6.853 Topics in Algorithmic Game Theory	September 22, 2011

Lecture 5

Lecturer: Constantinos Daskalakis — *Scribe: Constantinos Daskalakis*

We continue our discussion from last lecture, looking at the performance of *Multiplicative-Weights-Updates*. We then see how to use MWU in the context of zero-sum games hence concluding our discussion of zero-sum games. In the end of the lecture, we switch contexts to general games, presenting the elegant proof of Nash's theorem.

1 Performance of Multiplicative-Weights-Updates

Let us remember the mechanics of Multiplicative-Weights-Updates:

- At every time t, the learner maintains a weight vector $w_t \geq 0$ over the experts.
- Given the weight vector, the probability distribution over the experts is computed as $p_t = \frac{w_t}{w_t \cdot \mathbf{1}}$.
- The weights are initialized at $w_1 = \frac{1}{n} \cdot \mathbf{1}$.
- (Multiplicative-weights-update step.) Given the loss vector at time t the weights are updated as follows

  $$w_{t+1}(i) = w_t(i) \cdot u_\beta(l_t(i)), \forall i,$$

  where $u_\beta : [0,1] \to [0,1]$ is an update function satisfying

  $$\beta^x \leq u_\beta(x) \leq 1 - (1-\beta)x, \forall x \in [0,1],$$

  for some $\beta \in [0,1]$.

The reader is free to chose whatever function u_β s/he wants. For example, one can use $u_\beta(x) = \beta^x$, for any $\beta \in [0,1]$. In this case, $w_{t+1}(i) = w_t(i) \cdot \beta^{l_t(i)} = \ldots = w_1(i) \cdot \beta^{L_t(i)} \equiv \frac{1}{n}\beta^{L_t(i)}$.

We can give the following performance guarantee for this algorithm.

Theorem 1. *For all T and any sequence $l_1, l_2, \ldots, l_T$ of loss vectors, the cumulative loss suffered by "Multiplicative-Weights-Updates" satisfies*

$$L_T \leq \frac{\ln(n) + \min_i(L_T(i)) \cdot \ln(\frac{1}{\beta})}{1-\beta}.$$

For example, if we choose $\beta = \frac{1}{2}$, $L_T \leq 2\ln(n) + 2\ln(2) \cdot \min_i(L_T(i))$.

We proceed to give a proof of Theorem 1.

Proof: Let us define a potential function at time t to be $\ln\left(\sum_{i=1}^n w_t(i)\right)$. We have

$$\sum_{i=1}^n w_{t+1}(i) = \sum_{i=1}^n (w_t(i) \cdot u_b(\ell_t(i))) \leq \sum_{i=1}^n (w_t(i)(1-(1-b)\ell_t(i))).$$

We now note that $w_t(i) = p_t(i) \cdot \sum_i w_t(i)$, and hence the right-hand side above is just

$$= \left(\sum_i w_t(i)\right) \cdot \sum_i (p_t(i)(1-(1-b)\ell_t(i))).$$

We now take the natural log of both sides to obtain

$$\ln\left(\sum_{i=1}^{n} w_{t+1}(i)\right) \leq \ln\left(\sum_{i=1}^{n} w_t(i)\right) + \ln\left(1-(1-b)p_t\cdot \ell_t\right).$$

Since $ln(1-x) \leq -x$, we have:

$$\ln\left(\sum_{i=1}^{n} w_{t+1}(i)\right) \leq \ln\left(\sum_{i=1}^{n} w_t(i)\right) - (1-b)p_t\cdot \ell_t.$$

Summing both sides of the inequality from $t=1$ to T and cancelling the terms which appear on both sides yields:

$$\ln\left(\sum_{i=1}^{n} w_{T+1}(i)\right) \leq \ln\left(\sum_{i=1}^{n} w_1(i)\right) - (1-b)L_T \equiv -(1-b)L_T,$$

given that we chose the original weights to be $w_1(i) = \frac{1}{n}$, for all i. Hence:

$$L_T \leq \frac{-\ln\left(\sum_{i=1}^{n} w_{T+1}(i)\right)}{1-b}.$$

By monotonicity of the log function, the previous inequality implies:

$$L_T \leq \frac{-\ln(w_{T+1}(i))}{1-b}, \forall i. \tag{1}$$

We now observe that our update rule $w_{t+1}(i) \leftarrow w_t(i)u_b(\ell_t(i))$ combined with the inequality $u_b(x) \geq b^x$ implies that $w_{T+1}(i) \geq w_1(i)b^{\ell_1(i)}\cdot b^{\ell_2(i)}\cdots b^{\ell_T(i)} \equiv \frac{1}{n}b^{L_T(i)}$. Therefore, (1) implies

$$L_T \leq \frac{-\ln(\frac{1}{n}b^{L_T(i)})}{1-b} = \frac{\ln(n)}{1-b} - \frac{L_T(i)\ln(b)}{1-b}, \forall i.$$

□

Let us experiment a bit with Theorem 1:

- First, if we set $b = 1-\epsilon$ for some $\epsilon \in (0, 1/2)$, the bound becomes

$$L_T \leq (\min_i L_T(i))\frac{\ln(\frac{1}{1-\epsilon})}{\epsilon} + \frac{\ln(n)}{\epsilon}.$$

  which, using the inequality $-\ln(1-z) \leq z + z^2$ for all $z \in (0, 1/2)$, gives

$$L_T \leq \min_i L_T(i)(1+\epsilon) + \frac{\ln(n)}{\epsilon}.$$

- In particular, if we know the time horizon T for which the dynamics will be run in advance, we can set

$$\epsilon = \epsilon_T = \min\left(\sqrt{\frac{\ln(n)}{T}}, \frac{1}{2}\right) \tag{2}$$

  to obtain the bound

$$L_T \leq \min_i L_T(i) + 2\sqrt{T\cdot \ln(n)}.$$

  In this case, the average loss can be bounded by

$$\frac{L_T}{T} - \min_i \frac{L_T(i)}{T} \leq \frac{\sqrt{4\ln(n)}}{\sqrt{T}}.$$

  In other words, *the regret of the learner* goes down with rate $O(\frac{1}{\sqrt{T}})$.

- Even if we do not know the final time horizon T in advance, we can use a "doubling trick" to obtain a similar bound. The idea of the trick is to start agressively choosing ϵ_T for $T = 2$. If the time horizon exceeds 2, we can soften the learning rate by switching to ϵ_T for $T = 4$. If this time horizon is surpassed, we switch to ϵ_T for $T = 8$, and so on.

- Instead of the above "doubling trick," we could also change ϵ in each step. Using $\epsilon_t = \min\left(\sqrt{\frac{\ln(n)}{t}}, \frac{1}{2}\right)$ at time t, we can do a bit better than when using the doubling trick.

2 Importing Learning to Zero-sum Games

We saw that Fictitious Play can be viewed as the result of the two players of a zero-sum game choosing their strategies using the "Follow-the-Leader" rule. What would happen if they employed "Multiplicative-Weights-Updates" instead? In fact, how would they even employ MWU given that we defined the rule for losses in $[0, 1]$?

We start with answering the latter question. According to our completely-uncoupled dynamics assumption, the game $(R, C = -R)_{n\times n}$ is unknown to the players. [1] We assume however that the players know a bound $M > 0$, such that $|R_{ij}| \le M$, for all i, j. Now, given a strategy x of the row player the column player observes a loss vector of $R^{\mathrm{T}}x \in [-M, M]^n$, while given a strategy y of the column player the row player observes a loss vector of $Cy \in [-M, M]^n$. To apply the MWU rule the players apply an affine transformation on these loss vectors, using respectively the loss vectors $\frac{1}{2}\cdot\mathbf{1} + \frac{1}{2M}R^{\mathrm{T}}x$ and $\frac{1}{2}\cdot\mathbf{1} + \frac{1}{2M}Cy$. Here is what this gives:

Theorem 2. *Suppose that the players of a zero sum game $(R, C = -R)_{n\times n}$ choose strategies using the MWU rule, employing an affine transformation on their observed loss vectors as specified above, given an upper bound M on the absolute values of the entries of R. Let $x_1, \ldots, x_T$ and $y_1, \ldots, y_T$ be the sequences of mixed strategies produced by MWU using $\beta = 1 - \epsilon_T$, where ϵ_T is defined as in (2). Then*

$$\frac{1}{T}\sum_t x_t^{\mathrm{T}} R y_t \ge \max_i e_i^{\mathrm{T}} R\left(\frac{1}{T}\sum_t y_t\right) - 4M\sqrt{\frac{\ln n}{T}}; \tag{3}$$

$$\frac{1}{T}\sum_t x_t^{\mathrm{T}} C y_t \ge \max_j \left(\frac{1}{T}\sum_t x_t\right)^{\mathrm{T}} Ce_j - 4M\sqrt{\frac{\ln n}{T}}. \tag{4}$$

Proof: Assigned as an exercise. □

Theorem 3. *Suppose that two sequences of mixed strategies $x_1, \ldots, x_T$ and $y_1, \ldots, y_T$ for the row and column player respectively of a zero-sum game $(R, C = -R)_{n\times n}$ satisfy (3) and (4). Then*

- *the pair of strategies $\left(\frac{1}{T}\sum_t x_t, \frac{1}{T}\sum_t y_t\right)$ is a $8M\sqrt{\frac{\ln n}{T}}$-approximate Nash equilibrium;*

- *$\frac{1}{T}\sum_t x_t^{\mathrm{T}} R x_t$ is within an additive $4M\sqrt{\frac{\ln n}{T}}$ from the row player's value in the game;*

- *$\frac{1}{T}\sum_t x_t^{\mathrm{T}} C y_t$ is within an additive $4M\sqrt{\frac{\ln n}{T}}$ from the column player's value in the game.*

Proof: Let us start by proving the first assertion of the theorem. Equation (4) implies (replacing C by $-R$):

$$-\frac{1}{T}\sum_t x_t^{\mathrm{T}} R y_t \ge -\min_j \left(\frac{1}{T}\sum_t x_t\right)^{\mathrm{T}} Re_j - 4M\sqrt{\frac{\ln n}{T}}.$$

Summing this with (3) we get:

$$\min_j \left(\frac{1}{T}\sum_t x_t\right)^{\mathrm{T}} Re_j \ge \max_i e_i^{\mathrm{T}} R\left(\frac{1}{T}\sum_t y_t\right) - 8M\sqrt{\frac{\ln n}{T}}.$$

[1] The assumption that the number of strategies available to the players is the same is w.l.o.g.

Now observe that

$$\left(\frac{1}{T}\sum_t x_t\right)^{\mathrm{T}} R \left(\frac{1}{T}\sum_t y_t\right) \geq \min_j \left(\frac{1}{T}\sum_t x_t\right)^{\mathrm{T}} Re_j.$$

Put together, the last two inequalities give:

$$\left(\frac{1}{T}\sum_t x_t\right)^{\mathrm{T}} R \left(\frac{1}{T}\sum_t y_t\right) \geq \max_i e_i^{\mathrm{T}} R \left(\frac{1}{T}\sum_t y_t\right) - 8M\sqrt{\frac{\ln n}{T}}.$$

Similarly, we can show

$$\left(\frac{1}{T}\sum_t x_t\right)^{\mathrm{T}} C \left(\frac{1}{T}\sum_t y_t\right) \geq \max_j \left(\frac{1}{T}\sum_t x_t\right)^{\mathrm{T}} Ce_j - 8M\sqrt{\frac{\ln n}{T}}.$$

The above two inequalities imply the first assertion of the theorem.

Let us prove the second assertion of the theorem. The third assertion can be proven similarly. We start by observing that, for all mixed strategies x for the row player the following is true:

$$x^{\mathrm{T}} R \left(\frac{1}{T}\sum_t y_t\right) \geq \min_y x^{\mathrm{T}} Ry.$$

Hence:

$$\max_i e_i^{\mathrm{T}} R \left(\frac{1}{T}\sum_t y_t\right) \equiv \max_x x^{\mathrm{T}} R \left(\frac{1}{T}\sum_t y_t\right) \geq \max_x \min_y x^{\mathrm{T}} Ry.$$

Combining this with (3) we get:

$$\frac{1}{T}\sum_t x_t^{\mathrm{T}} R y_t \geq \max_x \min_y x^{\mathrm{T}} Ry - 4M\sqrt{\frac{\ln n}{T}},$$

which given the min-max theorem establishes the second assertion of the theorem. □

It follows as a corollary of Theorems 2 and 3 that

Corollary 1. *Under the assumptions of Theorem 2*

- *the pair of strategies* $\left(\frac{1}{T}\sum_t x_t, \frac{1}{T}\sum_t y_t\right)$ *is a* $8M\sqrt{\frac{\ln n}{T}}$*-approximate Nash equilibrium;*
- *the average payoff* $\frac{1}{T}\sum_t x_t^{\mathrm{T}} R x_t$ *of the row player is within an additive* $4M\sqrt{\frac{\ln n}{T}}$ *from the row player's value in the game;*
- *the average payoff* $\frac{1}{T}\sum_t x_t^{\mathrm{T}} C y_t$ *of the column player is within an additive* $4M\sqrt{\frac{\ln n}{T}}$ *from the column player's value in the game.*

3 Nash's Theorem

The following theorem was established by John Nash in 1950 [1].

Theorem 4 (Nash)**.** *Every game* $\langle [n], (S_p)_{p\in[n]}, (u_p)_{p\in[n]}\rangle$ *has a Nash equilibrium.*

Before we delve into the proof of this theorem, we need Brouwer's fixed point theorem. This will be proved (in the two-dimensional case) in the next lecture.

Theorem 5 (Brouwer)**.** *Let D be a convex, compact subset of the Euclidean space. If $f : D \to D$ is continuous, then there exists $x \in D$ such that $f(x) = x$.*

The idea behind Nash's proof is to construct a function $f : \Delta \to \Delta$ that satisfies the conditions of Brouwer's fixed point theorem such that the fixed point $\underset{\sim}{x}$ is a Nash equilibrium. To do so, we introduce the idea of a gain function.

Definition 1. *Suppose $\underset{\sim}{x} \in \Delta$ is given. For a player p and strategy $s_p \in S_p$, we define the gain as*

$$\mathrm{Gain}_{p;s_p}(\underset{\sim}{x}) = \max\{u_p(s_p; \underset{\sim}{x}_{-p}) - u_p(\underset{\sim}{x}), 0\}.$$

In other words, the gain is equal to the increase in payoff for player p if he were to switch to pure strategy s_p, unless the increase is negative in which case the gain is taken to equal 0.

Proof of Theorem 1: We define a function $f : \Delta \to \Delta$ as follows. For all $\underset{\sim}{x} \in \Delta$, $\underset{\sim}{x} \overset{f}{\mapsto} \underset{\sim}{y}$ where for all $p \in [n]$ and $s_p \in S_p$:

$$y_p(s_p) := \frac{x_p(s_p) + \mathrm{Gain}_{p;s_p}(\underset{\sim}{x})}{1 + \sum_{s'_p \in S_p} \mathrm{Gain}_{p;s'_p}(\underset{\sim}{x})}.$$

In words, function f tries to boost the probability mass that player p places on various pure strategies depending on the gains in payoff the player would get by switching to these strategies. The denominator just ensures that $\sum_{s_p \in S_p} y_p(s_p) = 1$.

It is easy to see that f is continuous. Moreover, Δ is a product of simplices, so is convex. At the same time, Δ is both closed and bounded, so it is also compact. Hence, Brouwer's fixed point theorem ensures the existence of a fixed point of f.

We claim that any fixed point of f is a Nash equilibrium. To establish this, it suffices to prove that a fixed point $\underset{\sim}{x} = f(\underset{\sim}{x})$ satisfies:

$$\mathrm{Gain}_{p;s_p}(\underset{\sim}{x}) = 0, \quad \forall p \in [n], s_p \in S_p.$$

We proceed by contradiction. Assume that there is some player p who can improve his payoff by switching to pure strategy s_p, i.e.

$$\mathrm{Gain}_{p;s_p}(\underset{\sim}{x}) > 0.$$

First, it is easy to see that $x_p(s_p) > 0$, otherwise $\underset{\sim}{x}$ cannot be a fixed point. Indeed, $x_p(s_p)$ would be 0, while $y_p(s_p)$ would be positive.

Given this, we argue next that there must exist some other pure strategy s''_p such that

$$x_p(s''_p) > 0 \tag{5}$$

and

$$u_p(s''_p; \underset{\sim}{x}_{-p}) - u_p(\underset{\sim}{x}) < 0. \tag{6}$$

Indeed, notice that

$$u_p(\underset{\sim}{x}) \equiv \sum_{s'_p \in S_p} x_p(s'_p) \cdot u_p(s'_p; \underset{\sim}{x}_{-p}).$$

Hence, because $x_p(s_p) > 0$ and $u_p(s_p; \underset{\sim}{x}_{-p}) > u_p(\underset{\sim}{x})$, there must exist some s''_p satisfying (5) and (6).

Now notice that a pure strategy s''_p satisfying (5) and (6) also satisfies $\mathrm{Gain}_{p;s''_p}(\underset{\sim}{x}) = 0$. So:

$$y_p(s''_p) = \frac{x_p(s''_p) + \mathrm{Gain}_{p;s''_p}(\underset{\sim}{x})}{1 + \sum_{s'_p \in S_p} \mathrm{Gain}_{p;s'_p}(\underset{\sim}{x})} < x_p(s''_p),$$

since the numerator is equal to $x_p(s''_p)$, while the denominator is greater than 1 as there is at least one non-zero gain in the summation—the one corresponding to pure strategy s_p. Therefore, $\underset{\sim}{x}$ is not a fixed point, a contradiction. It follows that $\underset{\sim}{x}$ is a Nash equilibrium, as desired. $\square$

References

[1] J. Nash. Equilibrium Points in n-Person Games. *Proceedings of the National Academy of Sciences*, 36(1):48–49, 1950.

[2] J. Nash. Noncooperative Games. *Annals of Mathematics*, 54:289–295, 1951.