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

Taira Tsuchiya

The University of Tokyo & RIKEN

Feb 27, 2026

UTokyo Joint Seminar on Machine Learning Theory

Learning in games

Multiple players interact in a shared environment,
each aiming to maximize their cumulative reward
by iteratively adapting their strategies based on repeated interactions

Learning in games

Multiple players interact in a shared environment,
each aiming to maximize their cumulative reward
by iteratively adapting their strategies based on repeated interactions

Broader applications

- Minimax optimization, finding equilibrium of games
- variational inequalities

Learning in games

Multiple players interact in a shared environment,
each aiming to maximize their cumulative reward
by iteratively adapting their strategies based on repeated interactions

Broader applications

- Minimax optimization, finding equilibrium of games
- variational inequalities
- Multi-agent reinforcement learning
- Superhuman AI for ppoker, Go, Stratego, ...
- Alignment of LLMs

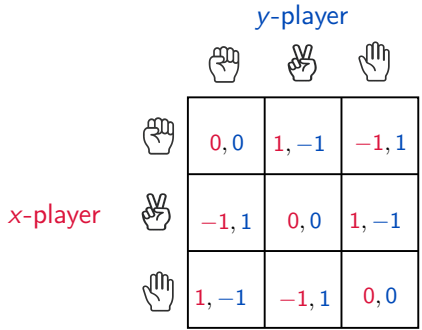
Two-player zero-sum games

- 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



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

A pair of probability distributions (x^*, y^*) over action sets $[m_x]$ and $[m_y]$ is an ε-**approximate Nash equilibrium** if

$$x^\top A y^* - \varepsilon \leq x^{*\top} A y^* \leq x^{*\top} A y + \varepsilon \quad \forall x \in \Delta_{m_x}, y \in \Delta_{m_y}.$$

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 $t = 1, \ldots, T$: (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 $t = 1, \ldots, T$: (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 $\langle x_t, A y_t \rangle$ and
   y-player incurs loss $\langle x_t, A y_t \rangle$; **(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 $t = 1, \ldots, T$: (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 $\langle x_t, A y_t \rangle$ and
   y-player incurs loss $\langle x_t, A y_t \rangle$; **(thus zero-sum)**
3. x-player observes reward vector $g_t = A y_t$ (**g**ain) 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 $t = 1, \ldots, T$: (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 $\langle x_t, A y_t \rangle$ and
   y-player incurs loss $\langle x_t, A y_t \rangle$; **(thus zero-sum)**
3. x-player observes reward vector $g_t = A y_t$ (**g**ain) and
   y-player observes loss vector $\ell_t = A^\top x_t$;

$\min_x \max_y \langle x, A y \rangle$

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 $t = 1, \ldots, T$: (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 $g_t = Ay_t$ (**g**ain) 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 $t = 1, \ldots, T$: (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 $g_t = Ay_t$ (**g**ain) 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!

Online convex optimization

A sequential formulation of (standard) convex optimization $\min_{z \in K} f(z)$

Online convex optimization

A sequential formulation of (standard) convex optimization $\min_{z \in K} f(z)$

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

1. Player chooses z_t from feasible set $K \subset \mathbb{R}^d$

Online convex optimization

A sequential formulation of (standard) convex optimization $\min_{z \in K} f(z)$

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

1. Player chooses z_t from feasible set $K \subset \mathbb{R}^d$
2. Environment chooses **convex loss function** $f_t \colon K \to \mathbb{R}$

Online convex optimization

A sequential formulation of (standard) convex optimization $\min_{z \in K} f(z)$

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

1. Player chooses z_t from feasible set $K \subset \mathbb{R}^d$
2. Environment chooses **convex loss function** $f_t \colon K \to \mathbb{R}$
3. Player incurs loss $f_t(z_t)$ and observes an element in $\partial f_t(z_t)$

Online convex optimization

A sequential formulation of (standard) convex optimization $\min_{z \in K} f(z)$

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

1. Player chooses z_t from feasible set $K \subset \mathbb{R}^d$
2. Environment chooses **convex loss function** $f_t \colon K \to \mathbb{R}$
3. Player incurs loss $f_t(z_t)$ and observes an element in $\partial f_t(z_t)$

Online convex optimization

A sequential formulation of (standard) convex optimization $\min_{z \in K} f(z)$

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

1. Player chooses z_t from feasible set $K \subset \mathbb{R}^d$
2. Environment chooses **convex loss function** $f_t \colon K \to \mathbb{R}$
3. Player incurs loss $f_t(z_t)$ and observes an element in $\partial f_t(z_t)$

Player's goal: minimize cumulative loss

$$\sum_{t=1}^{T} f_t(z_t)$$

Online convex optimization

A sequential formulation of (standard) convex optimization $\min_{z \in K} f(z)$

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

1. Player chooses z_t from feasible set $K \subset \mathbb{R}^d$
2. Environment chooses **convex loss function** $f_t \colon K \to \mathbb{R}$
3. Player incurs loss $f_t(z_t)$ and observes an element in $\partial f_t(z_t)$

Player's objective: minimize **regret** Reg_T

$$\mathrm{Reg}_T = \sum_{t=1}^{T} f_t(z_t) - \min_{z \in K} \sum_{t=1}^{T} f_t(z)$$

Player's cumulative loss — Best fixed-point cumulative loss

Online convex optimization

A sequential formulation of (standard) convex optimization $\min_{z \in K} f(z)$

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

1. Player chooses z_t from feasible set $K \subset \mathbb{R}^d$
2. Environment chooses **convex loss function** $f_t \colon K \to \mathbb{R}$
3. Player incurs loss $f_t(z_t)$ and observes an element in $\partial f_t(z_t)$

Player's objective: minimize **regret** Reg_T

$$\mathrm{Reg}_T = \sum_{t=1}^{T} f_t(z_t) - \min_{z \in K} \sum_{t=1}^{T} f_t(z)$$

Player's cumulative loss — Best fixed-point cumulative loss

Online convex optimization

A sequential formulation of (standard) convex optimization $\min_{z \in K} f(z)$

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

1. Player chooses z_t from feasible set $K \subset \mathbb{R}^d$
2. Environment chooses **convex loss function** $f_t \colon K \to \mathbb{R}$
3. Player incurs loss $f_t(z_t)$ and observes an element in $\partial f_t(z_t)$

Player's objective: minimize **regret** Reg_T

$$\mathrm{Reg}_T = \underbrace{\sum_{t=1}^{T} f_t(z_t)}_{\text{Player's cumulative loss}} - \underbrace{\min_{z \in K} \sum_{t=1}^{T} f_t(z)}_{\text{Best fixed-point cumulative loss}}$$

- **Online linear optimization**: $f_t(z) = \langle g_t, z \rangle$ for some $g_t \in \mathbb{R}^d$

Online convex optimization

A sequential formulation of (standard) convex optimization $\min_{z \in K} f(z)$

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

1. Player chooses z_t from feasible set $K \subset \mathbb{R}^d$
2. Environment chooses **convex loss function** $f_t \colon K \to \mathbb{R}$
3. Player incurs loss $f_t(z_t)$ and observes an element in $\partial f_t(z_t)$

Player's objective: minimize **regret** Reg_T

$$\mathrm{Reg}_T = \sum_{t=1}^{T} f_t(z_t) - \min_{z \in K} \sum_{t=1}^{T} f_t(z)$$

Player's cumulative loss — Best fixed-point cumulative loss

- **Online linear optimization**: $f_t(z) = \langle g_t, z \rangle$ for some $g_t \in \mathbb{R}^d$
- There exists an algorithm with $\mathrm{Reg}_T = O(\sqrt{T})$, which is optimal.

Learning in games as online linear optimization

At each round $t = 1, \ldots, T$: (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 $g_t = Ay_t$ (**g**ain) and
   y-player observes loss vector $\ell_t = A^\top x_t$;

Learning in games as online linear optimization

At each round $t = 1, \ldots, T$: (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 $g_t = Ay_t$ (**g**ain) and
   y-player observes loss vector $\ell_t = A^\top x_t$;

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

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

No-regret learning dynamics and 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.*

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

No-regret learning dynamics and 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.*

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

Immediate consequence:

$$\mathrm{Reg}_T^x, \mathrm{Reg}_T^y = \widetilde{O}(\sqrt{T}) \quad \Longrightarrow \quad \frac{\mathrm{Reg}_T^x + \mathrm{Reg}_T^y}{T} = \widetilde{O}(1/\sqrt{T}).$$

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

No-regret learning dynamics and 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.*

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

Immediate consequence:

$$\mathrm{Reg}_T^x, \mathrm{Reg}_T^y = \widetilde{O}(\sqrt{T}) \quad \Longrightarrow \quad \frac{\mathrm{Reg}_T^x + \mathrm{Reg}_T^y}{T} = \widetilde{O}(1/\sqrt{T}).$$

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

Q. Is this 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

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

cumulative gain of action i

cumulative loss of action i

$\to \widetilde{O}(1/\sqrt{T})$ 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

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

cumulative gain of action i

cumulative loss of action i

$\to \widetilde{O}(1/\sqrt{T})$ 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

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

which implies an

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

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

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

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

which implies an

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

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

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*

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

after T rounds.

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*

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

after T rounds.

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 the 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*

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

after T rounds.

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 the 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*

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

after T rounds.

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 the 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 (in submission)]

Q1. Optimal dependence on the number of actions?

Particularly important when the number of actions is large (*e.g.*, combinatorial set)

(recap) Optimistic Hedge with constant learning rates (Rakhlin and Sridharan 2013) gives

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

Q1. Optimal dependence on the number of actions?

Particularly important when the number of actions is large (*e.g.*, combinatorial set)

(recap) Optimistic Hedge with constant learning rates (Rakhlin and Sridharan 2013) gives

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

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

- This can be improved to

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

Q1. Optimal dependence on the number of actions?

Particularly important when the number of actions is large (*e.g.*, combinatorial set)

(recap) Optimistic Hedge with constant learning rates (Rakhlin and Sridharan 2013) gives

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

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

- This can be improved to

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

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

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

Q1. Optimal dependence on the number of actions?

Particularly important when the number of actions is large (*e.g.*, combinatorial set)
(recap) Optimistic Hedge with constant learning rates (Rakhlin and Sridharan 2013) gives

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

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

- This can be improved to

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

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

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

  $\rightarrow$ Resolves a question raised by Anagnostides–Kalavasis–Sandholm–Zampetakis (2024) on the optimal dependence on the number of actions for optimistic Hedge.

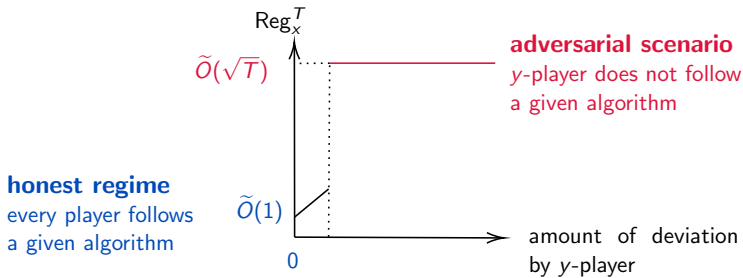
Q2. What if the opponent deviates?

Known solution (Syrgkanis 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 $\widetilde{O}(\sqrt{T})$

Q2. What if the opponent deviates?

Known solution (Syrgkanis 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 $\widetilde{O}(\sqrt{T})$

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 deviation of the opponent. In particular,

- Show that there exists a learning dynamic such that

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

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 deviation of the opponent. In particular,

- Show that there exists a learning dynamic such that

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

y-player's cumulative deviation

x-player's cumulative deviation

- Show that the above bounds are optimal

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

Q3. Is Scale-Free Fast Convergence Possible?

- **Scale-free**: no prior knowledge of the payoff scale is needed.
- **Scale-invariant**: if payoffs are rescaled by any constant $c > 0$, the strategy sequence is unchanged. Key propoerty of great success of regret matching!

Q3. Is Scale-Free Fast Convergence Possible?

- **Scale-free**: no prior knowledge of the payoff scale is needed.
- **Scale-invariant**: if payoffs are rescaled by any constant $c > 0$, the strategy sequence is unchanged. Key propoerty of great success of regret matching!

Main result (informal, from [T.–Luo–Ito 26 (in submission)]):

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

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

after T rounds.

Summary

Takeaway

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

More results in the papers

- Multiplayer general-sum games
  - Fast convergence to correlated equilibrium via swap regret
  - Technical tools: stability analysis of Markov-chain stationary distributions, doubling clipping technique, etc
- Last-iterate convergence
- Dynamic regret

Many interesting open problems remain!

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.