**CS 6840: Algorithmic Game Theory** **Spring 2017**

# Lecture 13: February 24

*Lecturer: Éva Tardos* *Scribe: Shijin Rajakrishnan*

In this lecture, we introduce Two-Player Zero-Sum games, and use the learning algorithm from last lecture to prove an interesting result regarding the payoffs of the players.

## 13.1 Two-Player Zero-Sum Game

A game with two agents is called a **Two-Player Zero-Sum** Game if, for every possible outcome of this game, the payoff of one player is the negative of the payoff received by the other player. The gain of a player in such a game is the loss of another player, and thus it is an extreme example of a competitive game.

We need only one cost(or payoff) matrix to represent the game, and the costs(or payoffs) of the other player can be inferred by negating the matrix. By convention, the cost/payoff matrix associated with the game is taken to represent the row player.

A classic example of such a game is that of Rock-Paper-Scissors, and the payoff matrix $\mathbf{A}$ for this is shown below.

| | Rock | Paper | Scissors |
|---|---|---|---|
| Rock | 0 | -1 | 1 |
| Paper | 1 | 0 | -1 |
| Scissors | -1 | 1 | 0 |

Table 13.1: Rock-paper-Scissor: Payoff matrix for the row player

A strategy for a player here is a vector of probabilities, and we will denote these vectors as $\mathbf{x}$ for the row player(Player 1) and as $\mathbf{y}$ for the column player(Player 2).

In this case, given vectors $\mathbf{x}$, $\mathbf{y}$, and a payoff matrix $\mathbf{A}$ we can compactly express the payoff for Player 1 as

$$\mathbb{E}\Big[\text{Payoff for Player 1}\Big] = \sum_i \sum_j x_i y_j a_{ij} = \mathbf{x}^T\mathbf{A}\mathbf{y}$$

and the expected payoff for Player 2 is the negative of this.

## 13.2 Minimax Theorem

If we analyze how the players choose their mixed strategies, we can see if we allow the players to go in order, and suppose that Player 2 chooses a strategy $\mathbf{y}$ first, then Player 1 will mix his strategies so as to maximize his payoffs, and thus choose the vector $\mathbf{x}$ that optimizes $\max_{\mathbf{x}} \mathbf{x}^T\mathbf{A}\mathbf{y}$. Knowing this, Player 2, who goes first will choose a vector $\mathbf{y}$ that minimizes this payoff Player 1 gets, and thus chooses the vector that optimizes

$$\min_{\mathbf{y}} \left(\max_{\mathbf{x}} \mathbf{x}^T\mathbf{A}\mathbf{y}\right)$$

On the other hand, if Player 1 chooses his strategy $\mathbf{x}$ first, then Player 2 will choose a $\mathbf{y}$ so that her expected payoff is maximized(and thus the payoff of Player 1 is minimized) and consequently Player 1 when going first chooses his vector to be the optimizer of

$$\max_{\mathbf{x}} \left( \min_{\mathbf{y}} \mathbf{x}^T \mathbf{A} \mathbf{y} \right)$$

While it may seem at first blush that going first provides a disadvantage to the player in that he allows the other to optimize based on his decision, we will see in the rest of this lecture the surprising result that the two quantitities above are in fact equal! This is also called the value of the game.

There are several ways of proving this statement, such as by using Strong LP-Duality for a primal dual pair associated with each player, or by using the fact that all finite games have Nash equilibria. But in this lecture, we will prove it by using the learning algorithm we explored in the last class. However, the algorithm from last time dealt with a cost matrix(and a minimization objective) with uniform signs, whereas here we have that the payoffs to the two players are equal in magnitude but opposite in sign.

We begin by transforming our game to a form suitable for applying the multipplicative weights algorithm we saw, whilst maintaining the incentives of the players across the games. Note that the important property of the costs of the players in this Zero-Sum game wasn't that they were of the equal in opposite directions but that they summed to a constant across outcomes. In this sense, it would be better to call these constant sum games. For an example of the transformation, we can look at our previous exmaple -We add 1 to the costs of each outcome, to get the following cost matrix.

| | Rock | Paper | Scissors |
|---|---|---|---|
| Rock | 1 | 2 | 0 |
| Paper | 0 | 1 | 2 |
| Scissors | 2 | 0 | 1 |

Finally, we normalize the matrix so that the sum of the costs to the players is 1, across outcomes, and get the following cost matrix:

| | Rock | Paper | Scissors |
|---|---|---|---|
| Rock | $\frac{1}{2}$ | 1 | 0 |
| Paper | 0 | $\frac{1}{2}$ | 1 |
| Scissors | 1 | 0 | $\frac{1}{2}$ |

Now we are ready to formally prove the result.

**Theorem 13.1** *For every two-player zero-sum game, with payoff matrix* $\mathbf{A}$*,*

$$\min_{\mathbf{y}} \left( \max_{\mathbf{x}} \mathbf{x}^T \mathbf{A} \mathbf{y} \right) = \max_{\mathbf{x}} \left( \min_{\mathbf{y}} \mathbf{x}^T \mathbf{A} \mathbf{y} \right)$$

**Proof:** Both of the quantities are the payoffs for Player 1 - the one on the left represents his payoff when Player 2 goes first and the one on the right when he goes first.

One of the directions is easy - going second cannot be worse than going first since if $\tilde{x}$ is an optimal mixed strategy for Player 1 when he goes first, then by playing the same $\tilde{x}$ when going second and so we should have $\min_{\mathbf{y}} \left( \max_{\mathbf{x}} \mathbf{x}^T \mathbf{A} \mathbf{y} \right) \geq \max_{\mathbf{x}} \left( \min_{\mathbf{y}} \mathbf{x}^T \mathbf{A} \mathbf{y} \right)$. In terms of the cost matrix $\mathbf{C} = -\mathbf{A}$, this can be expressed as $\max_{\mathbf{y}} \left( \min_{\mathbf{x}} \mathbf{x}^T \mathbf{C} \mathbf{y} \right) \leq \min_{\mathbf{x}} \left( \max_{\mathbf{y}} \mathbf{x}^T \mathbf{C} \mathbf{y} \right)$.

We prove the other direction by using the No-regret dynamics explored in the last chapter.

We start by assuming that the two quantities are not equal, and that their difference is a positive constant $4\epsilon$,

$$\min_{\mathbf{x}}\left(\max_{\mathbf{y}}\mathbf{x}^T\mathbf{C}\mathbf{y}\right)-\max_{\mathbf{y}}\left(\min_{\mathbf{x}}\mathbf{x}^T\mathbf{C}\mathbf{y}\right)=4\epsilon$$

Suppose that the players play this game by each using the Multiplicative weights algorithm for a time horizon $T$(to be fixed later), and that the mixed strategies of Player 1 are $\mathbf{p^1},\mathbf{p^2},...,\mathbf{p^T}$ and that of Player 2 are $\mathbf{q^1},\mathbf{q^2},...,\mathbf{q^T}$.

From the no regret guarantee, we have that the expected cost of Player 1 is bounded as

$$\sum_t\langle\mathbf{p^t},\mathbf{Cq^t}\rangle\leq(1+\epsilon)\min_i\left(\sum_t\sum_j c_{ij}q_j^t\right)+\frac{1}{\epsilon}\log(n_1)$$

where the first term on the right hand side uses that the cost of using a single strategy $i$ all the time is $\sum_t(\mathbf{Cq^t})_i=\sum_t\sum_j c_{ij}q_j^t$, and $n_1$ is the number of strategies avaiable to Player 1.

We can divide by $T$ and look at the average cost to Player 1, $C_1=\frac{1}{T}\sum_t\langle\mathbf{p^t},\mathbf{Cq^t}\rangle$,

$$C_1\leq(1+\epsilon)\frac{1}{T}\min_i\left(\sum_t\sum_j c_{ij}q_j^t\right)+\frac{1}{\epsilon}\log(n_1)=(1+\epsilon)\min_i\sum_j c_{ij}\left(\frac{1}{T}\sum_t q_j^t\right)+\frac{1}{T\epsilon}\log(n_1)$$

where we interchanged the summations to obtain the equality.

Now note that the vector $\mathbf{q}=\frac{1}{T}\sum_t\mathbf{q^t}$ is the average distribution that Player 2 plays, and substituting this into the above expression we get

$$C_1\leq(1+\epsilon)\min_i\sum_j c_{ij}q_j+\frac{1}{T\epsilon}\log(n_1)$$

Using fact that all the costs were normalized so that $c_{ij}\leq 1\forall i,j$, and $\sum_j q_j=1$, we have

$$C_1\leq\min_i\sum_j c_{ij}q_j+\epsilon+\frac{1}{T\epsilon}\log(n_1)\tag{13.1}$$

We can obtain a similar equation for the average expected cost to Player 2,

$$C_2\leq\min_j\sum_j(1-c_{ij})p_i+\epsilon+\frac{1}{T\epsilon}\log(n_2)$$

In the above equation, we have used $C_2$ to denote the average expected cost of Player 2, $n_2$ to denote the number of strategies available to Player 2, and $\mathbf{p}=\frac{1}{T}\sum_t\mathbf{p^t}$ is the average of the distributions played by Player 1. Further, we have used that this is a constant sum game, and that the cost to Player 2 for the outcome $ij$ is 1 minus the cost to Player 1 for the same outcome.

We can simplify this by using the fact that $\sum_i p_i=1$ to get

$$C_2\leq 1-\max_j\sum_j c_{ij}p_i+\epsilon+\frac{1}{T\epsilon}\log(n_2)\tag{13.2}$$

Now we pick the time horizon $T$ to be large enough so that

$$\epsilon + \frac{1}{T\epsilon}\log(n_1) < 2\epsilon \quad \text{and} \quad \epsilon + \frac{1}{T\epsilon}\log(n_2) < 2\epsilon$$

Since the cost of every outcome to the two players sums to 1, the cost that the two players experience at any time $t$ also sums to 1, and thus the average costs $C_1 + C_2 = 1$. This yields

$$1 - C_1 = C_2 \leq 1 - \max_j \sum_j c_{ij} p_i + \epsilon + \frac{1}{T\epsilon}\log(n_2) < 1 - \max_j \sum_j c_{ij} p_i + 2\epsilon$$

which on rearrangement gives

$$C_1 > \max_j \sum_j c_{ij} p_i - 2\epsilon$$

Combining this with Equation 13.1 produces

$$\max_j \sum_j c_{ij} p_i - 2\epsilon < C_1 \leq \min_i \sum_j c_{ij} q_j + \epsilon + \frac{1}{T\epsilon}\log(n_1) < \min_i \sum_j c_{ij} q_j + 2\epsilon$$

so we have

$$\max_j \sum_j c_{ij} p_i - \min_i \sum_j c_{ij} q_j < 4\epsilon$$

This is a contradiction as $\max_j \sum_j c_{ij} p_i \geq \min_{\mathbf{x}} \max_j \sum_j c_{ij} x_i$ and $\min_i \sum_j c_{ij} q_j \leq \max_{\mathbf{y}} \min_i \sum_j c_{ij} y_j$, so we also get

$$\min_{\mathbf{x}} \max_j \sum_j c_{ij} x_i - \max_{\mathbf{y}} \min_i \sum_j c_{ij} y_j < 4\epsilon$$

This yields the desired contradiction, since we initially assumed that the difference in this gap was $4\epsilon$, but the no regret learning leads us to a strictly smaller gap.

This thus proves the other direction as well.

■