

Online Learning and Game Theory: Regret Lower Bounds and Adaptive Learning Dynamics

Taira Tsuchiya

The University of Tokyo & RIKEN

May 7–8, 2026

ELLIS Milan + RIKEN AIP Joint ELIAS Workshop, Milano

Multiple players interact in a shared environment,
each adapting their strategy to maximize their cumulative reward.

Multiple players interact in a shared environment, each adapting their strategy to maximize their cumulative reward.

Broader applications

- Minimax optimization, finding equilibria of games
- Variational inequalities

Multiple players interact in a shared environment, each adapting their strategy to maximize their cumulative reward.

Broader applications

- Minimax optimization, finding equilibria of games
- Variational inequalities
- Multi-agent reinforcement learning
- Superhuman AI for poker, Go, Stratego, ...
- Alignment of LLMs







Two-player zero-sum games

- x -player and y -player, each having m_x , m_y actions
- characterized by a **payoff matrix** $A \in [-1, 1]^{m_x \times m_y}$

Two-player zero-sum games

- x -player and y -player, each having m_x , m_y actions
- characterized by a **payoff matrix** $A \in [-1, 1]^{m_x \times m_y}$

Example. Rock-Paper-Scissors

		y-player		
				
x-player		0, 0	1, -1	-1, 1
		-1, 1	0, 0	1, -1
		1, -1	-1, 1	0, 0

payoff matrix of the game

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

Two-player zero-sum games

- x -player and y -player, each having m_x , m_y actions
- characterized by a **payoff matrix** $A \in [-1, 1]^{m_x \times m_y}$

1. x -player chooses **strategy** $x \in \Delta_{m_x}$ and y -player chooses strategy $y \in \Delta_{m_y}$.

(Δ_m : $(m - 1)$ -dimensional probability simplex)

Two-player zero-sum games

- x -player and y -player, each having m_x, m_y actions
- characterized by a **payoff matrix** $A \in [-1, 1]^{m_x \times m_y}$

1. x -player chooses **strategy** $x \in \Delta_{m_x}$ and y -player chooses strategy $y \in \Delta_{m_y}$.

(Δ_m : $(m - 1)$ -dimensional probability simplex)

2. x -player gains reward $\langle x, Ay \rangle$, and y -player incurs loss $\langle x, Ay \rangle$ (thus zero-sum).

Two-player zero-sum games

- x -player and y -player, each having m_x , m_y actions
- characterized by a **payoff matrix** $A \in [-1, 1]^{m_x \times m_y}$

1. x -player chooses **strategy** $x \in \Delta_{m_x}$ and
 y -player chooses strategy $y \in \Delta_{m_y}$.

(Δ_m : $(m - 1)$ -dimensional probability simplex)

2. x -player gains reward $\langle x, Ay \rangle$, and
 y -player incurs loss $\langle x, Ay \rangle$ (thus zero-sum).

Solution of games?

→ **Nash equilibrium**: a pair of strategies in which no player has an incentive to deviate

Two-player zero-sum games

- x -player and y -player, each having m_x, m_y actions
- characterized by a **payoff matrix** $A \in [-1, 1]^{m_x \times m_y}$

1. x -player chooses **strategy** $x \in \Delta_{m_x}$ and
 y -player chooses strategy $y \in \Delta_{m_y}$.

(Δ_m : $(m - 1)$ -dimensional probability simplex)

2. x -player gains reward $\langle x, Ay \rangle$, and
 y -player incurs loss $\langle x, Ay \rangle$ (thus zero-sum).

Solution of games?

→ **Nash equilibrium**: a pair of strategies in which no player has an incentive to deviate

payoff matrix A is unknown before the game starts

→ learning in games!

Learning in two-player zero-sum games

A **sequential** formulation of two-player zero-sum games, characterized by an **unknown** payoff matrix $A \in [-1, 1]^{m_x \times m_y}$

(m_x, m_y : the number of actions of x - and y -players)

At each round $t = 1, \dots, T$:

(Δ_m : $(m - 1)$ -dimensional probability simplex)

1. x -player selects a strategy $x_t \in \Delta_{m_x}$ and y -player selects $y_t \in \Delta_{m_y}$;

Learning in two-player zero-sum games

A **sequential** formulation of two-player zero-sum games, characterized by an **unknown** payoff matrix $A \in [-1, 1]^{m_x \times m_y}$

(m_x, m_y : the number of actions of x - and y -players)

At each round $t = 1, \dots, T$:

(Δ_m : $(m - 1)$ -dimensional probability simplex)

1. x -player selects a strategy $x_t \in \Delta_{m_x}$ and y -player selects $y_t \in \Delta_{m_y}$;
2. x -player gains reward $\langle x_t, Ay_t \rangle$ and y -player incurs loss $\langle x_t, Ay_t \rangle$; **(thus zero-sum)**

Learning in two-player zero-sum games

A **sequential** formulation of two-player zero-sum games, characterized by an **unknown** payoff matrix $A \in [-1, 1]^{m_x \times m_y}$

(m_x, m_y : the number of actions of x - and y -players)

At each round $t = 1, \dots, T$:

(Δ_m : $(m - 1)$ -dimensional probability simplex)

1. x -player selects a strategy $x_t \in \Delta_{m_x}$ and y -player selects $y_t \in \Delta_{m_y}$;
2. x -player gains reward $\langle x_t, Ay_t \rangle$ and y -player incurs loss $\langle x_t, Ay_t \rangle$; **(thus zero-sum)**
3. x -player observes reward vector $g_t = Ay_t$ (**gain**) and y -player observes loss vector $\ell_t = A^\top x_t$;

Learning in two-player zero-sum games

A **sequential** formulation of two-player zero-sum games, characterized by an **unknown** payoff matrix $A \in [-1, 1]^{m_x \times m_y}$

(m_x, m_y : the number of actions of x - and y -players)

At each round $t = 1, \dots, T$:

(Δ_m : $(m - 1)$ -dimensional probability simplex)

1. x -player selects a strategy $x_t \in \Delta_{m_x}$ and y -player selects $y_t \in \Delta_{m_y}$;
2. x -player gains reward $\langle x_t, Ay_t \rangle$ and y -player incurs loss $\langle x_t, Ay_t \rangle$; **(thus zero-sum)**
3. x -player observes reward vector $g_t = Ay_t$ (**gain**) and y -player observes loss vector $l_t = A^\top x_t$;

$$\min_x \max_y \langle x, Ay \rangle$$

Learning in two-player zero-sum games

A **sequential** formulation of two-player zero-sum games, characterized by an **unknown** payoff matrix $A \in [-1, 1]^{m_x \times m_y}$

(m_x, m_y : the number of actions of x - and y -players)

At each round $t = 1, \dots, T$:

(Δ_m : $(m - 1)$ -dimensional probability simplex)

1. x -player selects a strategy $x_t \in \Delta_{m_x}$ and y -player selects $y_t \in \Delta_{m_y}$;
2. x -player gains reward $\langle x_t, Ay_t \rangle$ and y -player incurs loss $\langle x_t, Ay_t \rangle$; (**thus zero-sum**)
3. x -player observes reward vector $g_t = Ay_t$ (**gain**) and y -player observes loss vector $\ell_t = A^\top x_t$;

$$\min_x \max_y \langle x, Ay \rangle$$

Goal of x -/ y - players: maximize their **cumulative reward** (without knowing A):

- x -player: maximize $\sum_{t=1}^T \langle x_t, Ay_t \rangle$,
- y -player: minimize $\sum_{t=1}^T \langle x_t, Ay_t \rangle$.

Learning in two-player zero-sum games

A **sequential** formulation of two-player zero-sum games, characterized by an **unknown** payoff matrix $A \in [-1, 1]^{m_x \times m_y}$

(m_x, m_y : the number of actions of x - and y -players)

At each round $t = 1, \dots, T$:

(Δ_m : $(m - 1)$ -dimensional probability simplex)

1. x -player selects a strategy $x_t \in \Delta_{m_x}$ and y -player selects $y_t \in \Delta_{m_y}$;
2. x -player gains reward $\langle x_t, Ay_t \rangle$ and y -player incurs loss $\langle x_t, Ay_t \rangle$; **(thus zero-sum)**
3. x -player observes reward vector $g_t = Ay_t$ (**gain**) and y -player observes loss vector $\ell_t = A^\top x_t$;

$$\min_x \max_y \langle x, Ay \rangle$$

Goal of x -/ y - players: maximize their **cumulative reward** (without knowing A):

- x -player: maximize $\sum_{t=1}^T \langle x_t, Ay_t \rangle$,
- y -player: minimize $\sum_{t=1}^T \langle x_t, Ay_t \rangle$.

Deep connection with online learning!

At each round $t = 1, \dots, T$:

(Δ_m : $(m - 1)$ -dimensional probability simplex)

1. x -player selects strategy $x_t \in \Delta_{m_x}$ and y -player selects $y_t \in \Delta_{m_y}$;
2. x -player gains reward $\langle x_t, Ay_t \rangle = \langle x_t, g_t \rangle$ and
 y -player incurs loss $\langle x_t, Ay_t \rangle = \langle y_t, \ell_t \rangle$; **(thus zero-sum)**
3. x -player observes reward vector $g_t = Ay_t$ (**gain**) and
 y -player observes loss vector $\ell_t = A^\top x_t$;

Learning in games as online linear optimization

At each round $t = 1, \dots, T$:

(Δ_m : $(m - 1)$ -dimensional probability simplex)

1. x-player selects strategy $x_t \in \Delta_{m_x}$ and y-player selects $y_t \in \Delta_{m_y}$;
2. x-player gains reward $\langle x_t, Ay_t \rangle = \langle x_t, g_t \rangle$ and
y-player incurs loss $\langle x_t, Ay_t \rangle = \langle y_t, l_t \rangle$; **(thus zero-sum)**
3. x-player observes reward vector $g_t = Ay_t$ (gain) and
y-player observes loss vector $l_t = A^\top x_t$;

Each player solves **online linear optimization over the probability simplex** with regret

- $\text{Reg}_T^x = \max_{x^* \in \Delta_{m_x}} \left\{ \sum_{t=1}^T \langle x^*, g_t \rangle - \sum_{t=1}^T \langle x_t, g_t \rangle \right\}$,
- $\text{Reg}_T^y = \max_{y^* \in \Delta_{m_y}} \left\{ \sum_{t=1}^T \langle y_t, l_t \rangle - \sum_{t=1}^T \langle y^*, l_t \rangle \right\}$.

Theorem (Freund and Schapire 1999)

Let $\bar{x}_T = \frac{1}{T} \sum_{t=1}^T x_t$ and $\bar{y}_T = \frac{1}{T} \sum_{t=1}^T y_t$ be the average plays.

(\bar{x}_T, \bar{y}_T) is a $\frac{\text{Reg}_T^x + \text{Reg}_T^y}{T}$ -approximate Nash equilibrium.

Theorem (Freund and Schapire 1999)

Let $\bar{x}_T = \frac{1}{T} \sum_{t=1}^T x_t$ and $\bar{y}_T = \frac{1}{T} \sum_{t=1}^T y_t$ be the average plays.

(\bar{x}_T, \bar{y}_T) is a $\frac{\text{Reg}_T^x + \text{Reg}_T^y}{T}$ -approximate Nash equilibrium.

Immediate consequence:

$$\text{Reg}_T^x, \text{Reg}_T^y = \tilde{O}(\sqrt{T}) \implies \frac{\text{Reg}_T^x + \text{Reg}_T^y}{T} = \tilde{O}(1/\sqrt{T}).$$

So (\bar{x}_T, \bar{y}_T) converges to a Nash equilibrium at rate $\tilde{O}(1/\sqrt{T})$ under uncoupled dynamics.

Theorem (Freund and Schapire 1999)

Let $\bar{x}_T = \frac{1}{T} \sum_{t=1}^T x_t$ and $\bar{y}_T = \frac{1}{T} \sum_{t=1}^T y_t$ be the average plays.

(\bar{x}_T, \bar{y}_T) is a $\frac{\text{Reg}_T^x + \text{Reg}_T^y}{T}$ -approximate Nash equilibrium.

Immediate consequence:

$$\text{Reg}_T^x, \text{Reg}_T^y = \tilde{O}(\sqrt{T}) \implies \frac{\text{Reg}_T^x + \text{Reg}_T^y}{T} = \tilde{O}(1/\sqrt{T}).$$

So (\bar{x}_T, \bar{y}_T) converges to a Nash equilibrium at rate $\tilde{O}(1/\sqrt{T})$ under uncoupled dynamics.

Q. Is this the optimal rate in learning in games?

Fast convergence in games

Hedge algorithm (recall $g_t = Ay_t$ and $\ell_t = A^\top x_t$):

$\eta_x, \eta_y \simeq 1/\sqrt{T}$: learning rate

$$x_t(i) \propto \exp\left(\eta_x \sum_{s=1}^{t-1} g_s(i)\right) \forall i \in [m_x], \quad y_t(i) \propto \exp\left(-\eta_y \sum_{s=1}^{t-1} \ell_s(i)\right) \forall i \in [m_y]$$

cumulative gain
of action i
cumulative loss
of action i

$\rightarrow \tilde{O}(1/\sqrt{T})$ convergence rate to Nash equilibrium ☹ ☹

Fast convergence in games

Hedge algorithm (recall $g_t = Ay_t$ and $\ell_t = A^\top x_t$):

$\eta_x, \eta_y \simeq 1/\sqrt{T}$: learning rate

$$x_t(i) \propto \exp\left(\eta_x \sum_{s=1}^{t-1} g_s(i)\right) \quad \forall i \in [m_x], \quad y_t(i) \propto \exp\left(-\eta_y \sum_{s=1}^{t-1} \ell_s(i)\right) \quad \forall i \in [m_y]$$

cumulative gain
of action i
cumulative loss
of action i

$\rightarrow \tilde{O}(1/\sqrt{T})$ convergence rate to Nash equilibrium ☹️ ☹️

Optimistic Hedge algorithm (Rakhlin and Sridharan 2013; Syrgkanis et al. 2015):

$$x_t(i) \propto \exp\left(\eta_x \left(\sum_{s=1}^{t-1} g_s(i) + g_{t-1}(i)\right)\right), \quad y_t(i) \propto \exp\left(-\eta_y \left(\sum_{s=1}^{t-1} \ell_s(i) + \ell_{t-1}(i)\right)\right)$$

Fast convergence in games

Optimistic Hedge algorithm (Rakhlin and Sridharan 2013; Syrgkanis et al. 2015):

$$x_t(i) \propto \exp\left(\eta_x \left(\sum_{s=1}^{t-1} g_s(i) + g_{t-1}(i)\right)\right), \quad y_t(i) \propto \exp\left(-\eta_y \left(\sum_{s=1}^{t-1} \ell_s(i) + \ell_{t-1}(i)\right)\right)$$

Theorem (Rakhlin and Sridharan 2013)

If the x - and y -players follow optimistic Hedge with learning rates $\eta_x = \eta_y = 1/4$, then

$$\text{Reg}_T^x = O(\log(m_x m_y)), \quad \text{Reg}_T^y = O(\log(m_x m_y))$$

which implies an

$O(\log(m_x m_y)/T)$ convergence rate to a Nash equilibrium.

Fast convergence in games

Optimistic Hedge algorithm (Rakhlin and Sridharan 2013; Syrgkanis et al. 2015):

$$x_t(i) \propto \exp\left(\eta_x \left(\sum_{s=1}^{t-1} g_s(i) + g_{t-1}(i)\right)\right), \quad y_t(i) \propto \exp\left(-\eta_y \left(\sum_{s=1}^{t-1} \ell_s(i) + \ell_{t-1}(i)\right)\right)$$

Theorem (Rakhlin and Sridharan 2013)

If the x - and y -players follow optimistic Hedge with learning rates $\eta_x = \eta_y = 1/4$, then

$$\text{Reg}_T^x = O(\log(m_x m_y)), \quad \text{Reg}_T^y = O(\log(m_x m_y))$$

which implies an

$$O(\log(m_x m_y)/T) \text{ convergence rate to a Nash equilibrium.}$$

Rough intuition: If the opponent uses a no-regret algorithm, then we can predict the opponent's next strategy y_{t+1} (and thus gradient $g_{t+1} = Ay_{t+1}$).

Natural questions

Theorem (Rakhlin and Sridharan 2013)

If the x - and y -players *fully* follow optimistic Hedge with **constant** learning rates $\eta_x = \eta_y = 1/4$ in games with $A \in [-1, 1]^{m_x \times m_y}$, then we obtain an

$O(\log(m_x m_y)/T)$ -approximate Nash equilibrium

after T rounds.

Optimal in T ! (Daskalakis, Deckelbaum, and Kim 2011)

This result looks great, but ...

Q1 Is the dependence on the number of actions m_x, m_y optimal?

Natural questions

Theorem (Rakhlin and Sridharan 2013)

If the x - and y -players **fully** follow optimistic Hedge with **constant** learning rates $\eta_x = \eta_y = 1/4$ in games with $A \in [-1, 1]^{m_x \times m_y}$, then we obtain an

$O(\log(m_x m_y)/T)$ -approximate Nash equilibrium

after T rounds.

Optimal in T ! (Daskalakis, Deckelbaum, and Kim 2011)

This result looks great, but ...

Q1 Is the dependence on the number of actions m_x, m_y optimal?

Q2 What if the opponent deviates from optimistic Hedge with a constant learning rate?

Natural questions

Theorem (Rakhlin and Sridharan 2013)

If the x - and y -players *fully* follow optimistic Hedge with **constant** learning rates $\eta_x = \eta_y = 1/4$ in games with $A \in [-1, 1]^{m_x \times m_y}$, then we obtain an

$O(\log(m_x m_y)/T)$ -approximate Nash equilibrium

after T rounds.

Optimal in T ! (Daskalakis, Deckelbaum, and Kim 2011)

This result looks great, but ...

Q1 Is the dependence on the number of actions m_x, m_y optimal?

Q2 What if the opponent deviates from optimistic Hedge with a constant learning rate?

Q3 The algorithm assumes a known payoff scale, $A \in [-1, 1]^{m_x \times m_y}$. Can we achieve fast convergence without knowing the scale?

Natural questions

Theorem (Rakhlin and Sridharan 2013)

If the x - and y -players *fully* follow optimistic Hedge with **constant** learning rates $\eta_x = \eta_y = 1/4$ in games with $A \in [-1, 1]^{m_x \times m_y}$, then we obtain an

$O(\log(m_x m_y)/T)$ -approximate Nash equilibrium

after T rounds.

Optimal in T ! (Daskalakis, Deckelbaum, and Kim 2011)

This result looks great, but ...

Q1 Is the dependence on the number of actions m_x, m_y optimal?

[Tsuchiya, AISTATS2026 Spotlight]

Q2 What if the opponent deviates from optimistic Hedge with a constant learning rate?

[Tsuchiya-Ito-Luo, COLT2025]

Q3 The algorithm assumes a known payoff scale, $A \in [-1, 1]^{m_x \times m_y}$. Can we achieve fast convergence without knowing the scale?

[Tsuchiya-Luo-Ito, 2026]

Q1. Optimal dependence on the number of actions?

(recap) Optimistic Hedge with constant learning rates gives

$O(\log(m_x m_y) / T)$ -approximate Nash equilibrium

Q1. Optimal dependence on the number of actions?

(recap) Optimistic Hedge with constant learning rates gives

$O(\log(m_x m_y)/T)$ -approximate Nash equilibrium

Main results (informal, [T. AISTATS'26, Spotlight]):

- This can be improved to

$$O(\sqrt{\log m_x \log m_y}/T).$$

→ Particularly important when m_x and m_y are imbalanced

Q1. Optimal dependence on the number of actions?

(recap) Optimistic Hedge with constant learning rates gives

$O(\log(m_x m_y)/T)$ -approximate Nash equilibrium

Main results (informal, [T. AISTATS'26, Spotlight]):

- This can be improved to

$$O(\sqrt{\log m_x \log m_y}/T).$$

→ Particularly important when m_x and m_y are imbalanced

- This is optimal for optimistic Hedge with constant learning rates:

$$\text{Reg}_T^x = \Omega(\sqrt{\log m_x \log m_y}), \quad \text{Reg}_T^y = \Omega(\sqrt{\log m_x \log m_y}).$$

Q1. Optimal dependence on the number of actions?

(recap) Optimistic Hedge with constant learning rates gives

$O(\log(m_x m_y)/T)$ -approximate Nash equilibrium

Main results (informal, [T. AISTATS'26, Spotlight]):

- This can be improved to

$$O(\sqrt{\log m_x \log m_y}/T).$$

→ Particularly important when m_x and m_y are imbalanced

- This is optimal for optimistic Hedge with constant learning rates:

$$\text{Reg}_T^x = \Omega(\sqrt{\log m_x \log m_y}), \quad \text{Reg}_T^y = \Omega(\sqrt{\log m_x \log m_y}).$$

→ (Partially) resolves a question raised by

Anagnostides–Kalavasis–Sandholm–Zampetakis (2024) on the optimal dependence on the number of actions for optimistic Hedge.

Q1. Optimal dependence on the number of actions?

Also in the paper

- Refined upper bound via convex reformulation over (η, η', c, c') .
- Dynamic regret (connection with last-iterate convergence):
near-matching $\sqrt{\log m_x \log m_y} \cdot \log T$ regret upper and lower bounds.
- Extension to Hölder-smooth convex–concave games.

Q1. Optimal dependence on the number of actions?

Also in the paper

- Refined upper bound via convex reformulation over (η, η', c, c') .
- Dynamic regret (connection with last-iterate convergence):
near-matching $\sqrt{\log m_x \log m_y} \cdot \log T$ regret upper and lower bounds.
- Extension to Hölder-smooth convex-concave games.

Future work

- Learning-rate-independent lower bounds for optimistic Hedge.
- Algorithm-independent lower bounds for general strongly-uncoupled dynamics.

Q2. What if the opponent deviates?

Continuing with optimistic Hedge with constant learning rate may lead to a linear regret

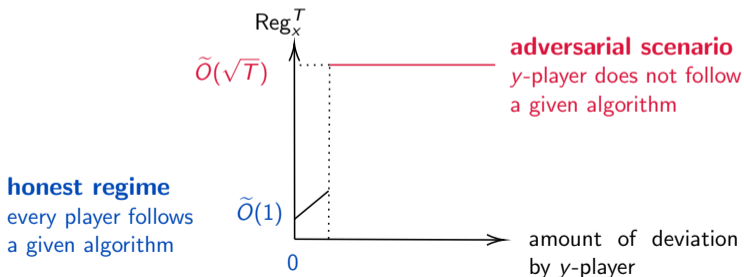
Q2. What if the opponent deviates?

Continuing with optimistic Hedge with constant learning rate may lead to a linear regret
Known solution (Syrkkanis et al. 2015): Monitor gradient variation $\sum_{s=1}^{t-1} \|g_s - g_{s+1}\|_1^2$, and if it exceeds a threshold, switch to an algorithm with a worst-case regret of $\tilde{O}(\sqrt{T})$

Q2. What if the opponent deviates?

Continuing with optimistic Hedge with constant learning rate may lead to a linear regret
 Known solution (Syrkkanis et al. 2015): Monitor gradient variation $\sum_{s=1}^{t-1} \|g_s - g_{s+1}\|_1^2$, and if it exceeds a threshold, switch to an algorithm with a worst-case regret of $\tilde{O}(\sqrt{T})$

Discontinuous behavior: A slight deviation of the y -player from a given algorithm can suddenly cause the x -player to suffer a regret of $O(\sqrt{T})$ 😞 😞



Q2. What if the opponent deviates? (cont'd)

Main results (informal, [T.-Ito-Luo COLT'25]):

There exists a learning dynamic robust against deviations by the opponent. In particular,

- Show that there exists a learning dynamic such that

$$\text{Reg}_T^x = \tilde{O}\left(\sqrt{1 + C_y} + C_x\right), \quad \text{Reg}_T^y = \tilde{O}\left(\sqrt{1 + C_x} + C_y\right).$$

y-player's cumulative
deviation

x-player's cumulative
deviation

Q2. What if the opponent deviates? (cont'd)

Main results (informal, [T.-Ito-Luo COLT'25]):

There exists a learning dynamic robust against deviations by the opponent. In particular,

- Show that there exists a learning dynamic such that

$$\text{Reg}_T^x = \tilde{O}\left(\sqrt{1 + C_y} + C_x\right), \quad \text{Reg}_T^y = \tilde{O}\left(\sqrt{1 + C_x} + C_y\right).$$

y-player's cumulative
deviation

x-player's cumulative
deviation

- Show that the above bounds are optimal

$$\text{Reg}_T^x = \Omega\left(\sqrt{1 + C_y} + C_x\right), \quad \text{Reg}_T^y = \Omega\left(\sqrt{1 + C_x} + C_y\right).$$

Q3. Is scale-free fast convergence possible?

(recap) Optimistic Hedge with constant learning rates gives

$O(A_{\max} \log(m_x m_y) / T)$ -approximate Nash equilibrium if A_{\max} is known

Q3. Is scale-free fast convergence possible?

(recap) Optimistic Hedge with constant learning rates gives

$O(A_{\max} \log(m_x m_y) / T)$ -approximate Nash equilibrium if A_{\max} is known

Scale-free learning dynamic (informal):

- no prior knowledge of the payoff scale is needed; and
- if payoffs are rescaled by any constant $c > 0$, the strategy sequence is unchanged.
(A key property behind the success of regret matching!)

Q3. Is scale-free fast convergence possible?

(recap) Optimistic Hedge with constant learning rates gives

$O(A_{\max} \log(m_x m_y) / T)$ -approximate Nash equilibrium if A_{\max} is known

Scale-free learning dynamic (informal):

- no prior knowledge of the payoff scale is needed; and
- if payoffs are rescaled by any constant $c > 0$, the strategy sequence is unchanged. (A key property behind the success of regret matching!)

Main result (informal, from [T.-Luo-Ito '26]):

There exists a scale-free learning dynamic such that for any payoff matrix $A \in [-A_{\max}, A_{\max}]^{m_x \times m_y}$, it gives

$\tilde{O}(A_{\max} / T)$ -approximate Nash equilibrium

after T rounds.

Q2, Q3

More results in the papers

Extension to multiplayer general-sum games

→ convergence to correlated equilibrium (CE) via swap regret

Q2, Q3

More results in the papers

Extension to multiplayer general-sum games

→ convergence to correlated equilibrium (CE) via swap regret

Technique

For Q2:

- corrupted learning dynamics in multiplayer general-sum games
- stability analysis of Markov-chain stationary distributions for optimistic FTRL with adaptive learning rates
- (dealing with adversarial corruption of the gradient feedback)

Q2, Q3

More results in the papers

Extension to multiplayer general-sum games

→ convergence to correlated equilibrium (CE) via swap regret

Technique

For Q2:

- corrupted learning dynamics in multiplayer general-sum games
- stability analysis of Markov-chain stationary distributions for optimistic FTRL with adaptive learning rates
- (dealing with adversarial corruption of the gradient feedback)

For Q3:

- Fast convergence to CE without knowing the scale
- A new stopping-time analysis to exploit negative terms without knowing the scale
- **doubling clipping**: A new scale-free learning algorithm for online linear optimization useful in swap regret minimization

Takeaway

Fast convergence in learning in games can be made
dimension-efficient • deviation-robust • scale-free

Thanks for listening!

References I

-  Daskalakis, Constantinos, Alan Deckelbaum, and Anthony Kim (2011). “Near-optimal no-regret algorithms for zero-sum games”. In: *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*. Society for Industrial and Applied Mathematics, pp. 235–254.
-  Freund, Yoav and Robert E. Schapire (1999). “Adaptive Game Playing Using Multiplicative Weights”. In: *Games and Economic Behavior* 29.1, pp. 79–103.
-  Rakhlin, Sasha and Karthik Sridharan (2013). “Optimization, Learning, and Games with Predictable Sequences”. In: *Advances in Neural Information Processing Systems*. Vol. 26. Curran Associates, Inc., pp. 3066–3074.
-  Syrgkanis, Vasilis et al. (2015). “Fast Convergence of Regularized Learning in Games”. In: *Advances in Neural Information Processing Systems*. Vol. 28. Curran Associates, Inc., pp. 2989–2997.