CSIE5410 Optimization algorithms

Lecture 10: Learning in games

Yen-Huan Li (yenhuan.li@csie.ntu.edu.tw)

20.12.2018

Department of Computer Science and Information Engineering
National Taiwan University

Abstract

Consider a two-player zero-sum repeated game.

If both players use on-line convex optimization algorithms to choose their strategies, what will happen?

Recommended reading

- D. Fudenberg and D. K. Levine. 1995. Consistency and cautious fictitious play.
- *D. P. Foster and R. V. Vohra. 1997. Calibrated learning and correlated equilibrium.
- Y. Freund and R. E. Schapire. 1999. Adaptive game playing using multiplicative weights.
- C. Daskalakis *et al.* 2015. Near-optimal no-regret algorithms for zero-sum games.

Table of contents

John von Neumann as an applied mathematician

John von Neumann
(1903–1957)

Known for:

- Minimax theorem (1928)
- *Mathematical Foundations of Quantum Mechanics* (1937)
- *Theory of Games and Economic Behavior* (1944)
- Von Neumann architecture (1945)
- Manhattan Project (1942–1946)
- Linear programming duality (1947)
- *Theory of Self-Reproducing Automata* (1966)
- Monte Carlo method, numerical weather forecast, etc.

Two-player zero-sum game

Setup of a two-player zero-sum game

Alice and Bob are playing a game.

- Alice can choose her action from the set $\mathcal{A} := \{ a_1, \dots, a_p \}$.
- Bob can choose his action from the set $\mathcal{B} := \{ b_1, \dots, b_q \}$.
- For every $(a, b) \in \mathcal{A} \times \mathcal{B}$, there is a number $\pi(a, b)$, which represents Alice's loss and Bob's pay-off.
- A (mixed) *strategy* is a randomized action.
  - Alice's strategy is specified by a probability distribution in $\Delta_p \subset \mathbb{R}^p$.
  - Bob's strategy is specified by a probability distribution in $\Delta_q \subset \mathbb{R}^q$.

Minimax strategy

Alice's minimax strategy is given by

$$x^\star \in \arg\min_x \max_y \left\{ \sum_{i,j} \pi(a_i, b_j) x^{(i)} y^{(j)} \;\middle|\; x \in \Delta_p, y \in \Delta_q \right\}.$$

Bob's minimax strategy is given by

$$y^\star \in \arg\max_y \min_x \left\{ \sum_{i,j} \pi(a_i, b_j) x^{(i)} y^{(j)} \;\middle|\; x \in \Delta_p, y \in \Delta_q \right\}.$$

Remark. Define $A \in \mathbb{R}^{p \times q}$, whose (a, b)-th entry is given by $\pi(a, b)$. Then the objective function can be written as $\langle x, Ay \rangle$.

Remark. It suffices to specify the game by the matrix A.

Von Neumann's minimax theorem

Theorem. Let $\mathcal{X}$ and $\mathcal{Y}$ be closed bounded convex sets in $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively. Let $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be a continuous convex-concave function. Then, it holds that

$$\begin{aligned}
&\min_x \max_y \{ f(x, y) \mid x \in \mathcal{X}, y \in \mathcal{Y} \} \\
&\quad = \max_y \min_x \{ f(x, y) \mid x \in \mathcal{X}, y \in \mathcal{Y} \} .
\end{aligned}$$

Corollary. It holds that

$$\begin{aligned}
&\min_x \max_y \{ \langle x, Ay \rangle \mid x \in \Delta_p, y \in \Delta_q \} \\
&\quad = \max_y \min_x \{ \langle x, Ay \rangle \mid x \in \Delta_p, y \in \Delta_q \} .
\end{aligned}$$

Remark. If the two players can take actions sequentially, then the order does not matter.

Solution and min-max pair of a two-player zero-sum game

Definition. The *value* of the game specified by the matrix $A \in \mathbb{R}^{p\times q}$ is given by

$$\begin{aligned} v &:= \min_x \max_y \left\{ \langle x, Ay \rangle \mid x \in \Delta_p, y \in \Delta_q \right\} \\ &= \max_y \min_x \left\{ \langle x, Ay \rangle \mid x \in \Delta_p, y \in \Delta_q \right\}. \end{aligned}$$

Definition. We say that $(x^\star, y^\star) \in \Delta_p \times \Delta_q$ is a *min-max pair* of the game specified by $A \in \mathbb{R}^{p\times q}$, if and only if

$$v = \langle x^\star, Ay^\star \rangle .$$

Nash equilibrium

Definition. A pair $(x^\star, y^\star) \in \Delta_p \times \Delta_q$ is called a *Nash equilibrium* of the game specified by $A \in \mathbb{R}^{p\times q}$, if and only if

$$\begin{aligned} \langle x^\star, Ay^\star\rangle &\leq \langle x, Ay^\star\rangle\,, \quad \forall x \in \Delta_p,\\ \langle x^\star, Ay^\star\rangle &\geq \langle x^\star, Ay\rangle\,, \quad \forall y \in \Delta_q. \end{aligned}$$

Proposition. In a two-player zero-sum game, a pair $(x^\star, y^\star) \in \Delta_p \times \Delta_q$ is a Nash equilibrium, if and only if it is a min-max pair.

Proof. The proposition follows from the definitions.

Facts about computing a Nash equilibrium

Consider a multiple-player possibly non-zero-sum game.

Theorem. There always exists a Nash equilibrium.

Proof. Apply Kakutani fixed-point theorem.

Remark. "That's trivial, you know. That's just a fixed-point theorem."—J. von Neumann.

Theorem. Computing a Nash equilibrium is PPAD-complete.

Theorem. Computing a Nash equilibrium in a two-player zero-sum game can be done in polynomial time.

J. Nash. 1950. Equilibrium points in n-person games.

C. Daskalakis *et al.* The complexity of computing a Nash equilibrium.

A. Gilpin *et al.* 2012. First-order algorithm with $O(\ln(1/\varepsilon))$ convergence for ε-equilibrium in two-person zero-sum games.

Learning in games (1/2)

The concept of a Nash equilibrium (NE) is so important to game theory that an extensive literature devoted to its defense and advancement exists.

Even so, there are aspects of the Nash equilibrium concept that are puzzling.

One is why any player should assume that the other will play their Nash equilibrium strategy?

Aumann (1987) says: "This is particularly perplexing when, as often happens, there are multiple equilibria; but it has considerable force even when the equilibrium is unique."

D. P. Foster. 1997. Calibrated learning and correlated equilibrium.

Learning in games (2/2)

One resolution is to argue that the assumption about an opponent's plays are the outcome of some learning process

Learning is modeled as recurrent updating.

Players choose a best reply on the basis of their forecasts of their opponents' future choices.

Much attention has focused on developing forecast rules by which a Nash equilibrium (or its refinements) may be learned.

Question. Does there exist an *uncoupled* dynamics that converges to a Nash equilibrium? If yes, how fast is the convergence?

D. P. Foster. 1997. Calibrated learning and correlated equilibrium.

Coupled vs. uncoupled dynamics

An uncoupled dynamics in a two-player game should follow the following template.

For $t = 1, \ldots, T$, the following happen *simultaneously*.

- Alice chooses a strategy x_t, given her losses and $x_1, \ldots, x_{t-1}$ and $y_1, \ldots, y_{t-1}$.
- Bob chooses a strategy y_t, given his pay-offs and $x_1, \ldots, x_{t-1}$ and $y_1, \ldots, y_{t-1}$.

Remark. In practice, typically only the other player's actions, instead of strategies, can be observed. Convergence to a Nash equilibrium can be achieved in this practical scenario.

As we would like to discuss convergence to a Nash equilibrium, we have to specify a definition of an approximate Nash equilibrium.

Definition. A pair $(\tilde{x}, \tilde{y}) \in \Delta_p \times \Delta_q$ is called an ε-approximate Nash equilibrium, if and only if

$$\langle \tilde{x}, A\tilde{y} \rangle \leq \min_i \langle e_i, A\tilde{y} \rangle + \varepsilon,$$
$$\langle \tilde{x}, A\tilde{y} \rangle \geq \max_j \langle \tilde{x}, Ae_j \rangle - \varepsilon.$$

Remark. Equivalently, the minimization and maximization in the definition can be modified to be over all $x \in \Delta_p$ and $y \in \Delta_q$, respectively. (Why?)

Fictitious play (1/3)

Definition. If fictitious play, both players choose the best responses with regard to the empirical frequency of other player's actions. That is,

$$x_{t+1} \in \arg\min_{x} \{ \langle x, A\hat{y}_t \rangle \mid x \in \Delta_p \},$$
$$y_{t+1} \in \arg\max_{y} \{ \langle \hat{x}_t, Ay \rangle \mid y \in \Delta_q \},$$

where $\hat{x}_t$ and $\hat{y}_t$ denote the empirical average of the actions of Alice and Bob, respectively.

Remark. This is also known as the *follow-the-leader* approach in online learning.

G. W. Brown. 1949. Some notes on computation of games solutions.

Fictitious play (2/3)

Theorem. Let $\bar{x}_t$ and $\bar{y}_t$ be the time average of x_t's and y_t's, respectively. Suppose that x_t's and y_t's are generated by the fictitious play in a two-player zero-sum game. Then, $(\bar{x}_t, \bar{y}_t)$ is an ε-Nash equilibrium with

$$\varepsilon = O\left(t^{-\frac{1}{p+q-2}}\right).$$

Question. What if we adopt the fictitious play in an online learning problem?

J. Robinson. 1951. An iterative method of solving a game. C. Daskalakis *et al.* 2014. A counter-example to Karlin's strong conjecture for fictitious play.

S. Karlin's conjecture

$$\varepsilon = O\left(t^{-\frac{1}{2}}\right).$$

Theorem. If a tie-breaking rule is not specified, then there exists a two-player zero-sum game where the pair $(\bar{x}_t, \bar{y}_t)$ generated by the fictitious play is an ε-Nash equilibrium with

$$\varepsilon = \Theta\left(t^{-\frac{1}{p}}\right),$$

assuming that $p = q$.

C. Daskalakis *et al.* 2014. A counter-example to Karlin's strong conjecture for fictitious play.

Learning in games

Multiplicative weights (MW)

Definition. The MW algorithm for Alice is given by

$$x_{t+1}^{(i)} \leftarrow \frac{x_t^{(i)} \mathrm{e}^{-\eta\langle e_i, A y_t\rangle}}{Z_t}, \quad \forall t \in \mathbb{N},$$

where $e_i \coloneqq (\delta_{i,j})_{1 \leq j \leq p} \in \mathbb{R}^p$, and Z_t is a positive real number such that $x_{t+1} \in \Delta_p$. The MW algorithm for Bob is defined analogously.

Question. How do we interpret the algorithm?

Y. Freund and R. E. Schapire. 1999. Adaptive game playing using multiplicative weights.

Cautious fictitious play

A slightly different, more practical variant is the following.

Definition. The cautious fictitious play for Alice is given by

$$x_{t+1}^{(i)} \leftarrow \frac{x_t^{(i)} \mathrm{e}^{-\eta \langle e_i, A\hat{y}_t \rangle}}{\tilde{Z}_t}, \quad \forall t \in \mathbb{N},$$

where $\hat{y}_t$ denotes the empirical distribution of Bob's actions, and $\tilde{Z}_t$ is a positive real number such that $x_{t+1} \in \Delta_p$. The cautious fictitious play for Bob is defined analogously.

D. Fudenberg and D. K. Levine. 1995. Consistency and cautious fictitious play.

Decision theoretic online learning (DTOL) perspective

Alice's side.

Write Alice's action set as $\mathcal{A} = \{ a_1, \ldots, a_p \}$. Let $L_0 = 0$. For $t = 1, \ldots, T$, the following happens sequentially.

1. $\textsc{Expert}$ i announces $a_i \in \mathcal{A}$, $i = 1, \ldots, p$.
2. Alice chooses an expert randomly according to $x_t \in \Delta_p$.
3. $\textsc{Reality}$ announces a loss $\ell_{t,i} = \langle e_i, Ay_t \rangle$ for $\textsc{Expert}$ i.
4. $L_t = L_{t-1} + \mathsf{E}\left[\ell_{t,i}\right] = \langle x_t, Ay_t \rangle$.

Remark. Bob faces a similar DTOL problem.

Online convex optimization perspective

Alice's side.

Let $L_0 = 0$. For $t = 1, \ldots, T$, the following happens sequentially.

1. Alice announces her strategy $x_t \in \Delta_p$.
2. $\textsc{Reality}$ announces a function $f_t(x) := \langle x, Ay_t \rangle$ for all $x \in \Delta_p$.
3. $L_t = L_{t-1} + f_t(x_t)$.

Remark. Bob faces a similar online convex optimization problem.

Question. What will happen if both Alice and Bob adopt online entropic mirror descent?

Convex-concave games (1/2)

Let $\mathcal{X} \subset \mathbb{R}^p$ and $\mathcal{Y} \subset \mathbb{R}^q$ be bounded closed convex sets. Let $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be continuous and convex-concave.

Definition. The value of the game is given by

$$\begin{aligned} v &\coloneqq \min_x \max_y \{\, f(x,y) \mid x \in \mathcal{X}, y \in \mathcal{Y} \,\} \\ &= \max_y \min_x \{\, f(x,y) \mid x \in \mathcal{X}, y \in \mathcal{Y} \,\}. \end{aligned}$$

Let $L_0 = G_0 = 0$. For $t = 1, \ldots, T$, the following happen.

1. Player 1 announces $x_t \in \mathcal{X}$.

1'. Simultaneously, Player 2 announces $y_t \in \mathcal{Y}$.

2. Then, update $L_t = L_{t-1} + f(x_t, y_t)$ and $G_t = G_{t-1} - f(x_t, y_t)$.

Convex-concave games (2/2)

Player 1's side.

Let $L_0 = 0$. For $t = 1, \ldots, T$, the following happen sequentially.

1. Player 1 announces $x_t \in \mathcal{X}$.
2. $\textsc{Reality}$ announces $f_t = f(\cdot, y_t)$.
3. Update $L_t = L_{t-1} + f_t(x_t)$.

Player 2's side.

Let $(-G)_0 = 0$. For $t = 1, \ldots, T$, the following happen sequentially.

1. Player 2 announces $y_t \in \mathcal{Y}$.
2. $\textsc{Reality}$ announces $(-g)_t = f(x_t, \cdot)$.
3. Update $(-G)_t = (-G)_{t-1} + (-g)_t(y_t)$.

Constructive proof of von Neumann's minimax theorem

Theorem. Let $\mathcal{X} \subset \mathbb{R}^p$ and $\mathcal{Y} \subset \mathbb{R}^q$ be bounded closed convex sets. Let $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be continuous and convex-concave. We have

$$\begin{aligned} &\min_x \max_y \{\, f(x, y) \mid x \in \mathcal{X}, y \in \mathcal{Y} \,\} \\ &\quad = \max_y \min_x \{\, f(x, y) \mid x \in \mathcal{X}, y \in \mathcal{Y} \,\} . \end{aligned}$$

Proof. Consider the convex-concave game where both players adopt online convex optimization algorithms of sublinear regrets.

Y. Freund and R. E. Schapire. 1999. Adaptive game playing using multiplicative weights.

J. Abernethy and J.-K. Wang. 2017. On Frank-Wolfe and equilibrium computation.

Proof detailed

Lemma. Suppose both players adopt online convex optimization algorithms of regrets $R_{T,1}$ and $R_{T,2}$, respectively, where $T \in \mathbb{N}$ denotes the time horizon. Then, it holds that

$$
\begin{aligned}
&\min_x \max_y \{\, f(x,y) \mid x \in \mathcal{X}, y \in \mathcal{Y} \,\} \\
&\qquad \le \max_y \min_x \{\, f(x,y) \mid x \in \mathcal{X}, y \in \mathcal{Y} \,\} + \frac{R_{T,1}}{T} + \frac{R_{T,2}}{T}.
\end{aligned}
$$

Proof of the theorem

Proof of the theorem. Suppose both players adopt online mirror descent. Then we have both $R_{T,1} = o(T)$ and $R_{T,2} = o(T)$. Letting $T \to \infty$, we obtain

$$\begin{aligned}&\min_x \max_y \{\, f(x,y) \mid x \in \mathcal{X}, y \in \mathcal{Y} \,\}\\&\quad \le \max_y \min_x \{\, f(x,y) \mid x \in \mathcal{X}, y \in \mathcal{Y} \,\}.\end{aligned}$$

Notice that by definition, it holds that (max-min inequality)

$$\begin{aligned}&\min_x \max_y \{\, f(x,y) \mid x \in \mathcal{X}, y \in \mathcal{Y} \,\}\\&\quad \ge \max_y \min_x \{\, f(x,y) \mid x \in \mathcal{X}, y \in \mathcal{Y} \,\}.\end{aligned}$$

The theorem follows.

Proof of the lemma (1/2)

Proof of the lemma. Denote Player 1's and Player 2's regrets by $R_{T,1}$ and $R_{T,2}$, respectively. Then we write

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} f(x_t, y_t) &= \min_x \left\{ \frac{1}{T}\sum_{t=1}^{T} f(x, y_t) \;\middle|\; x \in \mathcal{X} \right\} + \frac{R_{T,1}}{T} \\
&\leq \min_x \left\{ f\left(x, \frac{1}{T}\sum_{t=1}^{T} y_t\right) \;\middle|\; x \in \mathcal{X} \right\} + \frac{R_{T,1}}{T} \\
&\leq \max_y \min_x \{ f(x, y) \mid x \in \mathcal{X}, y \in \mathcal{Y} \} + \frac{R_{T,1}}{T}.
\end{aligned}$$

We have exploited the concavity of $f(x, \cdot)$ for the second inequality.

Proof of the lemma (2/2)

Proof of the lemma continued. Similarly, we write

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} f(x_t, y_t) &= \max_y \left\{ \frac{1}{T}\sum_{t=1}^{T} f(x_t, y) \;\middle|\; y \in \mathcal{Y} \right\} - \frac{R_{T,2}}{T} \\
&\geq \max_y \left\{ f\left(\frac{1}{T}\sum_{t=1}^{T} x_t, y \right) \;\middle|\; y \in \mathcal{Y} \right\} - \frac{R_{T,2}}{T} \\
&\geq \min_x \max_y \{\, f(x, y) \mid x \in \mathcal{X}, y \in \mathcal{Y} \,\} - \frac{R_{T,2}}{T}.
\end{aligned}
$$

Combining the upper and lower bounds of $(1/T)\sum_{t=1}^{T} f(x_t, y_t)$, the lemma follows.

Convergence to a Nash equilibrium

Notice that from the proof in the previous slides, we also specified an approach to computing an ε-approximate Nash equilibrium.

Corollary. Let $\bar{x}_t$ and $\bar{y}_t$ be the time average of x_t's and y_t's up to time t. Denote the regret up to time t for Player i as $R_{t,i}$ as in the proof in the previous slides. It holds that $(\bar{x}_t, \bar{y}_t)$ is an ε_t-approximate Nash equilibrium with

$$\varepsilon_t = 2 \max \left\{ \frac{R_{t,1}}{t}, \frac{R_{t,2}}{t} \right\}.$$

Remark. Therefore, an algorithm of sublinear regret implies convergence to a Nash equilibrium.

Proof of the corollary

Proof. Define

$$\bar{v}_T := \frac{1}{T}\sum_{t=1}^{T} f(x_t, y_t).$$

Notice that in the previous slides, we have shown that

$$\bar{v}_T \le \min_x \{\, f(x, \bar{y}_T) \mid x \in \mathcal{X} \,\} + \frac{\varepsilon_T}{2},$$
$$\bar{v}_T \ge \max_y \{\, f(\bar{x}_T, y) \mid y \in \mathcal{Y} \,\} - \frac{\varepsilon_T}{2}.$$

Therefore, we have

$$-\frac{\varepsilon_T}{2} \le \langle \bar{x}_T, A\bar{y}_T \rangle - \bar{v}_T \le \frac{\varepsilon_T}{2}.$$

Also, notice that any deviation can only cause a difference at most $(\varepsilon_T/2)$ in the expected loss.

Convergence to a Nash equilibrium by the MW

Recall that for the MW (or online entropic mirror descent), the regrets for Alice and Bob are

$$O\left(\sqrt{T \log p}\right) \quad \text{and} \quad O\left(\sqrt{T \log q}\right),$$

respectively.

Corollary. Let $\varepsilon > 0$. In a two-player zero-sum game, if the two players adopt the MW algorithm, then the time-averaged pair $(\bar{x}_t, \bar{y}_t)$ is an ε-Nash equilibrium with

$$\varepsilon = O\left(\sqrt{\frac{\log(p+q)}{t}}\right).$$

Y. Freund and R. E. Schapire. 1999. Adaptive game playing using multiplicative weights.

Is the MW optimal for learning in games?

Question. Is the rate $\varepsilon = O(t^{-1/2})$ optimal?

Answer. The answer is surprisingly (why?) no. A faster rate of $O\left(t^{-1}\log t\right)$ can be achieved by *Nesterov's excessive gap technique* and *optimistic online mirror descent*.

C. Daskalakis *et al.* 2015. Near-optimal no-regret algorithms for zero-sum games.
A. Rakhlin and K. Sridharan. 2013. Optimization, learning, and games with predictable sequences.

Conclusions

Summary

- We have introduced the problem of learning in games.
- We have proved that a sub-linear regret algorithm ensures convergence to a Nash equilibrium.
- We have studied the fictitious play and the multiplicative weights algorithm.

Next lecture

- Solving min-max problems, or
- Exp-concavity and/or optimistic online learning.