Lecture 10. Online Learning in Games

Advanced Optimization (Fall 2022)

Peng Zhao

zhaop@lamda.nju.edu.cn

Nanjing University

Outline

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

History about Game Theory

- **Emil Borel**

Emil Borel wrote a series of papers between 1921 and 1927 where he set out to investigate whether it is possible to determine ***a method of play that is better than all others***.

Emil Borel
1871-1956

History about Game Theory

- **John von Neumann**

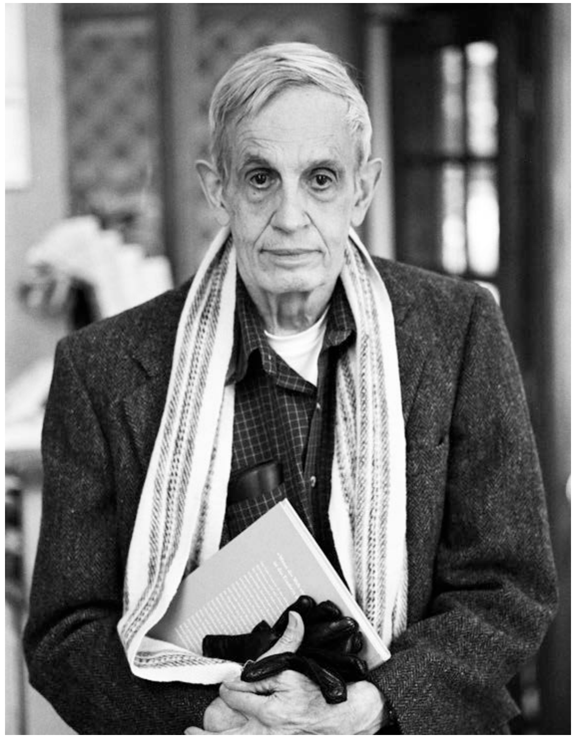
John von Neumann was a Hungarian mathematician. By 26, he had already published 32 papers. He has been credited with founding game theory based on a paper he wrote in ***1928***. In 1944, he wrote, alongside Oskar Morgestern, the seminal book ***Theory of Games and Economic Behavior***.

John von Neumann
1903-1957

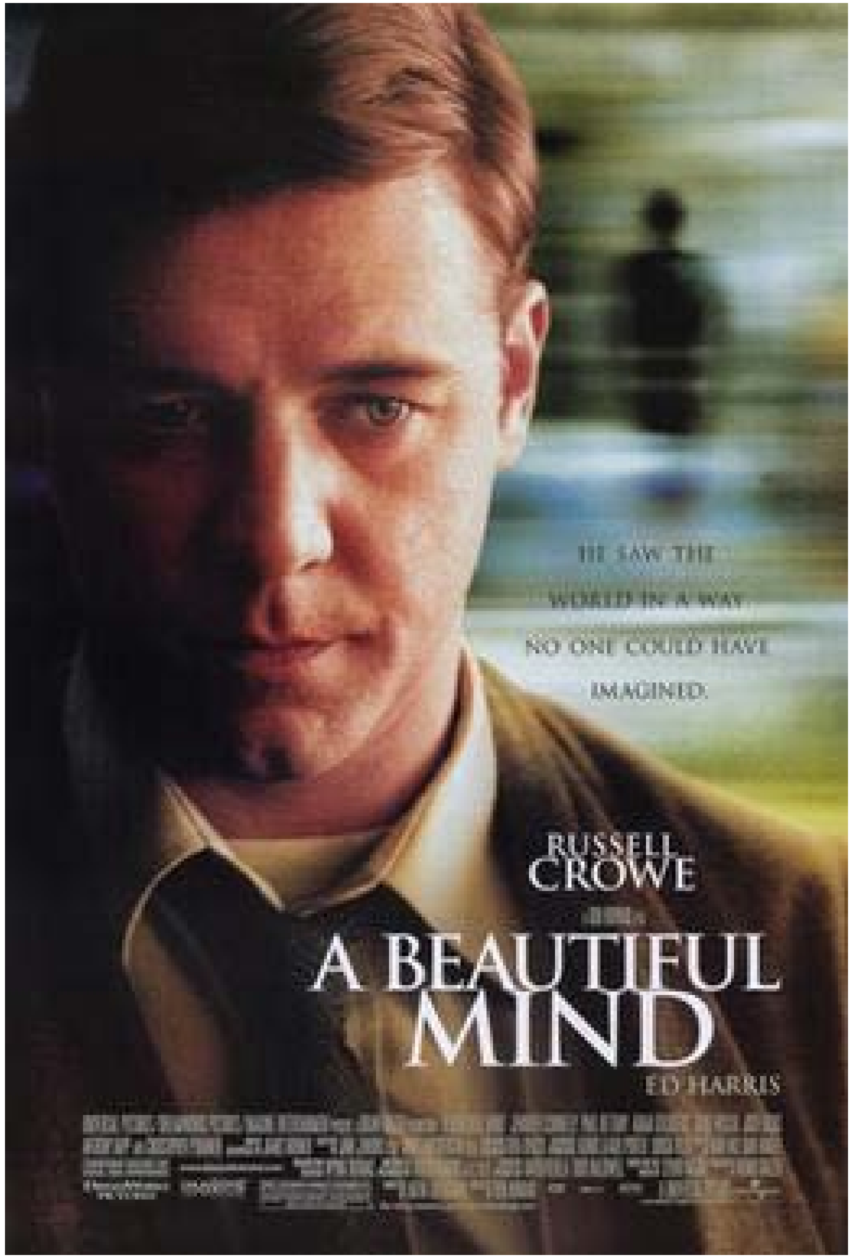
History about Game Theory

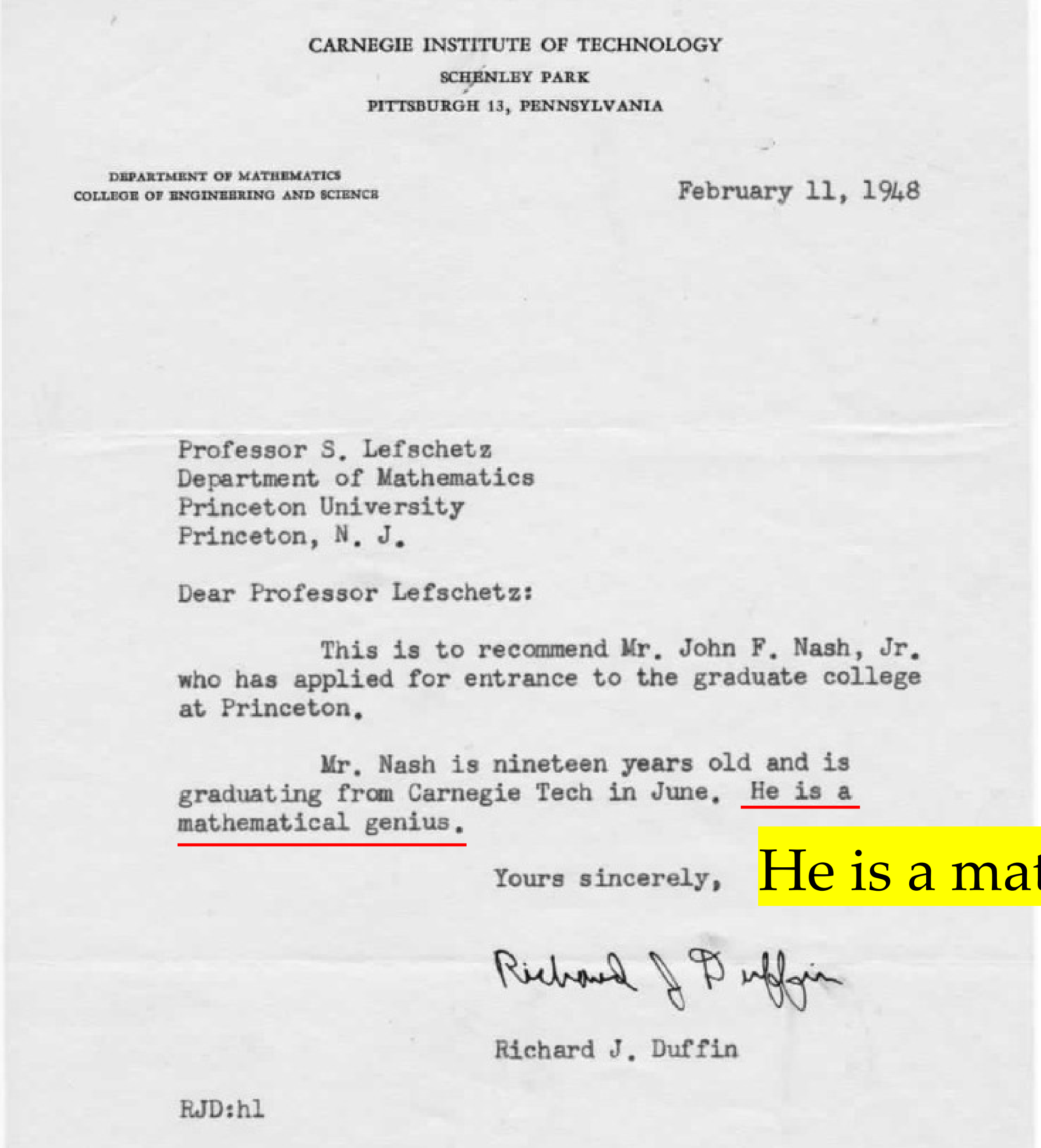
- **John Forbes Nash Jr.**

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



John Forbes Nash Jr.
1928-2015





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.

Two-Player Zero-Sum Games

- **Protocol**

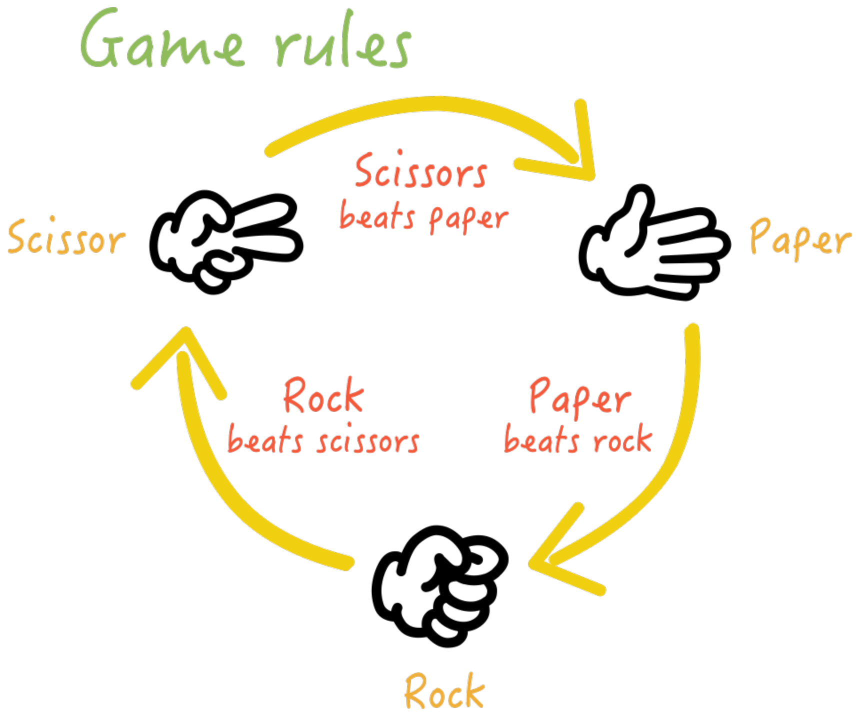
A two-player zero-sum game can be represented by a matrix $A \in [0, 1]^{m \times n}$:

- Player-x (row player) has m actions, and player-y (column player) has n actions

- The goal of player-x is to ***minimize her loss*** and the goal of player-y is to ***maximize her reward***.

Two-Player Zero-Sum Games

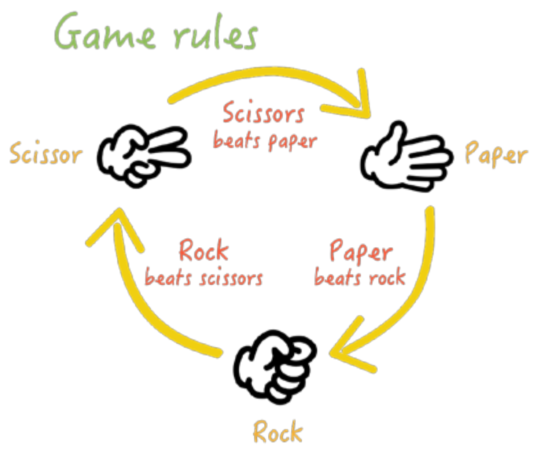
Classic example: ***Rock-Paper-Scissors game***



$$
\begin{array}{c} \\ \text{Rock} \\ \text{Paper} \\ \text{Scissors} \end{array}
\begin{array}{c} \begin{array}{ccc} \text{Rock} & \text{Paper} & \text{Scissors} \end{array} \\ \begin{pmatrix} 1/2 & 1 & 0 \\ 0 & 1/2 & 1 \\ 1 & 0 & 1/2 \end{pmatrix} \end{array}
$$

Two-Player Zero-Sum Games

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



- **Nash equilibrium**

Definition 2. A pair of mixed strategy $(\mathbf{x}^\star, \mathbf{y}^\star)$ is called a Nash equilibrium if neither player has a incentive to change his/her strategy given that the opponent is keeping his/hers, i.e.,

$$\mathbf{x}^{\star\top} A\mathbf{y}' \le \mathbf{x}^{\star\top} A\mathbf{y}^\star \le \mathbf{x}^\top A\mathbf{y}^\star, \forall \mathbf{x} \in \Delta_m, \mathbf{y} \in \Delta_n.$$

Two-Player Zero-Sum Games

- **Nash equilibrium**

Definition 2. A pair of mixed strategy $(\mathbf{x}^\star, \mathbf{y}^\star)$ is called a Nash equilibrium if neither player has a incentive to change his/her strategy given that the opponent is keeping his/hers, i.e.,

$$\mathbf{x}^{\star\top} A\mathbf{y}' \leq \mathbf{x}^{\star\top} A\mathbf{y}^\star \leq \mathbf{x}^\top A\mathbf{y}^\star, \forall \mathbf{x} \in \Delta_m, \mathbf{y} \in \Delta_n.$$

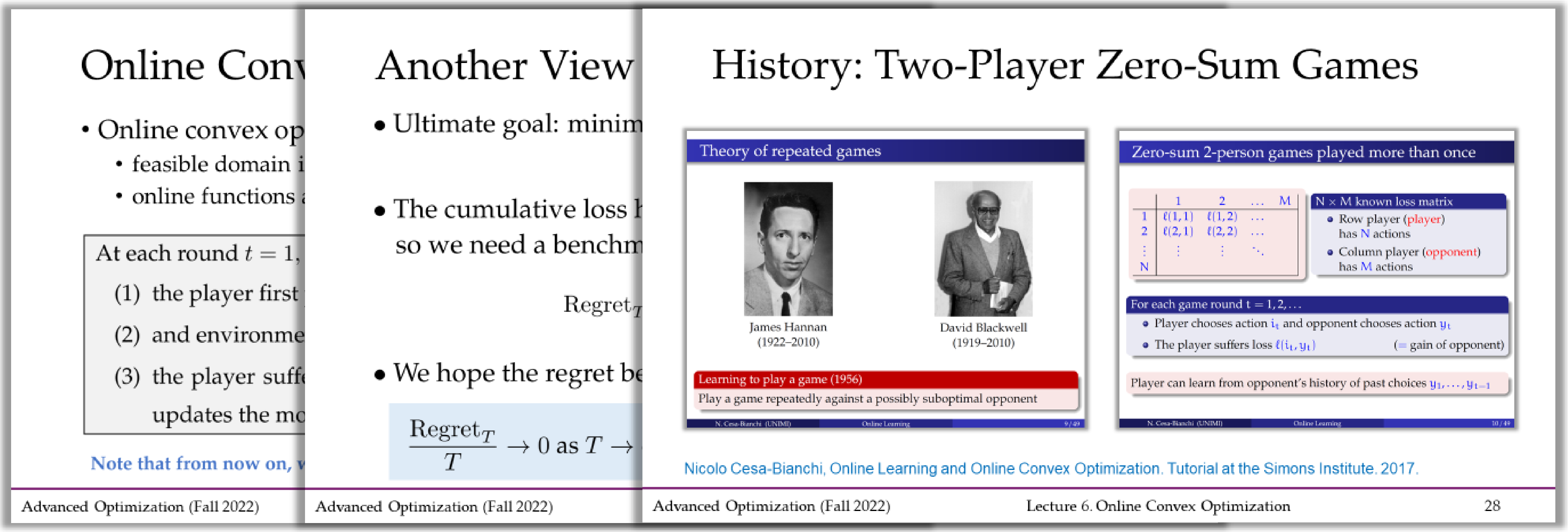
Player-y's goal is to *maximize* her reward, changing from $\mathbf{y}^\star$ to $\mathbf{y}$ will decrease reward.

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

⇒ *A natural question: is there always a Nash equilibrium?*

Connection with Online Learning

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



Minimax Strategy and Maximin Strategy

- ***minimax*** strategy

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

  in the worst case, playing $\mathbf{x}$ leads to a loss of at most $\max_{\mathbf{y}} \mathbf{x}^\top A \mathbf{y}$ for the x-player if the y-player sees $\mathbf{x}$ before making decisions

- ***maximin*** strategy

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

  in the worst case, playing $\mathbf{y}$ leads to a reward of at least $\min_{\mathbf{x}} \mathbf{x}^\top A \mathbf{y}$ for the y-player if the x-player sees $\mathbf{y}$ before making decisions

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

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

  in the worst case, playing $\mathbf{x}$ leads to a loss of at most $\max_{\mathbf{y}} \mathbf{x}^\top A \mathbf{y}$ for the x-player if the y-player sees $\mathbf{x}$ before making decisions

- ***maximin*** strategy

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

  in the worst case, playing $\mathbf{y}$ leads to a reward of at least $\min_{\mathbf{x}} \mathbf{x}^\top A \mathbf{y}$ for the y-player if the x-player sees $\mathbf{y}$ before making decisions

Proof: Define $\mathbf{x}^\star \in \arg\min_{\mathbf{x}} \max_{\mathbf{y}} \ \mathbf{x}^\top A \mathbf{y}$, then we have

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

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 $A \in [0,1]^{m \times n}$, 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).

In this lecture, we give a simple (constructive) proof by running online learning algo.

Repeated Play

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

> At each round $t = 1, 2, \ldots, T$:
>
> (1) player-x picks a mixed strategy $\mathbf{x}_t \in \Delta_m$
>
> (2) simitaneously player-y picks a mixed strategy $\mathbf{y}_t \in \Delta_n$
>
> (3) player-x and player-y submit their strategies together
>
> (4) player-x receives loss $\mathbf{x}_t^\top A \mathbf{y}_t$ and observes $A\mathbf{y}_t$; player-y receives loss $-\mathbf{x}_t^\top A \mathbf{y}_t$ and observes $-A\mathbf{x}_t$ ***assume full information***

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

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

Repeated Play

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

Our goal: prove $\min_{\mathbf{x}} \max_{\mathbf{y}} \ \mathbf{x}^\top A\mathbf{y} \leq \max_{\mathbf{y}} \min_{\mathbf{x}} \ \mathbf{x}^\top A\mathbf{y}$ 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 $\max_{\mathbf{y}} \min_{\mathbf{x}} \ \mathbf{x}^\top A\mathbf{y}$.

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

Repeated Play

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

Our goal: prove $\min_{\mathbf{x}} \max_{\mathbf{y}} \ \mathbf{x}^\top A\mathbf{y} \leq \max_{\mathbf{y}} \min_{\mathbf{x}} \ \mathbf{x}^\top A\mathbf{y}$ 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 $\max_{\mathbf{y}} \min_{\mathbf{x}} \ \mathbf{x}^\top A\mathbf{y}$.

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

Repeated Play

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

Our goal: prove $\min_{\mathbf{x}} \max_{\mathbf{y}} \ \mathbf{x}^\top A\mathbf{y} \leq \max_{\mathbf{y}} \min_{\mathbf{x}} \ \mathbf{x}^\top A\mathbf{y}$ 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) $\frac{1}{T}\sum_{t=1}^T \mathbf{x}_t^\top A\mathbf{y}_t \leq \max_{\mathbf{y}\in\Delta_n} \min_{\mathbf{x}\in\Delta_m} \mathbf{x}^\top A\mathbf{y} + \frac{\text{Reg}_T^{\mathbf{x}}}{T}$ (2) $-\frac{1}{T}\sum_{t=1}^T \mathbf{x}_t^\top A\mathbf{y}_t \leq -\min_{\mathbf{x}\in\Delta_m} \max_{\mathbf{y}\in\Delta_n} \mathbf{x}^\top A\mathbf{y} + \frac{\text{Reg}_T^{\mathbf{y}}}{T}$

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

If $\text{Reg}_T^{\mathbf{x}}, \text{Reg}_T^{\mathbf{y}}$ are sublinear $o(T)$, the gap becomes to 0 when $T \to \infty$. □

Repeated Play

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

Theorem 1. *A pair of mixed strategy* $(\mathbf{x}, \mathbf{y})$ *is a Nash equilibrium* ***if and only if*** *it is also a minimax solution, i.e., 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.,* $\mathbf{x}^\star \in \arg\min_{\mathbf{x}} \max_{\mathbf{y}} \ \mathbf{x}^\top A\mathbf{y}, \mathbf{y}^\star \in \arg\max_{\mathbf{y}} \min_{\mathbf{x}} \ \mathbf{x}^\top A\mathbf{y}$.

For simplicity, we denote by $(\mathbf{x}^\star, \mathbf{y}^\star)$ a Nash equilibrium, i.e., a minimax solution.

Proof: *(Nash ⇒ minimax solution)*

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

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

Repeated Play

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

> **Theorem 1.** *A pair of mixed strategy* $(\mathbf{x}, \mathbf{y})$ *is a Nash equilibrium* ***if and only if*** *it is also a minimax solution, i.e., 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.,* $\mathbf{x}^\star \in \arg\min_{\mathbf{x}} \max_{\mathbf{y}} \ \mathbf{x}^\top A\mathbf{y}, \mathbf{y}^\star \in \arg\max_{\mathbf{y}} \min_{\mathbf{x}} \ \mathbf{x}^\top A\mathbf{y}$.

For simplicity, we denote by $(\mathbf{x}^\star, \mathbf{y}^\star)$ a Nash equilibrium, i.e., a minimax solution.

Proof: *(minimax solution ⇒ Nash)*

$$\min_{\mathbf{x}} \max_{\mathbf{y}} \ \mathbf{x}^\top A\mathbf{y} \underset{\text{(minimax)}}{=} \max_{\mathbf{y}} \ \mathbf{x}^{\star\top} A\mathbf{y} \geq \mathbf{x}^{\star\top} A\mathbf{y}^\star \geq \min_{\mathbf{x}} \ \mathbf{x}^\top A\mathbf{y}^\star \underset{\text{(minimax)}}{=} \max_{\mathbf{y}} \min_{\mathbf{x}} \mathbf{x}^\top A\mathbf{y}$$

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

Repeated Play

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

> **Theorem 1.** *A pair of mixed strategy* $(\mathbf{x}, \mathbf{y})$ *is a Nash equilibrium* ***if and only if*** *it is also a minimax solution, i.e., 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.,* $\mathbf{x}^\star \in \arg\min_{\mathbf{x}} \max_{\mathbf{y}} \ \mathbf{x}^\top A\mathbf{y}, \mathbf{y}^\star \in \arg\max_{\mathbf{y}} \min_{\mathbf{x}} \ \mathbf{x}^\top A\mathbf{y}$.

For simplicity, we denote by $(\mathbf{x}^\star, \mathbf{y}^\star)$ a Nash equilibrium, i.e., a minimax solution.

- **Existence** of Nash equilibrium

A natural question: is there always a Nash equilibrium?

Since minximax solution always exist, by Theorem 1, Nash equilibrium also ***always exists***.

Repeated Play

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

The answer already lies in the proof of Von Neumann's minimax theorem.

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

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

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

(3) player-x and player-y submit their strategies together

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

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

Repeated Play

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

From previous analysis, we know that

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

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

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 player-x and player-y 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 Adaptivity

- Can we do **faster**?

Yes! The answer is Optimistic Online Mirror Descent (OOMD).

If player-x runs OOMD with gradients $\mathbf{g}_1^{\mathbf{x}} \triangleq A\mathbf{y}_1, \ldots, \mathbf{g}_T^{\mathbf{x}} \triangleq A\mathbf{y}_T$:

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

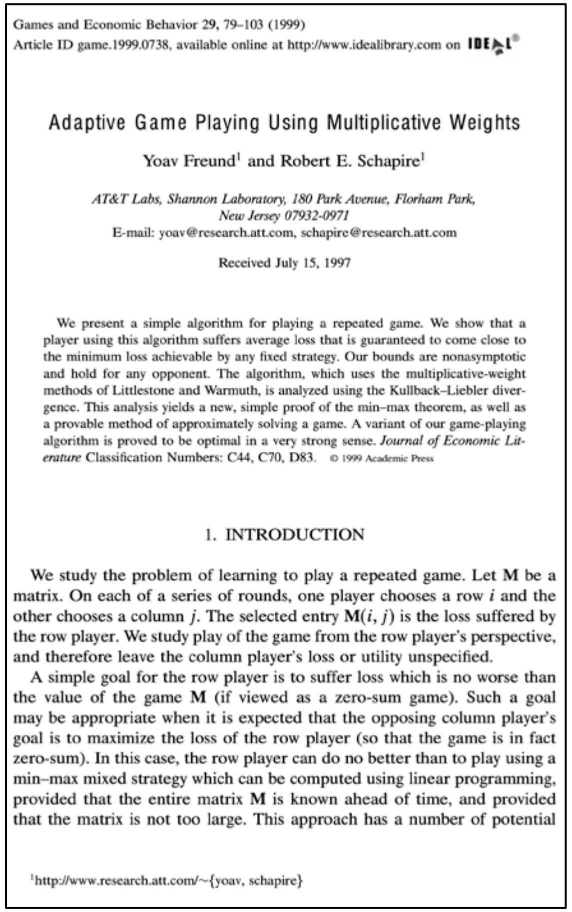
$$\mathrm{Reg}_T^{\mathbf{y}} = \sum_{t=1}^{T} \langle -A\mathbf{x}_t, \mathbf{y}_t - \mathbf{y} \rangle \lesssim \frac{1}{\eta^{\mathbf{y}}} + \eta^{\mathbf{y}} \sum_{t=2}^{T} \|A\mathbf{x}_t - A\mathbf{x}_{t-1}\|_\infty^2 - \frac{1}{\eta^{\mathbf{y}}} \sum_{t=2}^{T} \|\mathbf{y}_t - \mathbf{y}_{t-1}\|^2$$

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

History bits: online learning in games

- **Yoav Freund & Robert E. Schapire**

Yoav Freund and Robert E. Schapire's paper in 1999 reveals the 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. © 1999 Academic Press

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 E. Schapire
1963-now

Yoav Freund
1961-now

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

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.

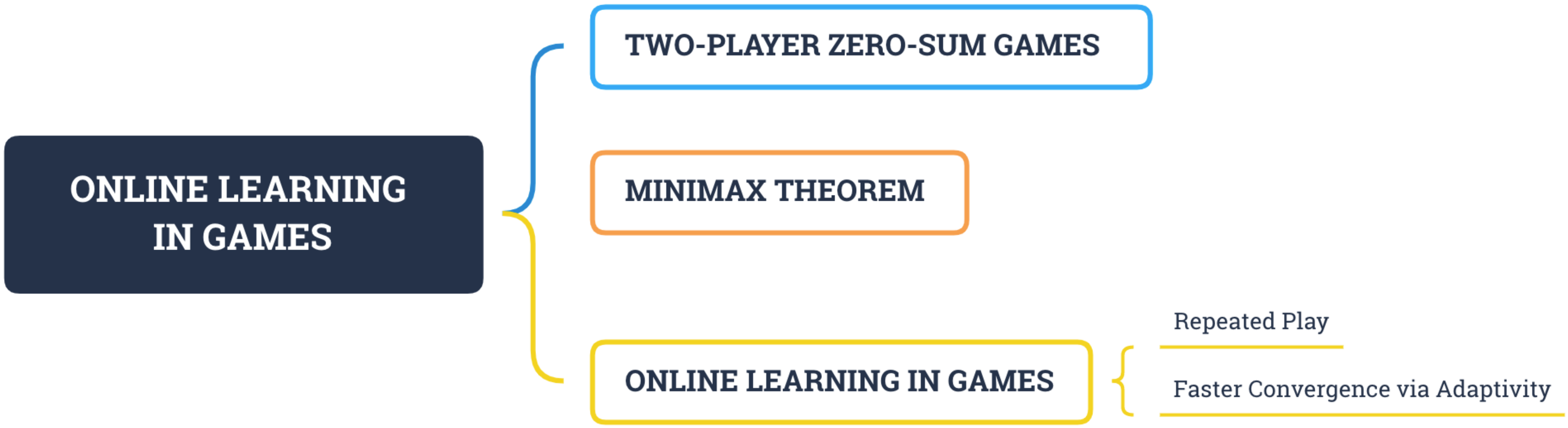
However, the players in our setting are facing other similar, predictable no-regret learning algorithms, a chink that hints at the possibility of improved convergence rates for such dynamics. This was first observed and exploited by Daskalakis et al. [4]. For two-player zero-sum games, they developed 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

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

Fast convergence of regularized learning in games. NIPS 2015.

Summary



Q & A

Thanks!