Optimization II **Winter 2009/10**

Lecture 2 — October 19

Lecturer: Khaled Elbassioni

2.1 ε-approximation of 2-player zero-sum games

In this lecture we give a randomized "fictitious play" algorithm for obtaining an approximate solution for 2-player zero-sum games.

2.1.1 matrix games

A *2-player zero-sum game*, or a *matrix game*, is defined by a matrix $A \in \mathbb{R}^{m \times n}$, called the *payoff* matrix. There are two players with opposing interests: the row player (minimizer) and the column player (maximizer). At each step of the play, the row player selects a row i, and the column player selects a column j, then the row player pays to the column player the value a_{ij} of the (i, j)th entry in the matrix. Suppose that the play continues forever, how to play such a game?

Example 1. *Consider the payoff matrix*

$$A = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}.$$

Let us see what happens if the row player chooses row 1. Then, the column player (being a maximizer) will choose the first column. But then the row player (being a minimizer) will switch to row 3. Then, the column player will find it more profitable to switch to column 3, after which the row player will switch to row 1, resulting in a cycle!

This situation describes what is not an equilibrium. *Examining the above matrix, the row maxima are* 1, 0, *and* 1, *respectively. So if the row player chooses row 1, the column player would guarantee 1, if the row player chooses row 2, then the column player would guarantee 0, and so on. In general, if the row player chooses row i, then the column player would guarantee* $\max_j a_{ij}$, *and thus the row player should choose the row that minimizes this maximum. Similarly, since the row minima are* $-1, 0, -1$, *respectively, the column player can guarantee* $\max_j \min_i a_{ij} = 0$. *Since it happens in this example that these two values are equal, there will be an equilibrium if the row player sticks to playing the 2nd row and the column player sticks to playing the 2nd column.*

An *equilibrium* or a *saddle point* is a pair of strategies for the two players such that no player has incentive to switch, assuming that the other player does not switch. But is there always a saddle-point in *pure strategies* as in Example 1. The answer is NO as the following well-known example shows.

Example 2. *Consider the payoff matrix*

$$A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

Then $\min_i \max_j a_{ij} = 1 \neq -1 = \max_j \min_i a_{ij}$.

So what to do? Play with mixed strategies. That is, the row player chooses a probability vector $x \in S_m = \{x \in \mathbb{R}^m : \ e^T x = 1, \ x \geq 0\}$, where e denotes the vector of all ones of the appropriate dimension, and plays row i with probability x_i. Similarly, the column vector plays according to a probability vector $y \in S_n = \{y \in \mathbb{R}^n : \ e^T y = 1, \ y \geq 0\}$. Let us denote by $A_1, \ldots, A_m$ the rows of A, by $A^1, \ldots, A^m$ the columns of A, and by e_i the i-unit vector with the appropriate dimension. Then the expected value that the row player would pay if she decided to play row i is $\sum_j a_{ij} y_j = A_i y = e_i^T A y$, and hence her expected payoff would be $\sum_i x_i A_i y = x^T A y$. Similarly, the expected payoff of the column player is $x^T A y$. For instance, in Example 2 above, if both players choose $(\frac{1}{2}, \frac{1}{2})$ as their strategy, then the expected payoff for both is 0. On the other hand, if the row player chooses $x = (\frac{1}{3}, \frac{2}{3})$, while the column player chooses $y = (\frac{2}{3}, \frac{1}{3})$, the then payoff is $-\frac{1}{9}$. Is any of these two pairs of strategies an equilibrium? And does such an equilibrium exists in general? The answer is YES as given by the following Theorem.

Theorem 2.1 (Von Neumann (1928)). *For any matrix $A \in \mathbb{R}^{m \times n}$,*

$$\min_{x \in S_m} \max_{y \in S_n} x^T A y = \max_{y \in S_n} \min_{x \in S_m} x^T A y. \tag{2.1}$$

Definition 2.2 (Saddle point). *A saddle point in a matrix game with payoff matrix $A \in \mathbb{R}^{m \times n}$, is a pair of strategies $x^* \in S_m$ and $y^* \in S_n$ such that*

$$min_{x \in S_m} x^T A y^* = \max_{y \in S_n} (x^*)^T A y. \tag{2.2}$$

Such a pair will be also called an optimal pair.

Exercise 1. *(i) Show that $\min_{x \in S_m} \max_{y \in S_n} x^T A y \geq \max_{y \in S_n} \min_{x \in S_m} x^T A y$.*

(ii) Show that a pair of strategies $x^* \in S_m$ and $y^* \in S_n$ are optimal if and only if for all i, j: $(x^*)^T A^j \leq A_i y^*$.

It is worth noting that a matrix game is equivalent to a pair of *packing-covering* linear programs (LP's).

Exercise 2. *Let v^* be the common value in the identity (2.1). Show that*

$$v^* = \min\{v \mid x^T A \leq v e^T, \ x \in S_m\} = \max\{v \mid Ay \geq ve, \ y \in S_n\}.$$

Let $\varepsilon \in [0, 1]$ be a given constant. We are interested in ε-optimal strategies, defined as follows.

Definition 2.3 (ε-optimal strategies). *A pair of strategies $x^* \in S_m$ and $y^* \in S_n$ is an ε-optimal pair for a matrix game with payoff matrix $A \in \mathbb{R}^{m \times n}$ if*

$$\max_{y \in S_n} (x^*)^T A y \leq \min_{x \in S_m} x^T A y^* + \varepsilon. \tag{2.3}$$

In this lecture, we consider the problem of finding approximate saddle points of matrix games.

ϵ-APPROXIMATION OF ZERO-SUM GAMES

Input: A matrix $A \in \mathbb{R}^{m \times n}$ and a desired accuracy ε

Output: A pair of strategies $x \in S_m$ and $y \in S_n$

Objective: x, y are ε-equilibria

2.1.2 fictitious play

Fictitious play is a method suggested by Brown in 1951 [Bro51] to obtain a saddle point for a given matrix game. Iteratively, the minimizer and the maximizer maintain in $X(t) \in \mathbb{Z}_+^m$ and $Y(t) \in \mathbb{Z}_+^n$ the frequencies by which rows and columns have been used, respectively, upto time t of the play. Then each player updates his/her strategy by applying the best response, given the current opponent's strategy. The procedure is given below.

Algorithm 1 FICTITIOUS PLAY(A)

1. $X(0) := 0$ and $Y(0) := 0$
2. **for** $t = 1, 2, \ldots$ **do**
3. $\quad i := \text{argmin}\{A_1 Y(t-1), \ldots, A_m Y(t-1)\}$; $X(t) := X(t-1) + e_i$
4. $\quad j := \text{argmax}\{X(t-1)^T A^1, \ldots, X(t-1)^T A^n\}$; $Y(t) := Y(t-1) + e_j$

Note that at each t, the vectors $\frac{X(t)}{t}$ and $\frac{Y(t)}{t}$ are feasible strategies. The convergence of such pair of strategies, $x^* = \lim_{t\to\infty} \frac{X(t)}{t}$, $y^* = \lim_{t\to\infty} \frac{Y(t)}{t}$, was established by Robinson [Rob51]. A bound of $\left(\frac{2^{m+n}\rho}{\varepsilon}\right)^{m+n-2}$, where $\rho = \max_{i,j} |a_{ij}|$, on the time needed for convergence to an ε-pair was obtained by Shapiro in 1958 [Sha58]. The tendency in the literature is to believe that this time is bounded by $O(\frac{\text{poly}(n,m)}{\varepsilon^2})$. A smoothed version of this fictitious play, introduced in the next section, archives such a bound.

2.1.3 Randomized fictitious play

Grigoriadis and Khachiyan (1995) [GK95] introduced a randomized version of fictitious play, in which the argmin and argmax operators in steps 3 and 4 above are replaced by a smoothed selection which picks a row (respectively, a column) with probability that decreases (respectively, increases) quickly as the current response of the opponent to this row (respectively, column) increases. More precisely, given the current vectors of frequencies $X(t) \in \mathbb{Z}_+^m$ and $Y(t) \in \mathbb{Z}_+^n$, the row and column players choose, respectively, a row i and a column j according to the (so-called *Gibbs*) distributions:

$$\frac{p_i(t)}{|p(t)|} \text{ where } p_i(t) = e^{-\frac{\varepsilon A_i Y(t-1)}{2}} \quad \text{and } |p(t)| = \sum_{i=1}^m p_i(t) \tag{2.4}$$

$$\frac{q_j(t)}{|q(t)|} \text{ where } q_j(t) = e^{\frac{\varepsilon X(t-1)^T A^j}{2}} \quad \text{and } |q(t)| = \sum_{j=1}^n q_j(t). \tag{2.5}$$

Here is the algorithm. This will be the theme of most of the algorithms described in the lectures on packing and covering LP's.

Algorithm 2 RANDOMIZED FICTITIOUS PLAY(A)

---

1. $X(0) := 0$ and $Y(0) := 0$
2. **for** $t = 1, 2, \ldots, T \stackrel{\text{def}}{=} \frac{6\ln(2nm)}{\varepsilon^2}$ **do**
3. Pick $i \in [m]$ and $j \in [n]$ independently, with probabilities $\frac{p_i(t)}{|p(t)|}$ and $\frac{q_j(t)}{|q(t)|}$, respectively
4. $X(t) := X(t-1) + e_i$; $Y(t) := Y(t-1) + e_j$
5. **return** $(\frac{X(t)}{t}, \frac{Y(t)}{t})$

It is the smoothing step in line 3 that makes it possible to prove better bounds on the number of iterations than those currently known for deterministic fictitious play.

The analysis, here and in all the algorithms considered in subsequent lectures, will follow more or less the same framework: we define a potential function

$$\Phi(t) = |p(t+1)||q(t+1)|, \tag{2.6}$$

and show that it does not increase by much from one iteration to the next. Then this implies, by iterating, that the expected potential after t time steps is bounded by the initial potential multiplied by some factor, which might depend exponentially on t. Since the initial potential is the sum of some non-negative exponentials, each term in the sum is bounded by the final potential. Taking logs allows us to bound the error in approximation at time t, as a function of t, and our choice of the terminating time T, when plugged in this function, guarantees to make the error less than ε as desired. The proof we give here uses ideas from Grigoriadis and Khachiyan (1995) [GK95] and Koufogiannakis and Young [KY07].

For the purpose of obtaining an approximation with an absolute error, we will assume that all the entries of the matrix A are in some fixed range, say $[-1, 1]$. Scaling the matrix A by $\frac{1}{\rho}$, where the "width" parameter ρ is defined as $\rho = \max_{i,j} |a_{ij}|$, and replacing ε by $\frac{\varepsilon}{\rho}$ in what follows, we get an algorithm that works without this assumption, but whose running time is proportional to ρ^2. We note that such dependence on the width is unavoidable in all known algorithms that obtain ε-approximate solutions and whose running time is proportional to poly$(\frac{1}{\varepsilon})$. An exception is when A is non-negative in which this dependence can be removed as we shall see in a later lecture.

Exercise 3. *Show that any matrix game (2.1) can be converted into an equivalent one in which each entry in the matrix A is in $[a, b]$, where $a, b \in \mathbb{R}$. Does the same reduction work if we are aiming at an ϵ-approximate saddle point?*

Theorem 2.4. *Assuming $A \in [-1, 1]^{m\times n}$, algorithm* RANDOMIZED FICTITIOUS PLAY *outputs ε-optimal strategies. The total running time is $O(\frac{(n+m)\log(n+m)}{\varepsilon^2})$.*

Proof: The bound on the running time is obvious. So it remains to show that the pair output by the algorithm is ε-optimal. As mentioned above, we analyze the change in the potential function (2.6). Note that, due to the random choices of the algorithm, the potential function is a random variable. We will prove the following bound.

Lemma 2.5. *For $t = 1, 2, \ldots,$*

$$\mathbb{E}[\Phi(t)] \le \mathbb{E}[\Phi(t-1)](1 + \frac{\varepsilon^2}{6})^2.$$

Then by iteration we get that $\mathbb{E}[\Phi(t)] \le \mathbb{E}[\Phi(0)](1+\frac{\varepsilon}{6}^2)^{2t} \le \Phi(0)e^{\frac{\varepsilon^2 t}{3}}$, where the last inequality follows by using the inequality $1+x \le e^x$, valid for all real x. This implies by Markov's inequality that, with probability at least $\frac{1}{2}$, after t iterations,

$$\Phi(t) \le 2e^{\frac{\varepsilon^2}{3}t}\Phi(0). \tag{2.7}$$

Note that $\Phi(t) = \sum_{i,j} e^{\frac{\varepsilon X(t)^T A^j}{2} - \frac{\varepsilon A_i Y(t)}{2}}$. Since each term in this sum is non-negative and the sum is bounded by $2e^{\frac{\varepsilon^2}{3}t}\Phi(0)$, we conclude that each term is also bounded by the same bound. Taking logs and using $\Phi(0) = nm$, we get that

$$\frac{\varepsilon X(t)^T A^j}{2} - \frac{\varepsilon A_i Y(t)}{2} \le \ln(2nm) + \frac{\varepsilon^2 t}{3}.$$

or

$$\frac{X(t)^T}{t} A^j \le A_i \frac{Y(t)}{t} + \frac{2\ln(2nm)}{\varepsilon t} + \frac{2\varepsilon}{3}.$$

Using the value of $t = T = \frac{6\ln(2nm)}{\varepsilon^2}$ at the end of the last iteration, we get that $\frac{X(t)^T}{t}A^j \le A_i\frac{Y(t)}{t} + \varepsilon$, implying (see Exercise 1(ii)) that the pair of strategies output by the algorithm is ε-optimal. $\square$

It remains to prove Lemma 2.5.

Proof of Lemma 2.5. Fix an iteration t. Denote by $\Delta X = \Delta X(t)$ and $\Delta Y = \Delta Y(t)$ the changes in the vectors X and Y in iteration t, that is, in step 4, we use the updates $X(t+1) = X(t) + \Delta X$ and $Y(t+1) = Y(t) + \Delta Y$.

In the following, we condition on the values of $X(t)$ and $Y(t)$ (so for the moment, we will think that the only random events are those in step 3, and hence $p(t)$, $q(t)$, and $\phi(t)$ are all constants). Let $p(t) = (p_1(t), \ldots, p_m(t))$ and $q(t) = (q_1(t), \ldots, q_n(t))$. Then

$$\mathbb{E}[\Delta X] = \frac{p(t)}{|p(t)|} \quad \text{and} \quad \mathbb{E}[\Delta Y] = \frac{q(t)}{|q(t)|}.$$

To estimate the change in $\Phi(t-1)$, we estimate the changes in $|p(t)|$ and $|q(t)|$.

$$\begin{aligned} |p(t+1)| &= \sum_{i=1}^m p_i(t+1) = \sum_{i=1}^m e^{-\frac{\varepsilon A_i Y(t)}{2}} = \sum_{i=1}^m e^{-\frac{\varepsilon A_i (Y(t-1)+\Delta Y)}{2}} \\ &= \sum_{i=1}^m p_i(t) e^{-\frac{\varepsilon A_i \Delta Y}{2}}. \end{aligned} \tag{2.8}$$

Exercise 4. *Show that, for all* $\delta \in [-\frac{1}{2}, \frac{1}{2}]$, $e^\delta \le 1 + \delta + \frac{2}{3}\delta^2$.

Note that $\frac{\varepsilon A_i \Delta Y}{2} \in [-\frac{1}{2}, \frac{1}{2}]$ since we have assumed that $|a_{ij}| \le 1$. Thus the fact in Exercise 4, together with (2.8), implies that

$$\begin{aligned} |p(t+1)| &\le \sum_{i=1}^m p_i(t)\left[1 - \frac{\varepsilon A_i \Delta Y}{2} + \frac{\varepsilon^2}{6}(A_i \Delta Y)^2\right] \\ &\le \sum_{i=1}^m p_i(t)\left[1 + \frac{\varepsilon^2}{6} - \frac{\varepsilon A_i \Delta Y}{2}\right] = |p(t)|\left[1 + \frac{\varepsilon^2}{6} - \frac{\varepsilon p(t)^T A \Delta Y}{2|p(t)|}\right], \end{aligned}$$

where in the second inequality we used again the assumption that $|a_{ij}| \leq 1$ and hence $(A_i \Delta Y)^2 \leq 1$. In fact, this is the only place where this assumption plays a role in the analysis.

Taking the expectation with respect to ΔY, we get by linearity of expectation

$$\mathbb{E}[|p(t+1)|] \leq |p(t)| \left[1 + \frac{\varepsilon^2}{6} - \frac{\varepsilon p(t)^T A q(t)}{2|p(t)||q(t)|}\right].$$

Similarly, we can derive

$$\mathbb{E}[|q(t+1)|] \leq |q(t)| \left[1 + \frac{\varepsilon^2}{6} + \frac{\varepsilon q(t)^T A^T p(t)}{2|q(t)||p(t)|}\right].$$

Thus, using independence of ΔX and ΔY, we have

$$\begin{aligned} \mathbb{E}[\Phi(t)] &= \mathbb{E}[|p(t+1)|]\mathbb{E}[|q(t+1)|] \leq |p(t)||q(t)| \left[\left(1 + \frac{\varepsilon^2}{6}\right)^2 + \right. \\ & \left. \frac{\varepsilon}{2}\left(1 + \frac{\varepsilon^2}{6}\right)\left(\frac{q(t)^T A^T p(t)}{|q(t)||p(t)|} - \frac{p(t)^T A q(t)}{|p(t)||q(t)|}\right) - \frac{\varepsilon^2}{4}\frac{q(t)^T A^T p(t)}{|q(t)||p(t)|} \cdot \frac{p(t)^T A q(t)}{|p(t)||q(t)|}\right]. \end{aligned}$$

Since $\frac{q(t)^T A^T p(t)}{|q(t)||p(t)|} = \frac{p(t)^T A q(t)}{|p(t)||q(t)|}$, we get that $\mathbb{E}[\Phi(t)] \leq \Phi(t-1)\left(1 + \frac{\varepsilon^2}{6}\right)^2$. Recalling that this expectation was conditional on the values of $X(t)$ and $Y(t)$, the lemma follows by taking the expectation of both sides of this inequality with respect to $X(t)$ and $Y(t)$. $\square$

Bibliography

[Bro51] George W. Brown. Iterative solution of games by fictitious play. *In: T.C. Koopmans, Editor, Activity Analysis of Production and Allocation*, pages 374–376, 1951.

[GK95] Michael D. Grigoriadis and Leonid G. Khachiyan. A sublinear-time randomized approximation algorithm for matrix games. *Operations Research Letters*, 18(2):53 – 58, 1995.

[KY07] Christos Koufogiannakis and Neal E. Young. Beating simplex for fractional packing and covering linear programs. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 494–504, 2007.

[Rob51] Julia Robinson. An iterative method of solving a game. *The Annals of Mathematics*, 54(2):296–301, 1951.

[Sha58] Harold N. Shapiro. Note on a computation method in the theory of games. *Communications on Pure and Applied Mathematics*, 11(4):587–593, 1958.