6.853 Topics in Algorithmic Game Theory September 20, 2011

Lecture 4

Lecturer: Constantinos Daskalakis *Scribe: Constantinos Daskalakis*

We have seen that Nash equilibria in two-player zero-sum games (and generalizations thereof) are polynomial-time tractable from a centralized computation perspective. We have also seen that the payoff matrix of a zero-sum game determines a unique value for the row player and a unique value for the column player (summing to zero), which specify their payoffs in *all* equilibria of the game. In this lecture, we investigate whether Nash equilibria can arise as a result of the *distributed interaction* between the players of a zero-sum game, and whether the values of the players in the game are descriptive of their long-term payoffs in the course of their interaction.

Clearly, if the players are aware of the details of the game (i.e. the game's payoff matrix), they can compute their min-max strategies on the side and just use these strategies forever. We envision a much weaker distributed scenario, of *completely-uncoupled dynamics* as follows:

- each player knows her own pure strategies, but does not know the game matrix, or even the number of strategies available to her opponent;
- players interact in rounds, and each player can choose a mixed strategy in each round;
- in the end of each round, each player is informed about the expected payoff she would have gotten had she played each of her pure strategies against the opponent's mixed strategy (but the mixed strategy of the opponent is not revealed to her).

1 Fictitious Play

We consider a type of completely-uncoupled dynamics called *fictitious play*. Fictitious play was defined by George W. Brown [1] who conjectured its convergence to the value of a zero-sum game, and its convergence properties were established by Julia Robinson [6] (yes, the same Julia Robinson of Hilbert's tenth problem [5]). We proceed to describe how fictitious play works. Let $(R, C = -R)_{m\times n}$ be a two player zero-sum game, but assume we are in a completely-uncoupled scenario where the players are ignorant of the game matrix. The players interact in rounds as follows:

- In round $t = 1$:
  - the row player plays an arbitrary strategy i_1 and the column player plays an arbitrary strategy j_1;
  - the row player observes Re_{j_1} and the column player observes $e_{i_1}^{\mathrm{T}} C$;
- In round $t = 2$:
  - the row player plays any strategy $i_2 \in \mathrm{argmax}_i\{e_i^{\mathrm{T}} Re_{j_1}\}$, and the column players plays any strategy $j_2 \in \mathrm{argmax}_j\{e_{i_1}^{\mathrm{T}} Ce_j\}$;
  - the row player observes Re_{j_2} and the column player observes $e_{i_2}^{\mathrm{T}} C$;
- ...
- In a general round t:
  - the row player plays any strategy

$$i_t \in \mathrm{argmax}_i \left\{ e_i^{\mathrm{T}} R \left(\frac{1}{t-1} \sum_{\tau \le t-1} e_{j_\tau} \right) \right\},$$

and the column player plays any strategy

$$j_t \in \text{argmax}_j \left\{ \left(\frac{1}{t-1} \sum_{\tau \leq t-1} e_{i_\tau} \right)^{\mathrm{T}} C e_j \right\};$$

*/*observe that i_t and j_t can be selected using information that the row and column player has observed, in particular average payoff vectors from previous rounds*/*

– the row player observes Re_{j_t} and the column player observes $e_{i_t}^{\mathrm{T}} C$;

• ...

For notational convenience in what follows we denote by $x_t = \frac{1}{t}(\sum_{\tau \leq t} e_{i_\tau})$ and $y_t = \frac{1}{t}(\sum_{\tau \leq t} e_{j_\tau})$ the *empirical, or historical, strategies* played by the row and column players respectively in the course of the dynamics. In a nutshell,

Definition 1. Fictitious play *is the completely uncoupled player interaction in which in every round each player plays a best response to the opponent's historical strategy as described above.*

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 $t = 1$ the row player plays $i_1 = 1$ and the column player plays $j_1 = 3$. Table 1 summarizes the first three rounds of fictitious play.

t	i_t	j_t	$te_1^{\mathrm{T}} R y_t$	$te_2^{\mathrm{T}} R y_t$	$te_3^{\mathrm{T}} R y_t$	$tx_t^{\mathrm{T}} C e_1$	$tx_t^{\mathrm{T}} C e_2$	$tx_t^{\mathrm{T}} C e_3$
1	1	3	0	<u>3</u>	-3	-2	-1	<u>0</u>
2	2	3	0	<u>6</u>	-6	-4	<u>-1</u>	-3
3	2	2	1	<u>6</u>	-3	-6	<u>-1</u>	-6

Table 1: Summary of the first three rounds of fictitious play. Underlined numbers indicate optimal cumulative pyoffs for a given round by each player of the game.

It is easy to establish the following:

Claim 1. *If the players of a zero-sum game $(R, C = -R)$ interact via fictitious play, then for all times $t \geq 1$:*

$$\max_i e_i^{\mathrm{T}} R y_t \geq v \geq \min_j x_t^{\mathrm{T}} R e_j,$$

where v is the value of the row player in the game.

Proof: The proof follows easily from the min-max theorem. Recall the linear program $LP(2)$ from Lecture 2:

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

In every optimal solution (y^*, z^*) of this linear program, at least one of the slack constraints must be tight. So we get $z^* = \max_i(e_i^{\mathrm{T}} R \cdot y^*)$. 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 $(y_t, \max_i(e_i^{\mathrm{T}} R \cdot y_t))$ is always a feasible solution of this linear program achieving value $\max_i(e_i^{\mathrm{T}} R \cdot y_t)$. Since the linear program is a minimization problem, we must have $\max_i(e_i^{\mathrm{T}} R \cdot y^t) \geq z^* = v$. Similarly, we can argue using $LP(1)$ of Lecture 2 that $v \geq \min_j(x_t^T \cdot Re_j)$. This concludes the proof. □

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 suffer 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 [6] showed that the answer to these questions is positive, namely

Theorem 1 (J. Robinson [6])**.** *If the players of a zero-sum game $(R, C = -R)$ interact via fictitious play, then:*

$$\lim_{t\to\infty} \max_i e_i^{\mathrm{T}} R y_t = \lim_{t\to\infty} \min_j x_t^{\mathrm{T}} R e_j = v,$$

where v is the value of the row player in 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 [6].
- 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 $t \geq \left(\frac{R_{max}}{\epsilon}\right)^{\Omega(m+n)}$ we have*

  $$\left| \max_i e_i^{\mathrm{T}} R y_t - \min_j x_t^{\mathrm{T}} R e_j \right| \leq \epsilon,$$

  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.

- Finally, there is nothing special about the row player; we can obviously state an analogous result for the column player.

And what about the empirical mixed strategies, do they also converge to some interesting object? Before discussing this, let us recall the notion of an ϵ-approximate Nash equilibrium from the previous lecture.

Definition 2. *A pair of mixed strategies (x, y) for the players of a two-player game $(R, C)_{m\times n}$ is an ϵ-approximate Nash Equilibrium if and only if*

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

That is, no player of the game can improve by more than an additive 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: Claim 1 and Theorem 2 imply that

$$0 \leq \max_i e_i^{\mathrm{T}} R y_t - \min_j x_t^{\mathrm{T}} R e_j \leq \epsilon.$$

But note that $\min_j x_t^{\mathrm{T}} R e_j \leq x_t^{\mathrm{T}} R y_t$. The reason is that the right hand side can be interpreted as a convex combination of the coordinates of $x_t^{\mathrm{T}} R$, and this convex combination must be at least as large as the minimum coordinate. Summing these inequalities we get

$$\max_i e_i^{\mathrm{T}} R y_t - x_t^{\mathrm{T}} R y_t \leq \epsilon$$

$$\Leftrightarrow x_t^{\mathrm{T}} R y_t \geq \max_i e_i^{\mathrm{T}} R y_t - \epsilon.$$

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. Similarly, we can argue that the column player 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 the players of a zero-sum game interact via fictitious play, then their empirical mixed strategies at round t constitute a $\left(R_{max} \cdot t^{-\frac{1}{O(m+n)}}\right)$- approximate Nash equilibrium. Can convergence be made faster? Samuel Karlin conjectured so...

Conjecture 1 (Samuel Karlin, 1959 [3])**.** *Fictitious play converges with rate* $\frac{1}{\sqrt{t}} \cdot f(|R|)$*, for some function* $f(|R|)$ *of the description complexity of the game matrix* R*.*

If the conjecture were true then, for all $\epsilon > 0$, the empirical strategies computed by fictitious play after time $t \geq \frac{1}{\epsilon^2} f^2(|R|)$ would constitute an ϵ-approximate Nash equilibrium of the game.

2 A Detour: Learning from Expert Advice

We temporarily postpone our study of games, switching contexts to optimization against an unknown future using expert advice. We 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; identify them with the elements of $[n] := \{1, \ldots, n\}$.
- At every time t:
  - The learner chooses a probability distribution over the experts $[n]$: p_t.
  - After the learner makes his choice, nature or an adversary outputs a loss vector suffered by the experts $l_t \in [0,1]^n$. (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 $p_t \cdot l_t$.
- The learner's cumulative loss up to time t is $L_t = \sum_{\tau \leq t} p_\tau \cdot l_\tau$.

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 $\sum_{\tau \leq t} \min_i(l_\tau(i))$. This is exactly the best we could do, if we knew the future. We argue that this is too ambitious. Indeed, as we have said, the adversary can in principle observe the learner's choice p_t before deciding l_t. Hence, she could give loss of 1 to all experts in the support of p_t, except for the expert with the smallest probability in p_t to which she would give loss of 0 (breaking ties arbitrarily). The learner's loss would grow linearly with time, while the benchmark $\sum_{\tau \leq t} \min_i(l_\tau(i))$ would remain 0.

It turns out that a more reasonable benchmark to compare against is the best fixed expert, incurring loss of $\min_i(\sum_{\tau \leq t} l_\tau(i))$. Below we consider a couple of learning algorithms, comparing them against this milder benchmark.

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 rule for choosing experts is called "Follow-the-Leader", formally defined as follows:

- For every expert i, let $L_t(i) = \sum_{\tau \leq t} l_\tau(i)$ be its cumulative loss up until time t, inclusive. (In these terms, we ultimately want to compare our cumulative loss L_t to $\min_i L_t(i)$.)
- At time t, pick some expert in $\operatorname{argmin}_i L_{t-1}(i)$, i.e. one of the best performing experts given the cumulative loss vectors observed so far.

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 t represents the loss vector l_t at time t. 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*

At time $t = n + 1$, the loss of "Follow the leader" is $L_{n+1} = L_1 + n$, while the benchmark achieves loss $\min_i(L_t(i)) = 1 + \frac{1}{n}$.

It looks then that the cumulative loss of "Follow the Leader" can be at least about n times larger than the benchmark $\min_i(L_t(i))$. In fact, this is essentially the worst possible performance by this algorithm.

Theorem 3. *For all t, "Follow-the-Leader" achieves*

$$L_t \leq n \cdot (\min_i L_t(i) + 1).$$

Proof: Assigned as an exercise problem. □

Remark 1. *Observe (exercise) that fictitious play can be viewed equivalently as the result of the two-players of a zero-sum using the "Follow-the-Leader" protocol to update their strategies.*

2.2 Hedging Method (a.k.a. "Multiplicative-Weights-Updates")

Instead of picking a single expert deterministically as in "Follow-the-Leader", wouldn't it be a better idea to spread risk across the various experts depending on their performance? This is the motivation behind the multiplicative-weights-update method, developed by Littlestone and Warmuth [4] and first used in a game-theoretic context by Freund and Schapire [2]. The method is described next:

- 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 4. *For all t and any sequence $l_1, l_2, \ldots, l_t$ of loss vectors,*

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

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 show Theorem 4 in the next lecture.

References

[1] George W. Brown. Some notes on computation of Games Solutions. *RAND Corporation Report P-78,* April 1949.

[2] Yoav Freund and Robert E. Schapire. Adaptive game playing using multiplicative weights. *Games and Economic Behavior,* 29(1-2):79–103, 1999.

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

[4] Nick Littlestone and Manfred K. Warmuth. The weighted majority algorithm. *Information and Computation,* 108(2):212–261, 1994.

[5] Yuri V. Matiyasevich. *Hilbert's Tenth Problem.* MIT Press, Cambridge, Massachusetts, 1993.

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