min_j \mathbf{M}_{ij} \leq \min_j \max_i \mathbf{M}_{ij}$$

6.2 Randomized Strategies

Now we consider randomized strategies, and see how things can change. Recall that $\Delta(S)$ denotes all probability distributions on a set S.

- **Row** plays $x \in \Delta(R)$
- **Col** plays $y \in \Delta(C)$
- Expected payoff: $\mathbb{E}_{i,j}[\mathbf{M}_{ij}] = \sum_{i,j} x_i y_j \mathbf{M}_{ij} = x^\intercal \mathbf{M} y$

Now, let us revisit the Rock-Paper-Scissor game. Suppose **Row** plays a uniform strategy $x = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. Then it is easy to see that $\min_y x^\intercal \mathbf{M} y = 0$, which suggests that the column player has no real advantage for playing second.

Actually, for any normal-form zero-sum game, the second player gains no advantage from playing second, provided that the first player can randomize their strategy and the second player does not observe the realization of this randomization.

Theorem 2 (Minimax Theorem - Von Neumann '28) *There exists a value $val(\mathbf{M})$ s.t:*

$$\max_{x\in\Delta(R)} \min_{y\in\Delta(C)} x^\intercal \mathbf{M} y = \min_{x\in\Delta(R)} \max_{y\in\Delta(C)} x^\intercal \mathbf{M} y = val(\mathbf{M})$$

Proof: To simplify notations, we will write $U(x,y) = x^\intercal M y$.

Note that by the "everyone wants to go second" theorem, we have

$$\min_y \max_x U(x,y) \geq \max_x \min_y U(x,y)$$

We will proceed by proof by contradiction. Assume $\min\limits_y \max\limits_x U(x,y) = \max\limits_x \min\limits_y U(x,y) + \delta$, for some $\delta > 0$. Consider a thought experiment. We will let the two players repeatedly play the game against each other over T rounds. For each round t: the two players choose a pair of strategies (x^t, y^t) via:

- Min player (previously referred to as **Col** player) plays according to a no-regret algorithm (e.g., FTRL or multiplicative weights), using $\ell^t(y) = U(x^t, y)$ as loss function;
- Max player best responds:
$$x^t = \arg\max_x U(x, y^t)$$

Let $\text{Reg}_y = \sum_{t=1}^T U(x^t, y^t) - \min_y \sum_{t=1}^T U(x^t, y)$. Note that standard algorithms achieve $\text{Reg}_y = O(\sqrt{T})$. Let $\bar{x} = \frac{1}{T}\sum_{t=1}^T x^t$ denote the average play by the max player.

1. By no regret of min player:
$$\begin{aligned}\frac{1}{T}\sum_{t=1}^T U(x^t, y^t) - \frac{1}{T}\text{Reg}_y &\leq \frac{1}{T}\min_y \sum_{t=1}^T U(x^t, y),\\ &= \min_y U(\bar{x}, y),\\ &\leq \max_x \min_y U(x,y)\end{aligned}$$

2. By best response of max player:
$$\begin{aligned}\frac{1}{T}\sum_{t=1}^T U(x^t, y^t) &= \frac{1}{T}\sum_{t=1}^T \max_x U(x, y^t)\\ &\geq \frac{1}{T}\sum_{t=1}^T \min_y \max_x U(x,y)\\ &= \min_y \max_x U(x,y)\end{aligned}$$

By our earlier assumption, we expect a gap of δ:

$$\min_y \max_x U = \max_x \min_y U + \delta$$

Combining 1. and 2.:

$$\min_y \max_x U(x, y) \leq \max_x \min_y U(x, y) + \frac{\text{Reg}_y}{T}$$

Note that the average regret term $\frac{\text{Reg}_y}{T}$ decreases at a rate of $\frac{1}{\sqrt{T}}$. The fact that the average regret goes to 0 contradicts our assumption. ■

Definition 1 *A pair* (x^*, y^*) *such that*

$$\begin{cases} \min_y x^{*T}\mathbf{M}y = val(M) \\ \max_x x^T\mathbf{M}y^* = val(M) \end{cases}$$

is known as a minimax equilibrium*.*