Lecture 10. Online Learning in Games

Advanced Optimization (Fall 2023)

Peng Zhao

zhaop@lamda.nju.edu.cn

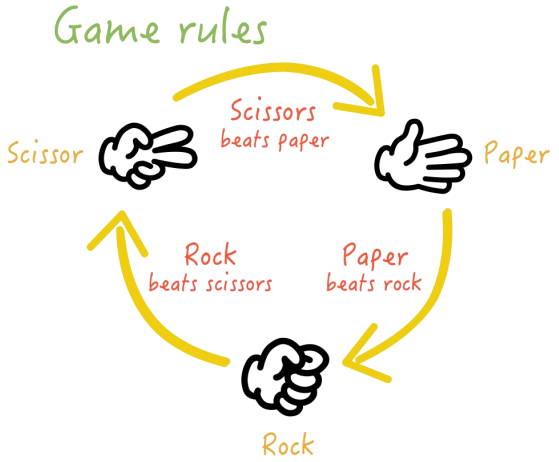
Nanjing University

Outline

- Two-player Zero-sum Games
- Minimax Theorem
- Repeated Play
- Faster Convergence via Adaptivity

Classic Game: *Rock-Paper-Scissors game*

- Rock-Paper-Scissors game



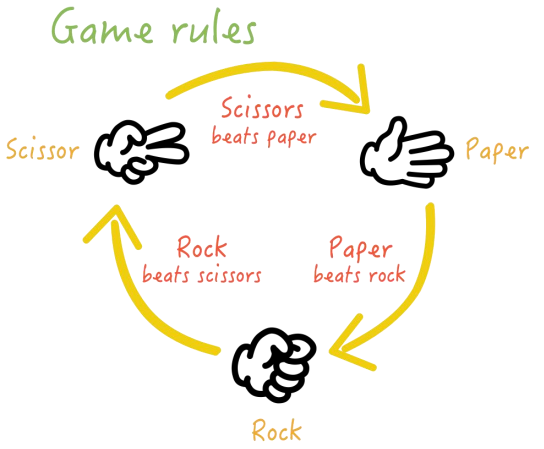
$$\begin{array}{c} \\ \textbf{Rock} \\ \textbf{Paper} \\ \textbf{Scissors} \end{array} \begin{array}{c} \begin{array}{ccc} \textbf{Rock} & \textbf{Paper} & \textbf{Scissors} \end{array} \\ \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix} \end{array}$$

- Strategy
  - ***Pure*** strategy: a fixed action, e.g., "Rock".
  - ***Mixed*** strategy: a ***distribution*** on all actions, e.g., ("Rock", "Paper", "Scissors") = $(1/3, 1/3, 1/3)$.

Two-Player Zero-Sum Games

- Terminology
  - ◇ game/payoff matrix $A \in [-1, 1]^{m \times n}$
  - ◇ two players
    - – player #1: **x**-player, row player, min player
    - – player #2: **y**-player, colume player, max player
  - ◇ action set (focusing on mixed strategy)
    - – player #1: $\Delta_m = \{\boldsymbol{p} \mid \sum_{i=1}^m p_i = 1, \text{ and } p_i \geq 0, \forall i \in [m]\}$.
    - – player #2: $\Delta_n = \{\boldsymbol{q} \mid \sum_{j=1}^n q_j = 1, \text{ and } q_j \geq 0, \forall j \in [n]\}$.

	Rock	**Paper**	**Scissors**
Rock	0	1	−1
Paper	−1	0	1
Scissors	1	−1	0



Two-Player Zero-Sum Games

- The protocol:
  - The repeated game is denoted by a (payoff) matrix $A \in [-1, 1]^{m \times n}$.
  - The $\mathbf{x}$-player has m actions, and the $\mathbf{y}$-player has n actions.
  - The goal of $\mathbf{x}$-player is to ***minimize her loss*** and the goal of $\mathbf{y}$-player is to ***maximize her reward***.

- Given the action $(\mathbf{x}, \mathbf{y}) \in \Delta_m \times \Delta_n$, the loss and reward are the same.
  - expected loss of $\mathbf{x}$-player is $\mathbb{E}[\text{loss}] = \sum_{i \in [m]} x_i \sum_{j \in [n]} y_j A_{ij} = \mathbf{x}^\top A \mathbf{y}$.
  - expected reward of $\mathbf{y}$-player is $\mathbb{E}[\text{reward}] = \sum_{i \in [m]} x_i \sum_{j \in [n]} y_j A_{ij} = \mathbf{x}^\top A \mathbf{y}$.

Nash Equilibrium

- What is a desired state for the two players in games?

Definition 2 (Nash equilibrium). A mixed strategy $(\mathbf{x}^\star, \mathbf{y}^\star)$ is called a Nash equilibrium if neither player has an incentive to change her strategy given that the opponent is keeping hers, i.e., for all $\mathbf{x} \in \Delta_m$ and $\mathbf{y} \in \Delta_n$, it holds that

$$\mathbf{x}^{\star\top} A \mathbf{y} \le \mathbf{x}^{\star\top} A \mathbf{y}^\star \le \mathbf{x}^\top A \mathbf{y}^\star.$$

Player-$\mathbf{y}$'s goal is to ***maximize*** her reward, changing from $\mathbf{y}^\star$ to $\mathbf{y}$ will decrease reward.

Player-$\mathbf{x}$'s goal is to ***minimize*** her loss, changing from $\mathbf{x}^\star$ to $\mathbf{x}$ will increase loss.

⇨ *Does the Nash equilibrium always exist for zero-sum games?*

Minimax Strategy and Maximin Strategy

- ***minimax*** strategy

$$\mathbf{x}^\star \in \arg\min_{\mathbf{x}} \max_{\mathbf{y}} \mathbf{x}^\top A \mathbf{y}$$

$\mathbf{x}$-player goes first, and given $\mathbf{x}$, the worst-case response of $\mathbf{y}$-player is $\max_{\mathbf{y}} \mathbf{x}^\top A \mathbf{y}$, so the best way for $\mathbf{x}$-player would be $\operatorname{argmin}_{\mathbf{x}} \max_{\mathbf{y}} \mathbf{x}^\top A \mathbf{y}$.

- ***maximin*** strategy

$$\mathbf{y}^\star \in \arg\max_{\mathbf{y}} \min_{\mathbf{x}} \mathbf{x}^\top A \mathbf{y}$$

$\mathbf{y}$-player goes first, and given $\mathbf{y}$, the worst-case response of $\mathbf{x}$-player is $\min_{\mathbf{x}} \mathbf{x}^\top A \mathbf{y}$, so the best way for $\mathbf{y}$-player would be $\operatorname{argmax}_{\mathbf{y}} \min_{\mathbf{x}} \mathbf{x}^\top A \mathbf{y}$.

Minimax Strategy and Maximin Strategy

- A natural consequence

$$\min_{\mathbf{x}} \max_{\mathbf{y}} \ \mathbf{x}^\top A\mathbf{y} \geq \max_{\mathbf{y}} \min_{\mathbf{x}} \ \mathbf{x}^\top A\mathbf{y}$$

***Intuition**: there should be no disadvantage of playing second*

- ***minimax*** strategy

  $\mathbf{x}^\star \in \arg\min_{\mathbf{x}} \max_{\mathbf{y}} \mathbf{x}^\top A\mathbf{y}$

  x-player goes first, and given $\mathbf{x}$, the worst-case response of y-player is $\max_{\mathbf{y}} \mathbf{x}^\top A\mathbf{y}$, so the best way for x-player would be $\operatorname{argmin}_{\mathbf{x}} \max_{\mathbf{y}} \mathbf{x}^\top A\mathbf{y}$.

- ***maximin*** strategy

  $\mathbf{y}^\star \in \arg\max_{\mathbf{y}} \min_{\mathbf{x}} \mathbf{x}^\top A\mathbf{y}$

  y-player goes first, and given $\mathbf{y}$, the worst-case response of x-player is $\min_{\mathbf{x}} \mathbf{x}^\top A\mathbf{y}$, so the best way for y-player would be $\operatorname{argmax}_{\mathbf{y}} \min_{\mathbf{x}} \mathbf{x}^\top A\mathbf{y}$.

Proof: Define $\mathbf{x}^\star \in \arg\min_{\mathbf{x}} \max_{\mathbf{y}} \ \mathbf{x}^\top A\mathbf{y}$ and $\mathbf{y}^\star \in \arg\max_{\mathbf{y}} \min_{\mathbf{x}} \ \mathbf{x}^\top A\mathbf{y}$.

$$\min_{\mathbf{x}} \max_{\mathbf{y}} \ \mathbf{x}^\top A\mathbf{y} = \max_{\mathbf{y}} \ \mathbf{x}^{\star\top} A\mathbf{y} \geq \mathbf{x}^{\star\top} A\mathbf{y}^\star \geq \min_{\mathbf{x}} \mathbf{x}^\top A\mathbf{y}^\star = \max_{\mathbf{y}} \min_{\mathbf{x}} \ \mathbf{x}^\top A\mathbf{y}.$$

(def) (def)

□

Von Neumann's Minimax Theorem

- For two-player zero-sum games, it is kind of surprising that the reverse direction is also true and thus minimax equals to maximin.

Theorem 1. *For any two-player zero-sum game $A \in [-1,1]^{m\times n}$, we have*

$$\min_{\mathbf{x}} \max_{\mathbf{y}} \ \mathbf{x}^\top A\mathbf{y} = \max_{\mathbf{y}} \min_{\mathbf{x}} \ \mathbf{x}^\top A\mathbf{y}.$$

The original proof relies on a fixed-point theorem (which is highly non-trivial).

Here gives a simple and ***constructive*** proof by running an online learning algo.

Connection with Online Learning

- Recall the OCO framework, regret notion, and the history bits.

Repeated Play

- It is often that a game is **repeatedly played for many times**

At each round $t = 1, 2, \ldots, T$:

(1) $\mathbf{x}$-player picks a mixed strategy $\mathbf{x}_t \in \Delta_m$

(2) simitaneously $\mathbf{y}$-player picks a mixed strategy $\mathbf{y}_t \in \Delta_n$

(3) $\mathbf{x}$-player and $\mathbf{y}$-player submit their strategies together

(4) $\mathbf{x}$-player receives loss $\mathbf{x}_t^\top A \mathbf{y}_t$ and observes $A\mathbf{y}_t$; $\mathbf{y}$-player receives loss $-\mathbf{x}_t^\top A \mathbf{y}_t$ and observes $-A\mathbf{x}_t$

The loss function that $\mathbf{x}$-player receives is $f_t^{\mathbf{x}}(\cdot) \triangleq \cdot^\top A\mathbf{y}_t$. *assume a gradient feedback*

$\Rightarrow$ $\mathbf{y}_t$ can depend on $\mathbf{x}_1, \ldots, \mathbf{x}_{t-1}$, meaning that $\mathbf{x}$-player is facing an *adptive adversary*.

Repeated Play

- Assume $\mathbf{x}$-player and $\mathbf{y}$-player run online algorithms with regret $\mathrm{Reg}_T^{\mathbf{x}}$ and $\mathrm{Reg}_T^{\mathbf{y}}$

Our goal: prove $\min_{\mathbf{x}} \max_{\mathbf{y}} \ \mathbf{x}^\top A\mathbf{y} \le \max_{\mathbf{y}} \min_{\mathbf{x}} \ \mathbf{x}^\top A\mathbf{y}$ via ***repeated play***.

Key idea: use the quantity $\frac{1}{T}\sum_{t=1}^T \mathbf{x}_t^\top A\mathbf{y}_t$ as a bridge between $\min_{\mathbf{x}} \max_{\mathbf{y}} \ \mathbf{x}^\top A\mathbf{y}$ and $\max_{\mathbf{y}} \min_{\mathbf{x}} \ \mathbf{x}^\top A\mathbf{y}$.

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^T \mathbf{x}_t^\top A\mathbf{y}_t &\le \min_{\mathbf{x}\in\Delta_m} \frac{1}{T}\sum_{t=1}^T \mathbf{x}^\top A\mathbf{y}_t + \frac{\mathrm{Reg}_T^{\mathbf{x}}}{T} \\
&= \min_{\mathbf{x}\in\Delta_m} \mathbf{x}^\top A\bar{\mathbf{y}}_T + \frac{\mathrm{Reg}_T^{\mathbf{x}}}{T} \qquad (\bar{\mathbf{y}}_T \triangleq \tfrac{1}{T}\textstyle\sum_{t=1}^T \mathbf{y}_t) \\
&\le \max_{\mathbf{y}\in\Delta_n} \min_{\mathbf{x}\in\Delta_m} \mathbf{x}^\top A\mathbf{y} + \frac{\mathrm{Reg}_T^{\mathbf{x}}}{T}
\end{aligned}$$

Repeated Play

- Assume $\mathbf{x}$-player and $\mathbf{y}$-player run online algorithms with regret $\mathrm{Reg}_T^{\mathbf{x}}$ and $\mathrm{Reg}_T^{\mathbf{y}}$

Our goal: prove $\min_{\mathbf{x}} \max_{\mathbf{y}} \ \mathbf{x}^\top A\mathbf{y} \le \max_{\mathbf{y}} \min_{\mathbf{x}} \ \mathbf{x}^\top A\mathbf{y}$ via ***repeated play***.

Key idea: use the quantity $\frac{1}{T}\sum_{t=1}^T \mathbf{x}_t^\top A\mathbf{y}_t$ as a bridge between $\min_{\mathbf{x}} \max_{\mathbf{y}} \ \mathbf{x}^\top A\mathbf{y}$ and $\max_{\mathbf{y}} \min_{\mathbf{x}} \ \mathbf{x}^\top A\mathbf{y}$.

$$
\begin{aligned}
-\frac{1}{T}\sum_{t=1}^T \mathbf{x}_t^\top A\mathbf{y}_t &\le \min_{\mathbf{y}\in\Delta_n} -\frac{1}{T}\sum_{t=1}^T \mathbf{x}_t^\top A\mathbf{y} + \frac{\mathrm{Reg}_T^{\mathbf{y}}}{T} \\
&= \min_{\mathbf{y}\in\Delta_n} -\bar{\mathbf{x}}_T^\top A\mathbf{y} + \frac{\mathrm{Reg}_T^{\mathbf{y}}}{T} \qquad (\bar{\mathbf{x}}_T \triangleq \tfrac{1}{T}\textstyle\sum_{t=1}^T \mathbf{x}_t) \\
&\le \max_{\mathbf{x}\in\Delta_m} \min_{\mathbf{y}\in\Delta_n} -\mathbf{x}^\top A\mathbf{y} + \frac{\mathrm{Reg}_T^{\mathbf{y}}}{T} = -\min_{\mathbf{x}\in\Delta_m} \max_{\mathbf{y}\in\Delta_n} \mathbf{x}^\top A\mathbf{y} + \frac{\mathrm{Reg}_T^{\mathbf{y}}}{T}
\end{aligned}
$$

Repeated Play

- Assume $\mathbf{x}$-player and $\mathbf{y}$-player run online algorithms with regret $\mathrm{Reg}_T^{\mathbf{x}}$ and $\mathrm{Reg}_T^{\mathbf{y}}$

Our goal: prove $\min_{\mathbf{x}} \max_{\mathbf{y}} \ \mathbf{x}^\top A\mathbf{y} \le \max_{\mathbf{y}} \min_{\mathbf{x}} \ \mathbf{x}^\top A\mathbf{y}$ via ***repeated play***.

Key idea: use the quantity $\frac{1}{T}\sum_{t=1}^T \mathbf{x}_t^\top A\mathbf{y}_t$ as a bridge between $\min_{\mathbf{x}} \max_{\mathbf{y}} \ \mathbf{x}^\top A\mathbf{y}$ and $\max_{\mathbf{y}} \min_{\mathbf{x}} \ \mathbf{x}^\top A\mathbf{y}$.

$$(1)\ \frac{1}{T}\sum_{t=1}^T \mathbf{x}_t^\top A\mathbf{y}_t \le \max_{\mathbf{y}\in\Delta_n} \min_{\mathbf{x}\in\Delta_m} \mathbf{x}^\top A\mathbf{y} + \frac{\mathrm{Reg}_T^{\mathbf{x}}}{T}$$

$$(2)\ -\frac{1}{T}\sum_{t=1}^T \mathbf{x}_t^\top A\mathbf{y}_t \le -\min_{\mathbf{x}\in\Delta_m} \max_{\mathbf{y}\in\Delta_n} \mathbf{x}^\top A\mathbf{y} + \frac{\mathrm{Reg}_T^{\mathbf{y}}}{T}$$

Repeated Play

- Assume $\mathbf{x}$-player and $\mathbf{y}$-player run online algorithms with regret $\mathrm{Reg}_T^{\mathbf{x}}$ and $\mathrm{Reg}_T^{\mathbf{y}}$

Our goal: prove $\min_{\mathbf{x}} \max_{\mathbf{y}} \ \mathbf{x}^\top A\mathbf{y} \le \max_{\mathbf{y}} \min_{\mathbf{x}} \ \mathbf{x}^\top A\mathbf{y}$ via ***repeated play***.

Key idea: use the quantity $\frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t^\top A\mathbf{y}_t$ as a bridge between $\min_{\mathbf{x}} \max_{\mathbf{y}} \ \mathbf{x}^\top A\mathbf{y}$ and $\max_{\mathbf{y}} \min_{\mathbf{x}} \ \mathbf{x}^\top A\mathbf{y}$.

$$(1)\ \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t^\top A\mathbf{y}_t \le \max_{\mathbf{y}\in\Delta_n} \min_{\mathbf{x}\in\Delta_m} \mathbf{x}^\top A\mathbf{y} + \frac{\mathrm{Reg}_T^{\mathbf{x}}}{T} \qquad (2)\ -\frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t^\top A\mathbf{y}_t \le -\min_{\mathbf{x}\in\Delta_m} \max_{\mathbf{y}\in\Delta_n} \mathbf{x}^\top A\mathbf{y} + \frac{\mathrm{Reg}_T^{\mathbf{y}}}{T}$$

$$\min_{\mathbf{x}\in\Delta_m} \max_{\mathbf{y}\in\Delta_n} \mathbf{x}^\top A\mathbf{y} \overset{(2)}{\le} \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t^\top A\mathbf{y}_t + \frac{\mathrm{Reg}_T^{\mathbf{y}}}{T} \overset{(1)}{\le} \max_{\mathbf{y}\in\Delta_n} \min_{\mathbf{x}\in\Delta_m} \mathbf{x}^\top A\mathbf{y} + \frac{\mathrm{Reg}_T^{\mathbf{x}}}{T} + \frac{\mathrm{Reg}_T^{\mathbf{y}}}{T}$$

If $\mathrm{Reg}_T^{\mathbf{x}}$ and $\mathrm{Reg}_T^{\mathbf{y}}$ are sublinear in T, the gap becomes to 0 when $T \to \infty$. $\square$

Minimax Solution and Nash equilibrium

Definition 2 (Nash equilibrium). A mixed strategy $(\mathbf{x}^\star, \mathbf{y}^\star)$ is called a Nash equilibrium if neither player has an incentive to change her strategy given that the opponent is keeping hers, i.e., for all $\mathbf{x} \in \Delta_m$ and $\mathbf{y} \in \Delta_n$, it holds that

$$\mathbf{x}^{\star\top} A \mathbf{y} \le \mathbf{x}^{\star\top} A \mathbf{y}^\star \le \mathbf{x}^\top A \mathbf{y}^\star.$$

- Relationship beween **Nash equilibrium** and **minimax solution.**

Theorem 2. *A pair of mixed strategy* $(\mathbf{x}^\star, \mathbf{y}^\star)$ *is a Nash equilibrium* ***if and only if*** *it is also a minimax solution (optimizer of* $\min_{\mathbf{x}} \max_{\mathbf{y}} \ \mathbf{x}^\top A \mathbf{y} = \max_{\mathbf{y}} \min_{\mathbf{x}} \ \mathbf{x}^\top A \mathbf{y}$*), i.e.,* $\mathbf{x}^\star \in \arg\min_{\mathbf{x}} \max_{\mathbf{y}} \ \mathbf{x}^\top A \mathbf{y}, \mathbf{y}^\star \in \arg\max_{\mathbf{y}} \min_{\mathbf{x}} \ \mathbf{x}^\top A \mathbf{y}$.

We denote by $(\mathbf{x}^\star, \mathbf{y}^\star)$ a Nash equilibrium, which will be proved as a minimax solution.

Proof

Proof: We denote by $(\mathbf{x}^\star, \mathbf{y}^\star)$ a Nash equilibrium, which will be proved as a minimax solution.

- *(Nash ⇒ minimax solution)*

$$\min_{\mathbf{x}} \max_{\mathbf{y}} \ \mathbf{x}^\top A\mathbf{y} \le \max_{\mathbf{y}} \ \mathbf{x}^{\star\top} A\mathbf{y} \underbrace{=}_{\text{(Nash)}} \mathbf{x}^{\star\top} A\mathbf{y}^\star \underbrace{=}_{\text{(Nash)}} \min_{\mathbf{x}} \ \mathbf{x}^\top A\mathbf{y}^\star \le \max_{\mathbf{y}} \min_{\mathbf{x}} \mathbf{x}^\top A\mathbf{y}$$

  By Von Neumann's minimax theorem, the above inequality is in fact an equality.

- *(minimax solution ⇒ Nash)*

$$\min_{\mathbf{x}} \max_{\mathbf{y}} \ \mathbf{x}^\top A\mathbf{y} \underbrace{=}_{\text{(minimax)}} \max_{\mathbf{y}} \ \mathbf{x}^{\star\top} A\mathbf{y} \ge \mathbf{x}^{\star\top} A\mathbf{y}^\star \ge \min_{\mathbf{x}} \ \mathbf{x}^\top A\mathbf{y}^\star \underbrace{=}_{\text{(minimax)}} \max_{\mathbf{y}} \min_{\mathbf{x}} \mathbf{x}^\top A\mathbf{y}$$

  By Von Neumann's minimax theorem, the above inequality is in fact an equality. □

Minimax Solution and Nash equilibrium

- Relationship beween **Nash equilibrium** and **minimax solution**

Theorem 2. *A pair of mixed strategy* $(\mathbf{x}^\star, \mathbf{y}^\star)$ *is a Nash equilibrium* ***if and only if*** *it is also a minimax solution (optimizer of* $\min_{\mathbf{x}} \max_{\mathbf{y}} \ \mathbf{x}^\top A\mathbf{y} = \max_{\mathbf{y}} \min_{\mathbf{x}} \ \mathbf{x}^\top A\mathbf{y}$*), i.e.,* $\mathbf{x}^\star \in \arg\min_{\mathbf{x}} \max_{\mathbf{y}} \ \mathbf{x}^\top A\mathbf{y}, \mathbf{y}^\star \in \arg\max_{\mathbf{y}} \min_{\mathbf{x}} \ \mathbf{x}^\top A\mathbf{y}$.

- Existence of Nash equilibrium

  Since the minximax solution always exists, by Theorem 2, Nash equilibrium also ***always exists*** in the two-player zero-sum games.

Nash Equilibrium Calculation

- How to **compute** an approximate a Nash equilibrium?

At each round $t = 1, 2, \ldots, T$:

(1) $\mathbf{x}$-player picks a mixed strategy $\mathbf{x}_t \in \Delta_m$

(2) simitaneously $\mathbf{y}$-player picks a mixed strategy $\mathbf{y}_t \in \Delta_n$

(3) $\mathbf{x}$-player and $\mathbf{y}$-player submit their strategies together

(4) $\mathbf{x}$-player receives loss $\mathbf{x}_t^\top A\mathbf{y}_t$ and observes $A\mathbf{y}_t$; $\mathbf{y}$-player receives loss $-\mathbf{x}_t^\top A\mathbf{y}_t$ and observes $-A\mathbf{x}_t$

Submit $\bar{\mathbf{x}}_T \triangleq \frac{1}{T}\sum_{t=1}^T \mathbf{x}_t$, ***and*** $\bar{\mathbf{y}}_T \triangleq \frac{1}{T}\sum_{t=1}^T \mathbf{y}_t$

Nash Equilibrium Calculation

- How to **compute** an approximate a Nash equilibrium?

  From the previous analysis, we know that the algorithm ensures:

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t^\top A\mathbf{y}_t \le \min_{\mathbf{x}\in\Delta_m}\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}^\top A\mathbf{y}_t + \frac{\mathrm{Reg}_T^{\mathbf{x}}}{T} = \min_{\mathbf{x}\in\Delta_m}\mathbf{x}^\top A\bar{\mathbf{y}}_T + \frac{\mathrm{Reg}_T^{\mathbf{x}}}{T} \le \max_{\mathbf{y}\in\Delta_n}\min_{\mathbf{x}\in\Delta_m}\mathbf{x}^\top A\mathbf{y} + \frac{\mathrm{Reg}_T^{\mathbf{x}}}{T}$$

$$-\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t^\top A\mathbf{y}_t \le \min_{\mathbf{y}\in\Delta_n} -\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t^\top A\mathbf{y} + \frac{\mathrm{Reg}_T^{\mathbf{y}}}{T} = \min_{\mathbf{y}\in\Delta_n} -\bar{\mathbf{x}}_T^\top A\mathbf{y} + \frac{\mathrm{Reg}_T^{\mathbf{y}}}{T} \le \max_{\mathbf{x}\in\Delta_m}\min_{\mathbf{y}\in\Delta_n} -\mathbf{x}^\top A\mathbf{y} + \frac{\mathrm{Reg}_T^{\mathbf{y}}}{T}$$

$$\Rightarrow \quad \mathbf{x}^{\star\top} A\mathbf{y}^\star \le \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t^\top A\mathbf{y}_t + \frac{\mathrm{Reg}_T^{\mathbf{y}}}{T} \le \min_{\mathbf{x}\in\Delta_m}\mathbf{x}^\top A\bar{\mathbf{y}}_T + \frac{\mathrm{Reg}_T^{\mathbf{x}}}{T} + \frac{\mathrm{Reg}_T^{\mathbf{y}}}{T}$$

$$\Rightarrow \quad \max_{\mathbf{y}\in\Delta_n}\bar{\mathbf{x}}_T^\top A\mathbf{y} \le \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t^\top A\mathbf{y}_t + \frac{\mathrm{Reg}_T^{\mathbf{y}}}{T} \le \mathbf{x}^{\star\top} A\mathbf{y}^\star + \frac{\mathrm{Reg}_T^{\mathbf{x}}}{T} + \frac{\mathrm{Reg}_T^{\mathbf{y}}}{T}$$

Nash Equilibrium Calculation

- How to **compute** an approximate a Nash equilibrium?

From the previous analysis, we know that the algorithm ensures:

$$\mathbf{x}^{\star\top} A \mathbf{y}^\star \le \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t^\top A \mathbf{y}_t + \frac{\mathrm{Reg}_T^{\mathbf{y}}}{T} \le \min_{\mathbf{x}\in\Delta_m} \mathbf{x}^\top A \bar{\mathbf{y}}_T + \frac{\mathrm{Reg}_T^{\mathbf{x}}}{T} + \frac{\mathrm{Reg}_T^{\mathbf{y}}}{T}$$

$$\max_{\mathbf{y}\in\Delta_n} \bar{\mathbf{x}}_T^\top A \mathbf{y} \le \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t^\top A \mathbf{y}_t + \frac{\mathrm{Reg}_T^{\mathbf{y}}}{T} \le \mathbf{x}^{\star\top} A \mathbf{y}^\star + \frac{\mathrm{Reg}_T^{\mathbf{x}}}{T} + \frac{\mathrm{Reg}_T^{\mathbf{y}}}{T}$$

It shows that $\min_{\mathbf{x}\in\Delta_m} \mathbf{x}^\top A \bar{\mathbf{y}}_T$ and $\max_{\mathbf{y}\in\Delta_n} \bar{\mathbf{x}}_T^\top A \mathbf{y}$ converges to the minimax value of the game at a rate of $(\mathrm{Reg}_T^{\mathbf{x}} + \mathrm{Reg}_T^{\mathbf{y}})/T$.

⇨ If $\mathbf{x}$-player and $\mathbf{y}$-player both run ***Hedge*** ($\mathrm{Reg}_T^{\mathbf{x}} = \mathrm{Reg}_T^{\mathbf{y}} = \mathcal{O}(\sqrt{T})$), the convergence rate is $\mathcal{O}(T^{-1/2})$.

Faster Convergence via Gradient Variation

- Can we do ***faster*** than the $\mathcal{O}(\sqrt{T})$ rate?

  Yes! We can use the Optimistic Online Mirror Descent of the last lecture.

- Recall in gradient-variation regret, the negative term is crucial.

$$\mathbf{x}_t = \arg\min_{\mathbf{x}\in\mathcal{X}} \eta \langle \nabla f_{t-1}(\mathbf{x}_{t-1}), \mathbf{x}\rangle + \frac{1}{2}\|\mathbf{x} - \widehat{\mathbf{x}}_t\|_2^2$$

$$\widehat{\mathbf{x}}_{t+1} = \arg\min_{\mathbf{x}\in\mathcal{X}} \eta \langle \nabla f_t(\mathbf{x}_t), \mathbf{x}\rangle + \frac{1}{2}\|\mathbf{x} - \widehat{\mathbf{x}}_t\|_2^2$$

Gradient Variation

$$\Longrightarrow \sum_{t=1}^T f_t(\mathbf{x}_t) - \sum_{t=1}^T f_t(\mathbf{u}) \le \eta \sum_{t=1}^T \|\nabla f_t(\mathbf{x}_t) - \nabla f_{t-1}(\mathbf{x}_{t-1})\|_2^2 + \frac{D^2}{2\eta} - \frac{1}{4\eta}\sum_{t=1}^T \|\mathbf{x}_{t+1} - \mathbf{x}_t\|_2^2$$

(negative term)

Faster Convergence via Gradient Variation

- Can we do ***faster*** than the $\mathcal{O}(\sqrt{T})$ rate?

  Yes! We can use the Optimistic Online Mirror Descent of the last lecture.

If $\mathbf{x}$-player runs OOMD with NE-entropy and gradients $\mathbf{g}_t^{\mathbf{x}} \triangleq A\mathbf{y}_t$ for $t \in [T]$:

$$\mathrm{Reg}_T^{\mathbf{x}} = \sum_{t=1}^{T} \langle A\mathbf{y}_t, \mathbf{x}_t - \mathbf{x} \rangle \lesssim \frac{1}{\eta^{\mathbf{x}}} \boxed{+\eta^{\mathbf{x}} \sum_{t=2}^{T} \|A\mathbf{y}_t - A\mathbf{y}_{t-1}\|_\infty^2} \boxed{-\frac{1}{\eta^{\mathbf{x}}} \sum_{t=2}^{T} \|\mathbf{x}_t - \mathbf{x}_{t-1}\|_1^2}$$

Similarly,

$$\mathrm{Reg}_T^{\mathbf{y}} = \sum_{t=1}^{T} \langle -A\mathbf{x}_t, \mathbf{y}_t - \mathbf{y} \rangle \lesssim \frac{1}{\eta^{\mathbf{y}}} \boxed{+\eta^{\mathbf{y}} \sum_{t=2}^{T} \|A\mathbf{x}_t - A\mathbf{x}_{t-1}\|_\infty^2} \boxed{-\frac{1}{\eta^{\mathbf{y}}} \sum_{t=2}^{T} \|\mathbf{y}_t - \mathbf{y}_{t-1}\|_1^2}$$

$\Rightarrow$ $\mathrm{Reg}_T^{\mathbf{x}} + \mathrm{Reg}_T^{\mathbf{y}} = \mathcal{O}(1)$, which leads to a much faster $\mathcal{O}(T^{-1})$ convergence rate! $\square$

Recap: negative terms in gradient variation

Optimistic OMD for Gradient-Variation Bound

$$\mathbf{x}_t = \arg\min_{\mathbf{x}\in\mathcal{X}} \left\{ \eta_t \langle \nabla f_{t-1}(\mathbf{x}_{t-1}), \mathbf{x} \rangle + \frac{1}{2}\|\mathbf{x} - \widehat{\mathbf{x}}_t\|_2^2 \right\}$$

$$\widehat{\mathbf{x}}_{t+1} = \arg\min_{\mathbf{x}\in\mathcal{X}} \left\{ \eta_t \langle \nabla f_t(\mathbf{x}_t), \mathbf{x} \rangle + \frac{1}{2}\|\mathbf{x} - \widehat{\mathbf{x}}_t\|_2^2 \right\}$$

Theorem 4 (Gradient Variation Regret Bound). *Assume that* $\psi(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_2^2$ *and* f_t *is L-smooth for all* $t \in [T]$*, when setting* $\eta_t = \min\{\frac{1}{4L}, \frac{D}{\sqrt{1+\widetilde{V}_{t-1}}}\}$ *and* $M_t = \nabla f_{t-1}(\mathbf{x}_{t-1})$*, the regret of Optimistic OMD to any comparator* $\mathbf{u} \in \mathcal{X}$ *is*

$$\text{Regret}_T = \sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}) \leq \mathcal{O}\left(\sqrt{1 + V_T}\right)$$

where $\widetilde{V}_{t-1} = \sum_{s=2}^{t-1} \|\nabla f_s(\mathbf{x}_{s-1}) - \nabla f_{s-1}(\mathbf{x}_{s-1})\|_2^2$ *is the empirical estimates of* V_t*.*

Gradient-Variation Bound

Proof. $$\sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}) \le \sum_{t=1}^{T} \eta_t \left\| \nabla f_t(\mathbf{x}_t) - M_t \right\|_2^2 + \sum_{t=1}^{T} \frac{1}{2\eta_t} \left(\|\mathbf{u} - \widehat{\mathbf{x}}_t\|_2^2 - \|\mathbf{u} - \widehat{\mathbf{x}}_{t+1}\|_2^2 \right)$$
$$- \sum_{t=1}^{T} \frac{1}{2\eta_t} \left(\|\widehat{\mathbf{x}}_{t+1} - \mathbf{x}_t\|_2^2 + \|\mathbf{x}_t - \widehat{\mathbf{x}}_t\|_2^2 \right)$$

(negative term)

$$\texttt{term (a)} \le 2 \sum_{t=2}^{T} \eta_t L^2 \left\| \mathbf{x}_t - \mathbf{x}_{t-1} \right\|_2^2 + 4D\sqrt{1 + V_T} + (4D+1)G^2$$

$$\texttt{term (b)} \le \frac{1}{2} \max\{4(L+1)D, D\sqrt{1+V_T} + D\}$$

$$\texttt{term (c)} \ge \sum_{t=2}^{T} \frac{1}{4\eta_t} \left\| \mathbf{x}_t - \mathbf{x}_{t-1} \right\|_2^2 \qquad (\eta_t = \min\{\tfrac{1}{4L}, \tfrac{D}{\sqrt{1+\widetilde{V}_{t-1}}}\})$$

Proof of Gradient-Variation Bound

Proof. Finally, putting three terms together yields

$$\texttt{term (a)} \le 2\sum_{t=2}^{T} \eta_t L^2 \|\mathbf{x}_t - \mathbf{x}_{t-1}\|_2^2 + 4D\sqrt{1+V_T} + (4D+1)G^2$$

$$\texttt{term (b)} \le \frac{1}{2}\max\{4(L+1)D, D\sqrt{1+V_T} + D\}$$

$$\texttt{term (c)} \ge \sum_{t=2}^{T} \frac{1}{4\eta_t}\|\mathbf{x}_t - \mathbf{x}_{t-1}\|_2^2 \qquad (\eta_t = \min\{\tfrac{1}{4L}, \tfrac{D}{\sqrt{1+\widetilde{V}_{t-1}}}\})$$

$$\Rightarrow \quad \mathrm{Regret}_T = \texttt{term (a)} + \texttt{term (b)} - \texttt{term (c)}$$

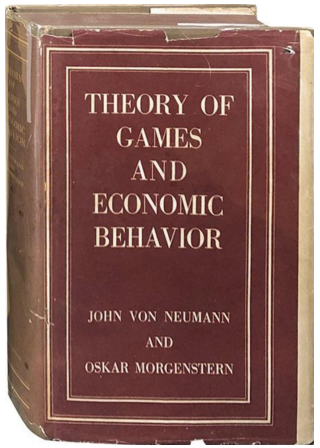
$$\le 5D\sqrt{1+V_T} + (4D+1)G^2 + 2LD = \mathcal{O}\big(\sqrt{1+V_T}\big). \qquad \square$$

History bits: Game Theory

- **John von Neumann**

John von Neumann was a Hungarian mathematician. By 26, he had already published 32 papers. He has been credited with founding game theory based on a paper he wrote in ***1928***. In 1944, he wrote, alongside Oskar Morgestern, the seminal book ***Theory of Games and Economic Behavior***.

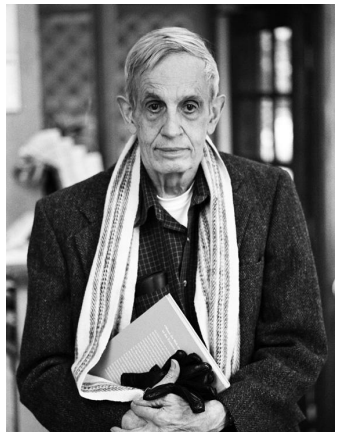
John von Neumann
1903-1957



History bits: Game Theory

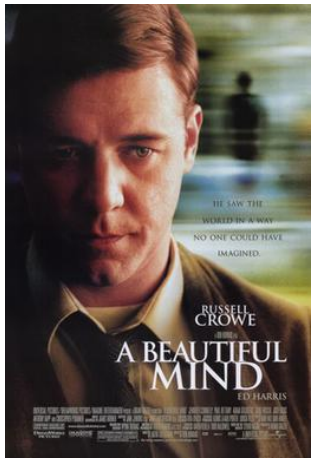
- **John Forbes Nash Jr.**

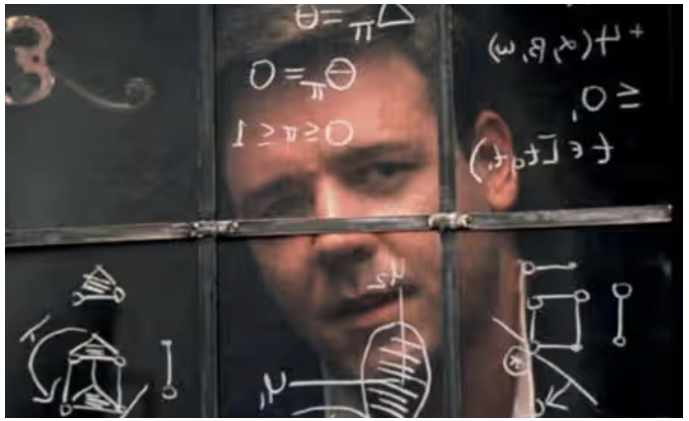
John Forbes Nash Jr., American mathematician who was awarded the ***1994 Nobel Prize*** for Economics.



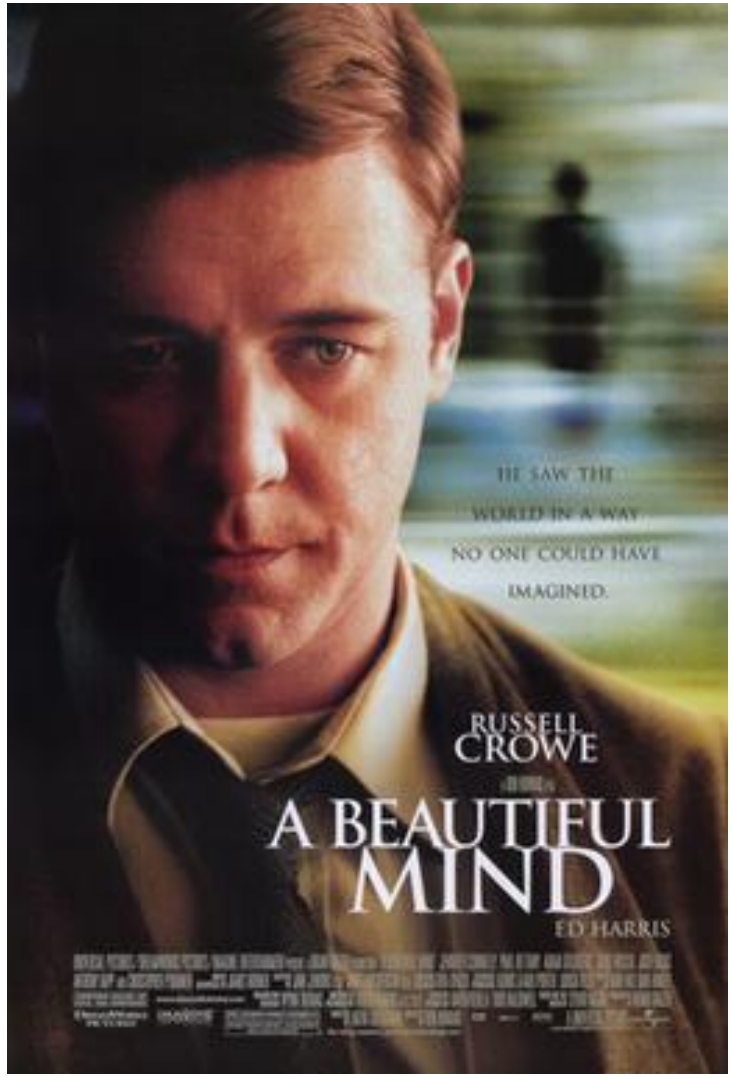
John Forbes Nash Jr.
1928-2015

He submitted a paper to the Proceedings of the National Academy of Sciences in 1949, where he proved that ***an equilibrium exists in every game***.





History bits: Game Theory



CARNEGIE INSTITUTE OF TECHNOLOGY
SCHENLEY PARK
PITTSBURGH 13, PENNSYLVANIA

DEPARTMENT OF MATHEMATICS
COLLEGE OF ENGINEERING AND SCIENCE

February 11, 1948

Professor S. Lefschetz
Department of Mathematics
Princeton University
Princeton, N. J.

Dear Professor Lefschetz:

This is to recommend Mr. John F. Nash, Jr. who has applied for entrance to the graduate college at Princeton.

Mr. Nash is nineteen years old and is graduating from Carnegie Tech in June. He is a mathematical genius.

Yours sincerely,

Richard J. Duffin

Richard J. Duffin

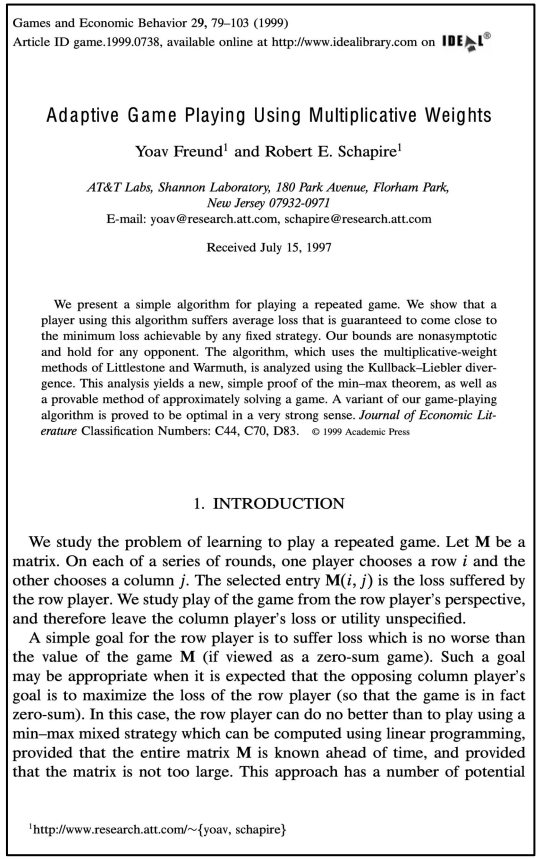
RJD:hl

He is a mathematical genius.

History bits: Online Learning in Games

- **Yoav Freund & Robert Schapire**

Yoav Freund and Robert Schapire's seminal paper in 1999 reveals the fundamental relationship between game theory and online learning, specifically, ***"a simple proof of the min-max theorem"***.



Games and Economic Behavior 29, 79–103 (1999)
Article ID game.1999.0738, available online at http://www.idealibrary.com on IDEAL®

Adaptive Game Playing Using Multiplicative Weights

Yoav Freund[1] and Robert E. Schapire[1]

AT&T Labs, Shannon Laboratory, 180 Park Avenue, Florham Park, New Jersey 07932-0971
E-mail: yoav@research.att.com, schapire@research.att.com

Received July 15, 1997

We present a simple algorithm for playing a repeated game. We show that a player using this algorithm suffers average loss that is guaranteed to come close to the minimum loss achievable by any fixed strategy. Our bounds are nonasymptotic and hold for any opponent. The algorithm, which uses the multiplicative-weight methods of Littlestone and Warmuth, is analyzed using the Kullback–Liebler divergence. This analysis yields a new, simple proof of the min–max theorem, as well as a provable method of approximately solving a game. A variant of our game-playing algorithm is proved to be optimal in a very strong sense. *Journal of Economic Literature* Classification Numbers: C44, C70, D83.

1. INTRODUCTION

We study the problem of learning to play a repeated game. Let **M** be a matrix. On each of a series of rounds, one player chooses a row i and the other chooses a column j. The selected entry $\mathbf{M}(i, j)$ is the loss suffered by the row player. We study play of the game from the row player's perspective, and therefore leave the column player's loss or utility unspecified.

A simple goal for the row player is to suffer loss which is no worse than the value of the game **M** (if viewed as a zero-sum game). Such a goal may be appropriate when it is expected that the opposing column player's goal is to maximize the loss of the row player (so that the game is in fact zero-sum). In this case, the row player can do no better than to play using a min–max mixed strategy which can be computed using linear programming, provided that the entire matrix **M** is known ahead of time, and provided that the matrix is not too large. This approach has a number of potential

[1] http://www.research.att.com/~{yoav, schapire}

Robert Schapire
1963-now

Yoav Freund
1961-now

Reference: Y. Freund and R. Schapire. Adaptive Game Playing Using Multiplicative Weights. Games and Economic Behavior, 1999.

History bits: Prediction with Expert Advice

Yoav Freund

Robert Schapire

Goldel Prize 2003

This paper introduced AdaBoost, an adaptive algorithm to improve the accuracy of hypotheses in machine learning. The algorithm demonstrated novel possibilities in analyzing data and is a permanent contribution to science even beyond computer science.

JOURNAL OF COMPUTER AND SYSTEM SCIENCES **55**, 119–139 (1997)
ARTICLE NO. SS971504

A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting*

Yoav Freund and Robert E. Schapire†

AT&T Labs, 180 Park Avenue, Florham Park, New Jersey 07932

Received December 19, 1996

In the first part of the paper we consider the problem of dynamically apportioning resources among a set of options in a worst-case on-line framework. The model we study can be interpreted as a broad, abstract extension of the well-studied on-line prediction model to a general decision-theoretic setting. We show that the multiplicative weight-update Littlestone–Warmuth rule can be adapted to this model, yielding bounds that are slightly weaker in some cases, but applicable to a considerably more general class of learning problems. We show how the resulting learning algorithm can be applied to a variety of problems, including gambling, multiple-outcome prediction, repeated games, and prediction of points in $\mathbb{R}^n$. In the second part of the paper we apply the multiplicative weight-update technique to derive a new boosting algorithm. This boosting algorithm does not require any prior knowledge about the performance of the weak learning algorithm. We also study generalizations of the new boosting algorithm to the problem of learning functions whose range, rather than being binary, is an arbitrary finite set or a bounded segment of the real line.

converting a "weak" PAC learning algorithm that performs just slightly better than random guessing into one with arbitrarily high accuracy.

We formalize our *on-line allocation model* as follows. The allocation agent A has N options or *strategies* to choose from; we number these using the integers $1, \ldots, N$. At each time step $t = 1, 2, \ldots, T$, the allocator A decides on a distribution $\mathbf{p}^t$ over the strategies; that is $p_i^t \geqslant 0$ is the amount allocated to strategy i, and $\sum_{i=1}^N p_i^t = 1$. Each strategy i then suffers some *loss* ℓ_i^t which is determined by the (possibly adversarial) "environment." The loss suffered by A is then $\sum_{i=1}^n p_i^t \ell_i^t = \mathbf{p}^t \cdot \boldsymbol{\ell}^t$, i.e., the average loss of the strategies with respect to A's chosen allocation rule. We call this loss function the *mixture loss*.

In this paper, we always assume that the loss suffered by any strategy is bounded so that, without loss of generality, $\ell_i^t \in [0, 1]$. Besides this condition, we make no assumptions

Reference: Y. Freund and R. Schapire. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. JCSS 1997.

History bits: Online Learning in Games

Optimization, Learning, and Games with Predictable Sequences

Alexander Rakhlin
University of Pennsylvania

Karthik Sridharan
University of Pennsylvania

Abstract

We provide several applications of Optimistic Mirror Descent, an online learning algorithm based on the idea of predictable sequences. First, we recover the Mirror Prox algorithm for offline optimization, prove an extension to Hölder-smooth functions, and apply the results to saddle-point type problems. Next, we prove that a version of Optimistic Mirror Descent (which has a close relation to the Exponential Weights algorithm) can be used by two strongly-uncoupled players in a finite zero-sum matrix game to converge to the minimax equilibrium at the rate of $\mathcal{O}((\log T)/T)$. This addresses a question of Daskalakis et al [6]. Further, we consider a partial information version of the problem. We then apply the results to convex programming and exhibit a simple algorithm for the approximate Max Flow problem.

1 Introduction

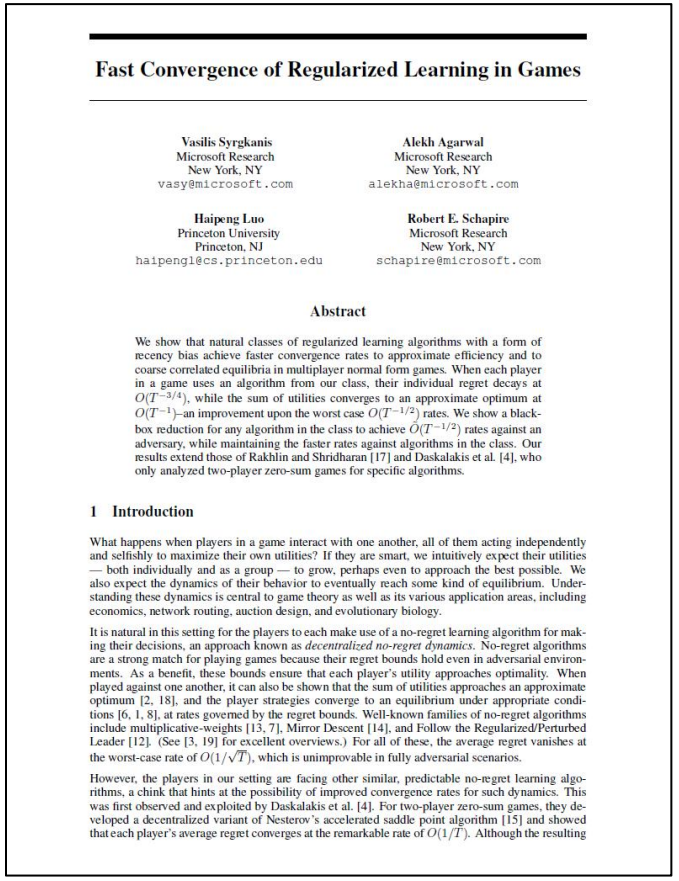
Recently, no-regret algorithms have received increasing attention in a variety of communities, including theoretical computer science, optimization, and game theory [3, 1]. The wide applicability of these algorithms is arguably due to the black-box regret guarantees that hold for arbitrary sequences. However, such regret guarantees can be loose if the sequence being encountered is not "worst-case". The reduction in "arbitrariness" of the sequence can arise from the particular structure of the problem at hand, and should be exploited. For instance, in some applications of online methods, the sequence comes from an additional computation done by the learner, thus being far from arbitrary.

One way to formally capture the partially benign nature of data is through a notion of predictable sequences [11]. We exhibit applications of this idea in several domains. First, we show that the Mirror Prox method [9], designed for optimizing non-smooth structured saddle-point problems, can be viewed as an instance of the predictable sequence approach. Predictability in this case is due precisely to smoothness of the inner optimization part and the saddle-point structure of the problem. We extend the results to Hölder-smooth functions, interpolating between the case of well-predictable gradients and "unpredictable" gradients.

Second, we address the question raised in [6] about existence of "simple" algorithms that converge at the rate of $\tilde{\mathcal{O}}(T^{-1})$ when employed in an uncoupled manner by players in a zero-sum finite matrix game, yet maintain the usual $\mathcal{O}(T^{-1/2})$ rate against arbitrary sequences. We give a positive answer and exhibit a fully adaptive algorithm that does not require the prior knowledge of whether the other player is collaborating. Here, the additional predictability comes from the fact that both players attempt to converge to the minimax value. We also tackle a partial information version of the problem where the player has only access to the real-valued payoff of the mixed actions played by the two players on each round rather than the entire vector.

Our third application is to convex programming: optimization of a linear function subject to convex constraints. This problem often arises in theoretical computer science, and we show that the idea of

Optimization, learning, and games with predictable sequences. NIPS 2013.



Fast Convergence of Regularized Learning in Games

Vasilis Syrgkanis
Microsoft Research
New York, NY
vasy@microsoft.com

Alekh Agarwal
Microsoft Research
New York, NY
alekha@microsoft.com

Haipeng Luo
Princeton University
Princeton, NJ
haipengl@cs.princeton.edu

Robert E. Schapire
Microsoft Research
New York, NY
schapire@microsoft.com

Abstract

We show that natural classes of regularized learning algorithms with a form of recency bias achieve faster convergence rates to approximate efficiency and to coarse correlated equilibria in multiplayer normal form games. When each player in a game uses an algorithm from our class, their individual regret decays at $O(T^{-3/4})$, while the sum of utilities converges to an approximate optimum at $O(T^{-1})$–an improvement upon the worst case $O(T^{-1/2})$ rates. We show a black-box reduction for any algorithm in the class to achieve $\tilde{O}(T^{-1/2})$ rates against an adversary, while maintaining the faster rates against algorithms in the class. Our results extend those of Rakhlin and Shridharan [17] and Daskalakis et al. [4], who only analyzed two-player zero-sum games for specific algorithms.

1 Introduction

What happens when players in a game interact with one another, all of them acting independently and selfishly to maximize their own utilities? If they are smart, we intuitively expect their utilities — both individually and as a group — to grow, perhaps even to approach the best possible. We also expect the dynamics of their behavior to eventually reach some kind of equilibrium. Understanding these dynamics is central to game theory as well as its various application areas, including economics, network routing, auction design, and evolutionary biology.

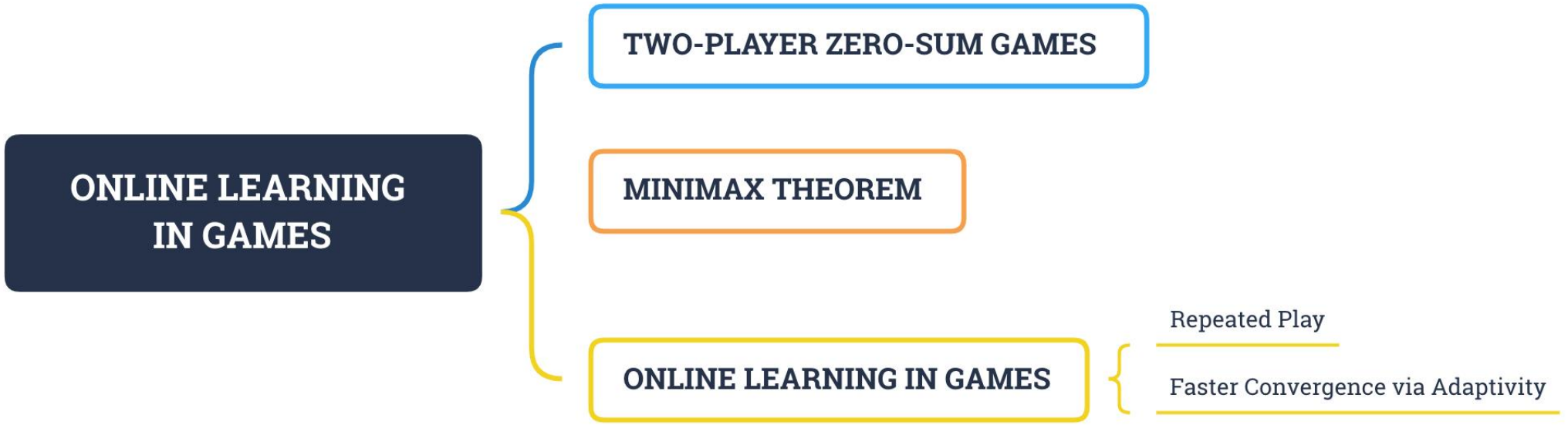
It is natural in this setting for the players to each make use of a no-regret learning algorithm for making their decisions, an approach known as *decentralized no-regret dynamics*. No-regret algorithms are a strong match for playing games because their regret bounds hold even in adversarial environments. As a benefit, these bounds ensure that each player's utility approaches optimality. When played against one another, it can also be shown that the sum of utilities approaches an approximate optimum [2, 18], and the player strategies converge to an equilibrium under appropriate conditions [6, 1, 8], at rates governed by the regret bounds. Well-known families of no-regret algorithms include multiplicative-weights [13, 7], Mirror Descent [14], and Follow the Regularized/Perturbed Leader [12]. (See [3, 19] for excellent overviews.) For all of these, the average regret vanishes at the worst-case rate of $O(1/\sqrt{T})$, which is unimprovable in fully adversarial scenarios.

However, the players in our setting are facing other similar, predictable no-regret learning algorithms, a chink that hints at the possibility of improved convergence rates for such dynamics. This was first observed and exploited by Daskalakis et al. [4]. For two-player zero-sum games, they developed a decentralized variant of Nesterov's accelerated saddle point algorithm [15] and showed that each player's average regret converges at the remarkable rate of $O(1/T)$. Although the resulting

NIPS 2015
best paper award

Fast convergence of regularized learning in games. NIPS 2015.

Summary



Q & A

Thanks!