CS 6840 Algorithmic Game Theory September 16, 2024

Lecture 9: Zero-Sum Games with 2 No Regret Learners

Instructor: Eva Tardos *Scribe: Adit Jain*

1 Recap

Recall that the reward/loss for the players in a two player zero-sum game can be represented compactly by the following matrix:

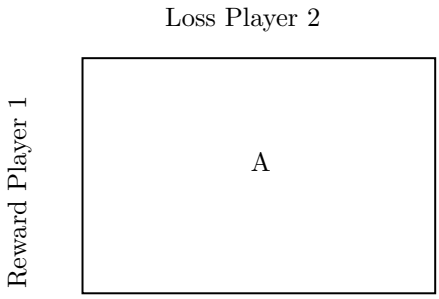

Figure 1: Compact matrix representation for a two player zero sum game. Here the rows denotes the reward (postive payoff) for player 1 (referred to as row player) and the columns denote the cost (negative payoff) for player 2 (referred to as column player).

Entries of this matrix are in the range $\in [-1, 1]$.

We denote a randomized strategy of player 1 by x, and if the strategy is different across different time instances by, $x^1, \ldots, x^t$. Each x is a probability distribution over the different strategies (actions) that the player can take.

Similarily we use $y^1, \ldots, y^t$ to denote the mixed strategies of player 2

Why are these strategies different across time? In case of learning, the players are given the choice to evolve their strategy in response to the strategy of the other player. However when we evaluate this strategy we benchmark it against about the best possible *stationary* (non-changing) strategy and look at the difference. In no-regret learning fro previous lecture we showed that the no-regret learning algorithm actually gets you an average strategy which performs better than the best possible strategy choosen in advance.

2 The average of the learning iterates is a nash

The loss of Player 2 at time t for action j is denoted by: $\sum_i a_{ij} x_i^t$. Since given the strategy x^t, this the expected loss that it encounters for action j at time t.

Similarly loss of Player 1 at time t for action i: $-\sum_j a_{ij} y_j^t$. Notice the negative sign which is due to the fact that player 1 has positive payoffs and so the loss will be negative to the payoff it receives (higher the payoff $\implies$ more smaller (more negative) the loss).

The net expected payoff of player 1 is : $\sum_t (x^t)^T A y^t$.

No regret learning for player 2 (previous lecture) :

$$\sum_t (x^t)^T A y^t \leq \min_j \sum_t \sum_i x_i^t a_{ij} \tag{1}$$

No regret learning for player 1 implies (previous lecture) :

$$-\sum_t (x^t)^T A y^t \leq \min_i \sum_t \sum_j y_j^t a_{ij} \tag{2}$$

Let us consider a strategy using the average of the intermediate strategies, $\bar{x} = \frac{1}{T}\sum_t x^t$

$\bar{y} = \frac{1}{T}\sum_t y^t$

Claim: $\bar{x}$ and $\bar{y}$ is Nash.

Let T be the time played for (Total time the learning algorithm is run)

$$v = \frac{1}{T}\sum_t (x^t)^T A y^t$$

where v is the value of the game.

We can derive the following about player 2:

$$\begin{aligned} & v \cdot T \leq \min_j \sum_i a_{ij} T \bar{x}_i \\ \implies & v \cdot T \leq \sum_i a_{ij} T \bar{x}_i \ \forall j \text{ if the previous inequality is true for min, then true for all} \\ \implies & v \leq \sum_i a_{ij} \bar{x}_i \end{aligned}$$

We can now infer that if Player 1 commits to play $\bar{x}$, then they are guaranteed to gain at least $\geq v$.

Performing a similar analysis for player 1 (taking negative on both sides of the no regret equation turns the negative into positive),

$$\begin{aligned} v \cdot T &\geq \max_i \sum_j a_{ij} \bar{y}_j \\ v &\geq \sum_j a_{ij} \bar{y}_j \ \forall i \end{aligned}$$

We can infer that if player 2 commits to play $\bar{y}$, then gauranteed to lose $\leq v$.

$(\bar{x}, \bar{y})$ is Nash and v is value.

Proof. First we show that the value of the game with a strategy $(\bar{x}, \bar{y})$, is exactly v.

1. $\bar{x}^T A \bar{y} \leq v$ (By Analysis of Player 1 of strategies)

2. $\bar{x}^T A\bar{y} \geq v$ (By Analysis of Player 2 using a convex combination of strategies)

Then we claim that this indeed is the best that player 1 can do (by analyzing the no regret equation of the player 1 Eq. (2)) and similarly for player 2 (Eq. (1)). ■

3 What if the learning is not truly no-regret?

We now introduce the error term in the equations (2) and (1) and show by contradiction that given enough learning rounds, any error can be recovered for by the solution of the learning.

Assume that the error is defined with respect to value for Player 1,

$\Delta :=$ value if row chooses first − value if column chooses first.

Recall: Regret bound from multiplicative weight algorithm: $\sqrt{2ln(K)T}$.

$k = \#\text{strategies}, T = \text{Time period}.$

Now, Eq. (2) gets added term of $\sqrt{\frac{2\ln(K)}{T}}$

Eq. (1) gets subtracted term of $\sqrt{\frac{2\ln(K)}{T}}$

This is from the fact that the learning algorithm has an error in minimizing the loss function (and the obtained loss function of the iterates are less than best by the error amount). And the difference in signs for the error arise from the fact that the two losses are in opposite directions.

We can choose T so that $\sqrt{\frac{2\ln(K)}{T}} < \frac{\Delta}{2}$

Therefore plugging the errors into equations (2) and (1)

$$v < \sum_i a_{ij}\bar{x}_i + \frac{\Delta}{2} \ \forall j$$

If Player 1 (row) commits to $\bar{x}$ guaranteed to gain $> v - \frac{\Delta}{2}$

$$v > \sum_j a_{ij}\bar{y}_j \ \forall i - \frac{\Delta}{2}$$

If Player 2 (column) commits to $\bar{y}$ guaranteed to lose $< v - \frac{\Delta}{2}$.

$\implies$ going first difference strictly less than $< v + \frac{\Delta}{2} - (v - \frac{\Delta}{2}) = \Delta$, which is a contradiction to our assumption.

Comment on Last Iterate Convergence: There is a section of algorithmic game theory literature which is focussed on convergence of the last iterate to the Nash, however we restrict ourself to the convergence of the mean of the iterates.

4 Alternate Version for constructing the value

:

We can also consider the loss defined by random instance of the strategy observed of opponent (rather than average using the mixed strategy)

Let $i_1, \ldots, i_t$ and $j_1, \ldots, j_t$ denote the strategies.

We can write the Value for player 1. $\sum a_{i_t, j_t}$

The loss is given by $l_t^c(j) = a_{i_t, j}$ and $l_t^r(j) = -a_{i, j_t}$

We can compute expectation for value for player one, $\mathbb{E}\sum_{t=1}^{T} a_{i_t, j_t} = \sum_{t=1}^{T} (x^t)^T A y^t$

And the expectations for $\mathbb{E}\sum_{t=1}^{T} a_{i_t, j} = \sum_{t=1}^{T} \sum_j x_i^t a_{ij}$ a particular action of player 2.

We can see that we recover the expectation terms for equations (2) and (1).