6.896 Topics in Algorithmic Game Theory February 10, 2010

Lecture 3

Lecturer: Constantinos Daskalakis *Scribe: Pablo Azar, Anthony Kim*

In the previous lecture we saw that there always exists a Nash equilibrium in two-player zero-sum games. Moreover, the equilibrium enjoys several attractive properties such as polynomial-time tractability, convexity of the equilibrium set, and uniqueness of players' payoffs in all equilibria. In the present and the following lecture we investigate whether simple and natural distributed protocols can find the value/equilibrium strategies of a zero-sum game. We have in mind very generic settings in which the players may be oblivious to the exact specifications of the zero-sum game they are playing. We only require that they know what strategies are available to them, and can observe how well each of their strategies performs against the choices of their opponent.

1 Fictitious Play

Let $(R, C)_{m \times n}$ be a two player zero-sum game. Suppose that the game is played repeatedly by its two players. Say that, at time $t = 0$, the row player plays strategy i_0 and the column player strategy j_0. For any row-player strategy i, define $V^0(i) = R_{i,j_0}$ to represent the payoff achieved by strategy i, given the current history of play by the other player (in this case the history has length 1). Similarly, for any column-player strategy j, define $U^0(j) = R_{i_0,j}$ to represent the *loss* incurred by strategy j given the history of play of the row player.

At time $t = 1$, the players need to decide what strategies to play. Suppose that the players are myopic and make their decisions greedily based on the current history of play. A myopic/greedy row player would choose some strategy

$$i_1 \in argmax_i V^0(i),$$

while a myopic/greedy column-player would choose some strategy

$$j_1 \in argmin_j U^0(j).$$

Given these strategies, the cumulative payoff and loss vectors should be updated as follows:

$$V^1(i) = V^0(i) + R_{i,j_1},$$
$$U^1(j) = U^0(j) + R_{i_1,j}.$$

At an arbitrary time t, assume that we are given the cumulative payoff and loss vectors up to that time; that is, we are given V^{t-1}, U^{t-1}. *Fictitious play* specifies that the choices made by the row and column player at time t satisfy respectively

$$i_t \in argmax_i V^{t-1}(i),$$
$$j_t \in argmin_j U^{t-1}(j).$$

Given i_t, j_t, we can update the payoff and loss vectors accordingly

$$V^t(i) = V^{t-1}(i) + R_{i,j_t},$$
$$U^t(j) = U^{t-1}(j) + R_{i_t,j}.$$

And the dynamics proceed ad infinitum or until some fixed period of time T is exhausted.

Example 1. *Let (R, C) be a two-player zero-sum game with three strategies per player. Suppose that the row player's payoffs are given by*

$$R = \begin{pmatrix} 2 & 1 & 0 \\ 2 & 0 & 3 \\ -1 & 3 & -3 \end{pmatrix}.$$

Suppose that at time $t = 0$ the row player plays $i_0 = 1$ and the column player plays $j_0 = 3$. Table 1 summarizes the first three rounds of the game if the players follow fictitious play.

t	i_t	j_t	$V^t(1)$	$V^t(2)$	$V^t(3)$	$U^t(1)$	$U^t(2)$	$U^t(3)$
0	1	3	0	$\underline{3}$	-3	2	1	$\underline{0}$
1	2	3	0	$\underline{6}$	-6	4	$\underline{1}$	3
2	2	2	1	$\underline{6}$	-3	6	$\underline{1}$	6

Table 1: Summary of first three rounds of game. Underlined numbers indicate optimal cumulative rewards/loesses for a given round by the two players of the game.

Observe that, in the first three rounds of the game shown in this table, $max_i V^t(i) \geq min_j U^t(j)$. Does this hold for all two player zero-sum games and for all times t? We show that this indeed is the case. We begin by defining the row player's *empirical mixed strategy*

$$x^t = \frac{1}{t+1} \sum_{\tau \leq t} e_{i_\tau},$$

where i_τ is the strategy played at time τ and e_i is a vector whose components are all zero, except for the i^{th} component, which is 1. Similarly, the column player's empirical mixed strategy is

$$y^t = \frac{1}{t+1} \sum_{\tau \leq t} e_{j_\tau}.$$

Given this definition, we can write the row player's payoff vector $V^t = (V^t(1), ..., V^t(m))$ as

$$V^t = \sum_{\tau \leq t} R \cdot e_{j_\tau} = (t+1) \cdot R \cdot y^t.$$

Similarly, we can write the column player's loss vector $U^t = (U^t(1), ..., U^t(n))$ as

$$U^t = \sum_{\tau \leq t} e_{i_\tau}^T R = (t+1) \cdot (x^t)^T R.$$

We can show the following.

Claim 1. *If a zero-sum game (R, C) is played repeatedly by two players following fictitious play, then for all times $t \geq 0$:*

$$max_i \frac{V^t(i)}{t+1} \geq v \geq min_j \frac{U^t(j)}{t+1},$$

where v is the value of the game.

Proof: It suffices to show that for all t:

$$\max R \cdot y^t \geq v \geq \min (x^t)^T R,$$

where the max and min operators pick the maximum, respectively minimum, coordinate values of their operand vectors.

Recall the linear program $LP(2)$ from the previous lecture:

$$\begin{aligned} & \min z \\ s.t. \quad & Ry \leq z \cdot \mathbb{1} \\ & \sum y_i = 1, y_i \geq 0. \end{aligned}$$

In every optimal solution (y^*, z^*) of this linear program, at least one of the slack constraints must be tight. So we get $z^* = \max(R \cdot y^*)$. We also argued in the previous lecture that the optimal value z^* of this LP is equal to the value v of the game.

Now notice that $(y^t, \max(R \cdot y^t))$ is always a feasible solution of this linear program achieving value $\max(R \cdot y^t)$. Since the linear program is a minimization problem, we must have $max(R \cdot y^t) \geq z^* = v$. Similarly, we can argue using $LP(1)$ that $v \geq \min((x^t)^T \cdot R)$. This concludes the proof. $\square$

1.1 Convergence of Fictitious Play

The above result gives an interesting property of fictitious play, namely that the maximum payoff that the row player can achieve against the empirical strategy of the column player is larger than the value of the game, which in turn is larger than the minimum loss that the column player could achieve against the empirical strategy of the row player. Do these values converge to the value of the game? And, do the empirical strategies converge to an equilibrium of the game? Julia Robinson showed that the max payoff and min loss converge to the value of the game.

Theorem 1 (J. Robinson, 1950 [1])**.** *If a zero-sum game (R, C) is played repeatedly by two players following fictitious play, then:*

$$\lim_{t\to\infty} \max \frac{V^t}{t+1} = \lim_{t\to\infty} \min \frac{U^t}{t+1} = v,$$

where v is the value of the game.

Discussion:

- Robinson's proof is a clever inductive argument on the number of strategies of the game. We do not provide the proof here, but encourage the interested reader to look at it [1].
- It is a priori not clear that the above limits exist. So in particular the above theorem informs us that these limits do exist.
- Robinson's proof does not discuss the speed of convergence to the value of the game. Unraveling her inductive argument we can establish the following.

  Theorem 2. *For all $\epsilon > 0$, for all $t \geq (\frac{R_{max}}{\epsilon})^{\Omega(m+n)}$ we have*

  $$\left| \max \frac{V^t}{t+1} - \min \frac{U^t}{t+1} \right| \leq \epsilon,$$

  where $R_{max} = max_{i,j}(|R_{ij}|)$, and m, n are respectively the number of rows and columns in the payoff matrices of the game.

And what about the empirical mixed strategies, do they converge to some interesting object? Before discussing this, let us define the notion of an approximate Nash equilibrium.

Definition 1. *A pair of strategies is an ϵ-approximate Nash Equilibrium if and only if*

1. $x^T R y \geq x'^T R y - \epsilon$ *for all* $x' \in \Delta_m$,
2. $x^T R y \geq x^T R y' - \epsilon$ *for all* $y' \in \Delta_n$.

That is, no player of the game can improve by more than ϵ by switching to a different mixed strategy.

We obtain the following corollary of Theorem 2, showing that the empirical strategies constitute an ϵ-approximate Nash equilibrium for all t large enough.

Corollary 1. *For all $\epsilon > 0$, for all $t \geq (\frac{R_{max}}{\epsilon})^{\Omega(m+n)}$, (x^t, y^t) is an ϵ-approximate Nash equilibrium of the game.*

Proof: Writing $\frac{V^t}{t+1}$ as Ry^t and $\frac{U^t}{t+1}$ as $(x^t)^T R$, we get from Theorem 2 that

$$0 \leq \max Ry^t - \min (x^t)^T R \leq \epsilon.$$

But note that $\min (x^t)^T R \leq (x^t)^T R y^t$. The reason is that the right hand side can be interpreted as an average of the coordinate-values of $(x^t)^T R$. This average is always greater than or equal to the minimum coordinate value of $(x^t)^T R$.

Summing the two inequalities above, we get

$$\max Ry^t - (x^t)^T Ry^t \leq \epsilon$$

$$\Leftrightarrow (x^t)^T Ry^t \geq \max Ry^t - \epsilon.$$

That is, if the column player uses her empirical mixed stretegy y^t, the row player cannot improve her payoff by more than ϵ by not using his empirical mixed strategy x^t. We can reason analogously to show that the column player performs cannot improve by more than ϵ by deviating from y^t. This establishes that the pair (x^t, y^t) is an ϵ-approximate Nash equilibrium. □

In other words, if we consider the empirical mixed strategies resulting from fictitious play at any time t, these constitute an $R_{max} \cdot t^{-\frac{1}{\Omega(m+n)}}$- approximate Nash equilibrium. Can convergence be made faster? Samuel Karlin conjectured so...

Conjecture 1 (Samuel Karlin, 1959 [2])**.** *Fictitious play converges with rate* $\frac{1}{\sqrt{t}} \cdot f(|R|, |C|)$*, for some function* f *that only depends on the description size of* R *and* C*.*

2 Learning under Expert Advice

We temporarily postpone our study of games, and switch contexts to optimization under expert advice. We will come back to zero-sum games in the next lecture. The setup we consider here is the following:

- n experts/strategies are available to a learner
- At every time t:
  - The learner chooses a probability distribution over $[n]$: $\underset{\sim}{p}^t$.
  - After the learner makes his choice, nature or an adversary outputs a loss vector suffered by the experts $\underset{\sim}{l}^t \in [0,1]^n$. (N.B. our limitation to $[0,1]$ is benign since we can always apply an affine transformation to bring the losses to $[0,1]$, as long as the losses are bounded)
  - The learner's loss in this round is $\underset{\sim}{p}^t \cdot \underset{\sim}{l}^t$.
- The learner's cumulative loss up to time t is $L^t = \sum_{\tau \leq t} \underset{\sim}{p}^\tau \cdot \underset{\sim}{l}^\tau$.

Our goal is to devise an algorithm for the learner so as to minimize the cumulative loss, L^t. But, what benchmark should we compare our algorithm's performance against? One possibility is $\sum_{\tau \leq t} \min(\underset{\sim}{l}^\tau)$. This is exactly the best we could do, if we knew the future. We argue that this is too ambitious. Indeed, the adversary could observe the learner's choice $\underset{\sim}{p}^t$ before deciding $\underset{\sim}{l}^t$. Then she could give loss of 1 to all experts in the support of $\underset{\sim}{p}^t$, except for the expert with the smallest probability in $\underset{\sim}{p}^t$ to which she would give loss of 0. The learner's loss would grow linearly in time, while the benchmark would get payoff of 0.

It turns out that a more reasonable benchmark to compare against is the best performing expert, incurring loss of $\min(\sum_{\tau \leq t} \underset{\sim}{l}^\tau)$. Below we consider a couple of learning algorithms.

2.1 "Follow the Leader"

Maybe the simplest strategy for the learner is to pick the strategy that has performed the best so far. This is called the "Follow the Leader" algorithm, whose outline is given below:

- Let $L_i^t = \sum_{\tau \leq t} l_i^\tau$.
- At time t, pick some expert in $\text{argmin}_i \, L_i^{t-1}$.

The following example shows that the performance of this algorithm can be poor.

Example 2. *In the table below, the rows are indexed by the n strategies available to the learner and the columns are indexed by the time step $t = 1, 2, \ldots$. Each column i represents the loss vector $\underset{\sim}{l}^t$ at time $t = i$. The empty cells of the table should be interpreted as carrying loss of* 0.

	$t=1$	$t=2$	$t=3$	$\cdots$	$t=n+1$
1	$\frac{1}{n}$	*1*			
2	$\frac{1}{n-1}$		*1*		
3	$\frac{1}{n-2}$				
$\vdots$	$\vdots$			$\ddots$	
n	*1*				*1*

After $n+1$ steps, the loss of our algorithm is $L^{n+1} = L^1 + n$, while $\min(\sum_{\tau \le n+1} \underset{\sim}{l}^\tau) = 1 + \frac{1}{n}$.

It looks then that the cumulative loss by the "Follow the Leader" algorithm can be at least about n times larger than the benchmark $\min(\sum_{\tau \le t} \underset{\sim}{l}^\tau)$. In fact, this is essentially the worst possible performance by this algorithm.

Theorem 3. *For all t,*

$$L^t \le n \cdot (\min_i L_i^t + 1).$$

Proof: Assigned as an exercise problem for 2 points. □

2.2 Hedging (a.k.a. Multiplicative Weights Update Method)

Instead of picking a single strategy deterministically as in the "Follow the Leader" algorithm, wouldn't it be a better idea to spread risk across the various experts depending on their performance? This is the motivation behind the hedging algorithm described next:

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

  $$w_i^{t+1} = w_i^t \cdot u_\beta(l_i^t), \forall i$$

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

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

For example, we can use $u_\beta(x) = \beta^x$. In this case, $w_i^{t+1} = w_i^t \cdot \beta^{l_i^t} = \ldots = w_i^1 \cdot \beta^{L_i^t}$.

We can give the following performance guarantee for this algorithm.

Theorem 4. *For all t and any sequence $\underset{\sim}{l}^1, \underset{\sim}{l}^2, \ldots, \underset{\sim}{l}^t$,*

$$L^t \le \frac{\ln(n) + \min_i(L_i^t) \cdot \ln(\frac{1}{\beta})}{1-\beta}.$$

For example, if we choose $\beta = \frac{1}{2}$, $L^t \le 2\ln(n) + 2\ln(2) \cdot \min_i(L_i^t)$. We show Theorem 4 in the next lecture.

3 Homework [2 points]

Prove Theorem 3.

References

[1] Julia Robinson. An iterative method of solving a game. *The Annals of Mathematics*, 54(2):296–301, 1951.

[2] Samuel Karlin. *Mathematical Methods and Theory in Games, Programming & Economics*. Addison-Wesley, 1959.