CS295 Introduction to Algorithmic Game Theory

Instructor: Ioannis Panageas
Scribed by: Sehwan Hong, Radhit Dedania, Rohith Reddy Gangam

Lecture 4. Online learning and a proof of minimax theorem.

1 Abstract

The previous lectures talk about LP duality and their application to Zero-sum games. They also show how the minimax theorem emerges as a corollary of LP Strong duality. In this lecture, the concepts of 'Expert games', 'No-regret algorithm' and 'Multiplicative Weights Update' are introduced. The results and properties of these algorithms will be used to give an alternative proof of the Von Neumann minimax theorem.

2 Playing The Expert Game[3, 4]

2.1 Definition

Expert game is a type of game where an expert predicts the result of the game. For example, we could think about the weather forecast. In this situation, the forecast center becomes the experts and predicts whether it will rain or not. This kind of game is called Expert game and it will be defined as following.

Definition 2.1 (Playing the Expert Game) *For each day $t = 1...T$, you have to choose between alternative A, B*

- *Choose A or B according to some rule.*
- *One of the alternatives realizes.*
- *If you choose correctly you are not penalized otherwise you lose one point.*
- *Imagine that there are n experts who on each day t, recommend either A or B*

In the expert game, expert cannot be always correct. However we could gather the expert's opinion, and try to perform close to best expert. In zero information, choose the majority's opinion, then penalize(give less weight to their future opinions) those who are incorrect. Repeating this, we are able to perform close to the best expert.

2.2 Algorithm

From above definition, we design Algorithm 1. This algorithm is called weighted Majority. The first step is initializing all the weights to 1 to find the majority.

After the first step, we update the weights for every time-step. For every mistake an expert makes, their weight will be decreased by a factor of ϵ. However, if they did not make the mistake, then the weight value will be maintained. ϵ is the step-size, though experiment, step-size will be chosen to make the best prediction.

As the time-steps increase, the algorithm will improve to get better result. This algorithm performs almost as good as "best" expert. In other words, the number of mistakes by this algorithm is bounded by the number of mistakes by the best expert from our pool of experts.

Algorithm 1 Weighted Majority

Initialize $w_i^0 = 1$ for all $i \in [n]$
for $t = 1...T$ **do**
  if $\Sigma_{i \text{ choose } A} w_i^{t-1} \geq \Sigma_{i \text{ choose } B} w_i^{t-1}$ **then**
    Choose A
  else
    Choose B
  end if
  for expert i that made mistake **do**
    $w_i^t = (1-\epsilon) w_i^{t-1}$
  end for
  for expert i that did not make mistake **do**
    $w_i^t = w_i^{t-1}$
  end for
end for

2.3 Theorem

Theorem 2.1 (Weighted Majority) *Let M_T, M_T^B be the total number of mistakes the algorithm and best expert make until time step T, respectively. It holds that*

$$M_T \leq 2(1+\epsilon) M_T^B + \frac{2 \log n}{\epsilon} \tag{1}$$

Proof: Let's define potential function $\phi = \Sigma_i w_i^t$ From this equation we are able to find two important aspect of the potential function:

- $\phi_0 = n$
- $\phi_{t+1} \leq \phi_t$

From the above algorithm 1, we could observe that at each time step, the value of phi could decrease when the experts makes a mistake, but when they do not make the mistake the weight remains. From this observation, we could see that $\phi_{t+1} \leq \phi_t$.

At time t,if there is a mistake, it means majority of the experts have made a mistake, so at the least $\frac{\phi_t}{2}$ will be multiplied by $(1-\epsilon)$ and remaining will maintain its value. From this, we get the following equation:

$$\phi_{t+1} \leq (1-\epsilon) \cdot \frac{\phi_t}{2} + \frac{\phi_t}{2} = (1 - \frac{\epsilon}{2})\phi_t \tag{2}$$

So, when we make a mistake, $\phi_{t+1} \leq (1 + \frac{\epsilon}{2})\phi_t$. On the other hand, when we do not make mistake, $\phi_{t+1} \leq \phi_t$. Since we are making M_T mistakes, expanding the result we get:

$$\phi_{t+1} \leq (1 - \frac{\epsilon}{2})^{M^T}\phi_1 \tag{3}$$

Moreover, assuming the best expert (say i^*) did M_T^B mistakes, we have $\phi_T > w_{i^*}^T = (1-\epsilon)^{M_T^B}$. Using this equation and the equation 3, we conclude:

$$(1-\epsilon)^{M_T^B} < \phi_t \leq (1 - \frac{\epsilon}{2})^{M^T}\phi_0$$

$$\implies (1-\epsilon)^{M_T^B} < (1 - \frac{\epsilon}{2})^{M^T} n \tag{4}$$

When we apply the logarithms on both side of inequality (4), we will get (5).

$$M_T^B \log(1-\epsilon) < M_T \log(1 - \frac{\epsilon}{2}) + \log n \tag{5}$$

$$\implies M_T^B(-\epsilon - \epsilon^2) < -M_T\epsilon/2 + \log n$$

Since $-x - x^2 < \log(1-x) < -x$, we could substitute (5) to eliminate the logarithm. By eliminating the logarithms, we are are able to simplify the result to get the (1).

$$\begin{aligned} M_T^B(-\epsilon - \epsilon^2) &< -M_T\epsilon/2 + \log n \\ \implies M_T\epsilon/2 &< M_T^B(\epsilon + \epsilon^2) + \log n \\ \implies M_T/2 &< M_T^B(1+\epsilon) + \frac{\log n}{\epsilon} \\ \implies M_T &< 2M_T^B(1+\epsilon) + \frac{2\log n}{\epsilon} \end{aligned}$$

$$M_T < 2M_T^B(1+\epsilon) + \frac{2\log n}{\epsilon} \tag{6}$$

This proves that the number of mistakes committed by the weighted majority theorem algorithm is bounded w.r.t. the number of mistakes by the best expert.

■

2.4 General Setting

The expert game setting can be generalised as below.

Definition 2.2 *At each time step* $t = 1...T$

- *Player choose* $x_t \in \Delta_n$
- *Adversary chooses* $u_t \in [-1, 1]^n$
- *Player gets payoff* $x_t^\top u_t$ *and observes* u_t

Player's goal is to minimize the (time average) Regret, that is :

$$\frac{1}{T}\left[\max_{x\in\Delta_n}\sum_{t=1}^{T} x^\top u_t - \sum_{t=1}^{T} x_t^\top u_t\right] = \frac{1}{T}\left[\max_{i^*\in[n]}\sum_{t=1}^{T} u_{t,i} - \sum_{t=1}^{T} x_t^\top u_t\right] \tag{7}$$

The goal of the player is to minimize the regret using different expert's prediction. As described in the equation above, the player is trying to minimize the regret and compare with the best expert's prediction. In this expert game, if Regret gets to zero as T becomes ∞, the algorithm is said to have no-regret.

3 Multiplicative Weights Update[3, 4]

3.1 Algorithm

Following up on the above algorithm, we define a variant of it below(Algorithm 2). In contrast to the above algorithm, this one updates the beliefs(with respect to each expert) of each player irrespective of the status of their mistakes. This algorithm is called Multiplicative Weights Update as, at each step, it tries to modify the current belief by adding a multiplicative factor of its payoff. Here, the player has a some belief in every expert and they sum to 1 at each time step as the beliefs constitute a probability distribution.

The algorithm starts by initializing belief in every expert to the same constant $\frac{1}{n}$. Then, for each time step and each expert, it updates the belief of the player in that expert at that time step by adding a fraction of its payoff(computed using the observed adversary realisation at that time step) to the current belief in that expert. A re-normalization factor is also included in each update to ensure that the updated beliefs represent a probability distribution. The algorithm is stated below (Algorithm 2). Here, ϵ refers to the step-size(which will be chosen later) and $Z^t = \sum_i p_i^t(1+\epsilon u_{t,i})$ is the re-normalization constant. This algorithm performs almost as good as the `best` expert(fewest mistakes).

Algorithm 2 Multiplicative Weights Update

Initialize $p_i^0 = \frac{1}{n}$ for all $i \in [n]$
for $t = 1...T$ **do**
  for each i that gives payoff $u_{t,i}$ **do**

$$p_i^{t+1} = p_i^t \frac{1+\epsilon u_{t,i}}{Z^t} \tag{8}$$

  end for
end for

3.2 Theorem

Theorem 3.1 (Multiplicative Weights Update) *It holds that*

$$\frac{1}{T}\sum_t u_t^T p^t \geq \max_x \sum_t x^T u_t - \frac{log\, n}{\epsilon T} - \epsilon \tag{9}$$

Proof: Let's define the potential function $\phi = \Sigma_i w_i^t$ where weights $w_i^t = \prod_{s=0}^{t}(1+\epsilon u_{s,i})$. It can be observed that the potential function obeys the following properties:

- $\phi_0 = n$
- $\phi_t \geq 0$, for all $t \in \{0, 1, \ldots T\}$

As $1 + \epsilon u_{s,i} \geq 0$ for every s and w_i^t is defined as the product of these terms, $\phi_t \geq 0$ as it is the sum over all such positive w_i^t values. As $w_i^0 = 1$ since $p_i^0 = \frac{1}{n}$ for all $i \in [n]$ and $\phi_0 = \Sigma_i w_i^0$, $\phi_0 = n$.

Let the best strategy be i^*, the following equation holds as ϕ_T which is the sum over w_i^T values for every $i \in [n]$ is always greater than a particular constituent term(here, it is the weight of the best expert, $w_{i^*}^T$)

$$\phi_T > w_{i^*}^T$$

Since $\log(1+x) \geq x - x^2$ or in other words, $1 + x \geq e^{x-x^2}$ where x = $\epsilon u_{s,i^*}$, we could rewrite the above equation as shown below.

$$\phi_T > w_{i^*}^T \geq e^{\epsilon \sum_{s=0}^{T} u_{s,i^*} - \epsilon^2 \sum_{s=0}^{T} u_{s,i^*}^2} \tag{10}$$

As ϕ_{t+1} is the sum over w_i^{t+1} values for every $i \in [n]$ and w_i^{t+1} can be written recursively in terms of w_i^t as stated in the equation below

$$\phi_{t+1} = \sum_i w_i^{t+1} = \sum_i w_i^t (1 + \epsilon u_{t,i})$$

As $p_i^t = \frac{w_i^t}{Z^t} = \frac{w_i^t}{\phi_t}$ since $\phi_t = \sum_i w_i^t = Z^t$ (using the fact that re-normalizing updated beliefs at each iteration yields the same result as re-normalizing the beliefs once at the end of all iterations along with the application of telescopic method to the belief update equation (8) for all $t \in \{0, 1, \ldots T\}$ to eliminate intermediate p_i^ts), we get the following equation.

$$\phi_{t+1} = \sum_i w_i^{t+1} = \sum_i w_i^t (1 + \epsilon u_{t,i}) = \sum_i \phi_t p_i^t (1 + \epsilon u_{t,i})$$

Since ϕ_t is independent of i, it can be taken out of the summation to get the below expression.

$$\phi_{t+1} = \sum_i w_i^{t+1} = \sum_i w_i^t (1 + \epsilon u_{t,i}) = \sum_i \phi_t p_i^t (1 + \epsilon u_{t,i}) = \phi_t \sum_i p_i^t (1 + \epsilon u_{t,i})$$

In other words,

$$\phi_{t+1} = \phi_t \sum_i p_i^t (1 + \epsilon u_{t,i})$$

Since beliefs p_i^t for every $i \in [n]$ constitute a probability distribution, $\sum_i p_i^t = 1$ and so the first term equals to 1 when $\sum_i p_i^t = 1$ is brought inside the sum term $(1 + \epsilon u_{t,i})$. As ϵ is a constant, it will stay out of $\sum_i p_i^t$ in the second term.

$$\phi_{t+1} = \phi_t(1 + \epsilon \sum_i p_i^t u_{t,i})$$

Using the fact that $\log(1 + x) \leq x$ or in other words, $1 + x \geq e^x$ where x $= \epsilon \sum_i p_i^t u_{t,i}$, we get,

$$\phi_{t+1} = \phi_t(1 + \epsilon \sum_i p_i^t u_{t,i}) \leq \phi_t e^{\epsilon \sum_i p_i^t u_{t,i}}$$

Using the vector notation, $\sum_i p_i^t u_{t,i}$ becomes $u_t^T p^t$.

$$\phi_{t+1} = \phi_t(1 + \epsilon \sum_i p_i^t u_{t,i}) \leq \phi_t e^{\epsilon \sum_i p_i^t u_{t,i}} = \phi_t e^{\epsilon u_t^T p^t} \tag{11}$$

Enumerating (11) by listing it at each time step $t \in \{0, 1, \ldots T\}$, we get,

$$\begin{aligned}
\cancel{\phi_1} &\leq \phi_0 e^{\epsilon u_0^T p^0} \\
\cancel{\phi_2} &\leq \cancel{\phi_1} e^{\epsilon u_1^T p^1} \\
\cancel{\phi_3} &\leq \cancel{\phi_2} e^{\epsilon u_2^T p^2} \\
\cancel{\phi_4} &\leq \cancel{\phi_3} e^{\epsilon u_3^T p^3} \\
&\vdots \\
\phi_T &\leq \cancel{\phi_{T-1}} e^{\epsilon u_{T-1}^T p^{T-1}}
\end{aligned}$$

Multiplying the above equations(using telescopic product), we get,

$$\phi_T \leq \phi_0 e^{\epsilon \sum_t u_t^T p^t}$$

Since $\phi_0 = n$ as shown above, ϕ_T becomes upper bounded by $n e^{\epsilon \sum_t u_t^T p^t}$

$$\phi_T \leq \phi_0 e^{\epsilon \sum_t u_t^T p^t} = n e^{\epsilon \sum_t u_t^T p^t} \tag{12}$$

Therefore, from (10)[lower bound of ϕ_T] and (12)[upper bound of ϕ_T], we can state that,

$$e^{\epsilon \sum_{s=0}^T u_{s,i^*} - \epsilon^2 \sum_{s=0}^T u_{s,i^*}^2} \leq n e^{\epsilon \sum_t u_t^T p^t}$$

Since $\sum_{s=0}^T u_{s,i^*}^2 \leq T$ as $\max u_{s,i^*} = 1$ for all $s \in \{0, 1, \ldots T\}$, we can get a lower bound of $e^{\epsilon \sum_{s=0}^T u_{s,i^*} - \epsilon^2 \sum_{s=0}^T u_{s,i^*}^2}$ as follows,

$$e^{\epsilon \sum_{s=0}^T u_{s,i^*} - \epsilon^2 T} \leq e^{\epsilon \sum_{s=0}^T u_{s,i^*} - \epsilon^2 \sum_{s=0}^T u_{s,i^*}^2} \leq n e^{\epsilon \sum_t u_t^T p^t}$$

As log is a strictly increasing function, by taking log on both sides of the inequality will not change the sign of it. Thus, it can be written as,

$$\epsilon \sum_{s=0}^T u_{s,i^*} - \epsilon^2 T \leq \epsilon \sum_{s=0}^T u_{s,i^*} - \epsilon^2 \sum_{s=0}^T u_{s,i^*}^2 \leq \log n + \epsilon \sum_t u_t^T p^t$$

Equivalently, the equation becomes,

$$\epsilon \sum_{s=0}^{T} u_{s,i^*} - \epsilon^2 T \leq \log n + \epsilon \sum_{t} u_t^T p^t$$

Dividing both sides of the equation by ϵT and rearranging it, we get the main equation which we had set out to prove.

$$\frac{1}{T} \sum_{t} u_t^T p^t \geq \max_{x} \sum_{t} x^T u_t - \frac{log\, n}{\epsilon T} - \epsilon \tag{13}$$

■

3.3 Choice of Step-size(Epsilon)

As we can choose ϵ arbitrarily, we can try the below mentioned value which gives us good interpretation of the player's strategy with respect to that of the best expert. Regret can be alternately understood as the number of additional mistakes made by the player in comparison to the best expert. As regret is always non-zero(as no player knows what their best strategy should be while playing), the player always gets worse payoff compared to the best expert.

- Setting $\epsilon \longrightarrow \sqrt{\frac{lnn}{T}}$, we get that the regret is bounded by $2\sqrt{\frac{lnn}{T}}$. As T tends to ∞, the above tends to zero, so this is a no-regret algorithm.

4 Minimax Theorem

Theorem 4.1 (Minimax Theorem by John Von Neumann) *Let A be a matrix of size $n \times m$.*

$$\min_{x \in \Delta_n} \max_{y \in \Delta_m} x^\top A y = \max_{y \in \Delta_m} \min_{x \in \Delta_n} x^\top A y \tag{14}$$

The above is defined to be the **value** of the game.

Von Neumann published the above result in 1928 [1]. We will prove the same result using the theory presented above.

Proof: To prove the above equality, we first prove that the LHS $\geq$ RHS. For this, we use the following well-known, slightly trivial lemma.

Lemma 4.2 *For all functions $f(x, y)$*

$$\inf_{x \in X} \sup_{y \in Y} f(x, y) \geq \sup_{y \in Y} \inf_{x \in X} f(x, y) \tag{15}$$

Proof: Define a function $g(y)$ based on the function $f(x, y)$ as follows.

$$g(y) = \inf_{x \in X} f(x, y)$$

Since $g(y)$ is defined as the infimum over x of f, and using inf and sup operators on the free variables, we get

$$\begin{aligned}
&\forall x, \forall y, g(y) \leq f(x,y) \\
&\implies \forall x, \sup_y g(y) \leq \sup_y f(x,y) \\
&\implies \sup_y g(y) \leq \inf_x \sup_y f(x,y) \\
&\implies \sup_y \inf_x f(x,y) \leq \inf_x \sup_y f(x,y)
\end{aligned}$$

■

And so, irrespective of the function $f(x,y)$, the above inequality always holds. Choosing $f(x,y) = x^\top Ay$ for $x \in \Delta_n$ and $y \in \Delta_m$ and the given matrix A, we get that

$$\min_{x \in \Delta_n} \max_{y \in \Delta_m} x^\top Ay \geq \max_{y \in \Delta_m} \min_{x \in \Delta_n} x^\top Ay \tag{16}$$

■

4.1 Using MWU to prove the minimax theorem

And to prove the other inequality, we will model the game from the viewpoint of "Multiplicative Weights Update for the general setting"(MWU). In the general setting, there was only one player playing against an adversary. So, when player x is playing, we view player y as its adversary and he will use the MWU iterates as his strategy. Now, player y responds to this using his own MWU iterates viewing player x as his adversary. Viewed as a whole, this is a "game" between the two players but separately this should satisfy the no-regret results obtained in the previous sections.

Let $x_1, \ldots, x_T$ and $y_1, \ldots, y_T$ be the iterates advised by MWU to each player. Lets also define $\hat{x} = \frac{1}{T}\sum_{i=1}^T x_i$ and $\hat{y} = \frac{1}{T}\sum_{i=1}^T y_i$ and let $T = \Theta\big(\frac{1}{\eta^2}\big)$ i.e., the game has been run for long enough time steps that the regret is bound by η.

And so, choosing any x, from the no-regret property for x, we get that,

$$\frac{1}{T}\sum_t x_t^\top Ay_t \leq \frac{1}{T}\sum_t x^\top Ay_t + \eta = x^\top A\left(\frac{\sum_t y_t}{T}\right) + \eta$$

Similarly, choosing any y, noting that he is trying to oppose of player x in terms of the 'game's value', the no-regret property gives,

$$\frac{1}{T}\sum_t x_t^\top Ay_t \geq \frac{1}{T}\sum_t x_t^\top Ay - \eta = \left(\frac{\sum_t x_t}{T}\right)^\top Ay - \eta$$

Joining both the above equations, for all x, y we have,

$$\left(\frac{\sum_t x_t}{T}\right)^\top Ay - 2\eta \leq x^\top A\left(\frac{\sum_t y_t}{T}\right)$$

Using the facts that x and y are free variables in the above equations, and maximum and minimum of a set sandwich the averages, we get

$$\min_x x^\top A\left(\frac{\sum_t y_t}{T}\right) \geq \max_y \left(\frac{\sum_t x_t}{T}\right)^\top Ay - 2\eta$$

$$\implies \max_y \min_x x^\top Ay \geq \min_x x^\top A\left(\frac{\sum_t y_t}{T}\right)$$

$$\implies \max_y \min_x x^\top Ay \geq \max_y \left(\frac{\sum_t x_t}{T}\right)^\top Ay - 2\eta$$

$$\max_y \min_x x^\top Ay \geq \min_x \max_y x^\top Ay - 2\eta \tag{17}$$

The last inequality, along with the previous lemma, tells us that, for any chosen η, we can reach to a stage where the value of the games is within a 2η interval. But, increasing the number if time steps $T \to \inf$, we can get η as close to zero and hence, proving that the value of the games is constant and equal to both $\min_{x\in\Delta_n} \max_{y\in\Delta_m} x^\top Ay$ and $\max_{y\in\Delta_m} \min_{x\in\Delta_n} x^\top Ay$.

Using MWU in the above way, leads us to the following trivial algorithm 3 to find an ϵ-approximate Nash Equilibrium for the case of two player Zero sum game.

Algorithm 3 Computing ϵ-approximate Nash Equilibrium

Initialize $x_i^0 = \frac{1}{n}$ for all $i \in [n]$ and $y_i^0 = \frac{1}{n}$ for all $i \in [n]$ and $T = \Theta\left(\frac{1}{\epsilon^2}\right)$
for $t = 1...T$ **do**
  for $i \in [n]$ **do**
    $x_i^{t+1} = x_i^t \frac{1-\epsilon(Ay^t)_i}{1-\epsilon {x^t}^\top Ay^t}$
  end for
  for $j \in [n]$ **do**
    $y_j^{t+1} = y_j^t \frac{1+\epsilon(A^\top x^t)_j}{1+\epsilon {x^t}^\top Ay^t}$
  end for
end for
Return $(\hat{x}, \hat{y}) = \left(\left(\frac{1}{T}\sum_{i=1}^T x_i\right), \left(\frac{1}{T}\sum_{i=1}^T y_i\right)\right)$

5 Summary

In this lecture, we have covered basic online learning using expert game. Expert game is a type of game where an expert predicts the result of the game. Since experts cannot be always correct, we use Weighted Majority Algorithms to get the best result using multiple experts. Weighted Majority Algorithm updates every weights by checking if the experts have made a mistake or not.

The variant of Weighted Majority Algorithms is Multiplicative Weights Update. This algorithms updates the beliefs with respect to each experts of each player. Compared to WM algorithm, the weights of the WMU algorithm must sum up to 1 and it represents the probability distribution of belief.

We use the results of Multiplicative Weights Update to show that, a two player zero-sum game has a **value** and give a proof of **minimax theorem**. A modified MWU algorithm to compute an ϵ-approximate Nash Equilibrium of the game is provided.

6 Do you know?

- The theory of online learning was proposed by Hannan in 1957 which is interestingly the same year when the game of battle of sexes was introduced and has an interesting application to the casino problem(also known as multi-arm bandit problem).
- Exploration(switching to another expert) and exploitation(continuing with the same expert) also derive their origin from the bandit problem wherein at each step the choice of continuing with the same machine or switching to another is to be made to maximize monetary gains.
- The first proof of minimax theorem was proposed by Von Neumann as early as 1928([1]) but at that time, he had no knowledge of the theory of linear inequalities and fixed-point theorem and hence, he came up with another proof in 1944([2]) using these topics which was much simpler to understand than the former.
- The most prominent usage of online learning nowadays is in online advertisements where the website has to instantly come up with an advertisement for the new user. If the user opens the banner, then the website earns revenue and their end goal is to maximize it by showing relevant advertisements to the visitor.

References

[1] John von Neumann. *Zur Theorie der Gesellschaftsspiele(On the theory of board games)*. Math. Ann. 100, 295–320 (1928).

[2] John von Neumann, Oscar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Presss, Princeton, NJ (1944).

[3] Arora, Sanjeev, Elad Hazan, and Satyen Kale. *The Multiplicative Weights Update Method: Meta Algorithm and Applications*. Theory of Computing 8.1 (2012): 121-164. `http://www.theoryofcomputing.org/articles/v008a006/v008a006.pdf`

[4] Si Yi Meng. *Multiplicative Weights Update*. Lecture Slides. University of British Columbia(2019).
`https://www.cs.ubc.ca/labs/lci/mlrg/slides/2019_summer_2_multiplicative_weight_update.pdf`

[5] Sanjeev Arora. *Lecture 8: Decision-making under total uncertainty: the multiplicative weight algorithm.* `https://www.cs.princeton.edu/courses/archive/fall13/cos521/lecnotes/lec8.pdf`.