Tight Regret Upper and Lower Bounds for Optimistic Hedge in Two-Player Zero-Sum Games

Taira Tsuchiya

The University of Tokyo & RIKEN

AISTATS 2026 (Spotlight)

Motivation: learning in games

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

Motivation: learning in games

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

Wide range of applications:

- Minimax optimization, variational inequalities
- Multi-agent reinforcement learning
- Superhuman AI for board games
- Alignment of LLMs

Two-player zero-sum games

Payoff matrix $A \in [-1, 1]^{m \times n}$, unknown ($m, n$: numbers of actions of the players).

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

1. x-player chooses $x_t \in \Delta_m$; y-player chooses $y_t \in \Delta_n$.
2. x-player gains reward $\langle x_t, Ay_t \rangle$; y-player incurs loss $\langle x_t, Ay_t \rangle$.
3. x-player observes $g_t = Ay_t$; y-player observes $\ell_t = A^\top x_t$.

Two-player zero-sum games

Payoff matrix $A \in [-1,1]^{m \times n}$, unknown ($m, n$: numbers of actions of the players).

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

1. x-player chooses $x_t \in \Delta_m$; y-player chooses $y_t \in \Delta_n$.
2. x-player gains reward $\langle x_t, A y_t \rangle$; y-player incurs loss $\langle x_t, A y_t \rangle$.
3. x-player observes $g_t = A y_t$; y-player observes $\ell_t = A^\top x_t$.

Goal: minimize external regret,

$$\mathrm{Reg}_T^x = \max_{x^* \in \Delta_m} \sum_{t=1}^{T} \langle x^* - x_t, g_t \rangle \,, \quad \mathrm{Reg}_T^y = \max_{y^* \in \Delta_n} \sum_{t=1}^{T} \langle y_t - y^*, \ell_t \rangle .$$

Two-player zero-sum games

Payoff matrix $A \in [-1,1]^{m \times n}$, unknown ($m, n$: numbers of actions of the players).

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

1. x-player chooses $x_t \in \Delta_m$; y-player chooses $y_t \in \Delta_n$.
2. x-player gains reward $\langle x_t, A y_t \rangle$; y-player incurs loss $\langle x_t, A y_t \rangle$.
3. x-player observes $g_t = A y_t$; y-player observes $\ell_t = A^\top x_t$.

Goal: minimize external regret,

$$\mathrm{Reg}_T^x = \max_{x^* \in \Delta_m} \sum_{t=1}^{T} \langle x^* - x_t, g_t \rangle \,, \quad \mathrm{Reg}_T^y = \max_{y^* \in \Delta_n} \sum_{t=1}^{T} \langle y_t - y^*, \ell_t \rangle .$$

Lemma ([Freund–Schapire, 1999]) The average plays form an $\mathrm{SocialReg}_T / T$-approximate Nash equilibrium, where $\mathrm{SocialReg}_T := \mathrm{Reg}_T^x + \mathrm{Reg}_T^y$ is the *social regret*.

Optimistic Hedge: state of the art

Optimistic Hedge with learning rates $\eta, \eta' > 0$:

$$x_t(i) \propto \exp\Big(\eta\Big(\textstyle\sum_{s=1}^{t-1} g_s(i) + g_{t-1}(i)\Big)\Big), \quad y_t(i) \propto \exp\Big(-\eta'\Big(\textstyle\sum_{s=1}^{t-1} \ell_s(i) + \ell_{t-1}(i)\Big)\Big).$$

= Standard Hedge update + one extra copy of the most recent gradient (one-step optimistic correction)

Optimistic Hedge: state of the art

Optimistic Hedge with learning rates $\eta, \eta' > 0$:

$$x_t(i) \propto \exp\Big(\eta\Big(\textstyle\sum_{s=1}^{t-1} g_s(i) + g_{t-1}(i)\Big)\Big), \quad y_t(i) \propto \exp\Big(-\eta'\Big(\textstyle\sum_{s=1}^{t-1} \ell_s(i) + \ell_{t-1}(i)\Big)\Big).$$

= Standard Hedge update + one extra copy of the most recent gradient (one-step optimistic correction)

Theorem (classical social regret, [Rakhlin–Sridharan, 2013]) With constant learning rates,

$$\mathrm{SocialReg}_T = O(\log(mn))$$

so the average play is an $O(\log(mn)/T)$-approximate Nash equilibrium.

Optimistic Hedge: state of the art

Optimistic Hedge with learning rates $\eta, \eta' > 0$:

$$x_t(i) \propto \exp\Big(\eta\Big(\textstyle\sum_{s=1}^{t-1} g_s(i) + g_{t-1}(i)\Big)\Big), \quad y_t(i) \propto \exp\Big(-\eta'\Big(\textstyle\sum_{s=1}^{t-1} \ell_s(i) + \ell_{t-1}(i)\Big)\Big).$$

= Standard Hedge update + one extra copy of the most recent gradient (one-step optimistic correction)

Theorem (classical social regret, [Rakhlin–Sridharan, 2013]) With constant learning rates,

$$\mathrm{SocialReg}_T = O(\log(mn))$$

so the average play is an $O(\log(mn)/T)$-approximate Nash equilibrium.

The $1/T$ rate is optimal [Daskalakis–Deckelbaum–Kim, 2011]

Q. Is the dependence on m, n really optimal in games?

Question and answers

What is the optimal dependence on the numbers of actions m, n (including the leading constants) for optimistic Hedge?

Question and answers

What is the optimal dependence on the numbers of actions m, n (including the leading constants) for optimistic Hedge?

Two concrete subquestions:

Q1. Can the $O(\log(mn))$ upper bound be further tightened?

Question and answers

What is the optimal dependence on the numbers of actions m, n (including the leading constants) for optimistic Hedge?

Two concrete subquestions:

Q1. Can the $O(\log(mn))$ upper bound be further tightened?

Q2. Are the social-regret bounds optimal for optimistic Hedge?

Q1. A tighter upper bound

cardinality-aware strongly-uncoupled learning:
each player additionally knows the opponent's number of actions

Q1. A tighter upper bound

cardinality-aware strongly-uncoupled learning:
each player additionally knows the opponent's number of actions

Theorem (cardinality-aware social regret) With appropriately tuned learning rates,

$$\mathrm{SocialReg}_T \;\le\; 2\sqrt{\log m\,(\log n + \tfrac{1}{2})} \;+\; 2\sqrt{\log n\,(\log m + \tfrac{1}{2})}.$$

So $O(\log(mn))$ improves to $O(\sqrt{\log m \log n})$.

Q1. A tighter upper bound

cardinality-aware strongly-uncoupled learning:
each player additionally knows the opponent's number of actions

Theorem (cardinality-aware social regret) With appropriately tuned learning rates,

$$\mathrm{SocialReg}_T \;\le\; 2\sqrt{\log m\,(\log n + \tfrac{1}{2})} \;+\; 2\sqrt{\log n\,(\log m + \tfrac{1}{2})}.$$

So $O(\log(mn))$ improves to $O(\sqrt{\log m \log n})$.
The improvement is largest when $\log m$ and $\log n$ are imbalanced, for example

- network interdiction
- extensive-form games
- games with submodular structure
- maximum flow

Q2. A matching lower bound

Theorem (main lower bound) For any $\eta, \eta' > 0$, there exists a game such that

$$\mathrm{Reg}_T^x \;\geq\; \frac{\log m}{\eta} - o(1)\,, \quad \mathrm{Reg}_T^y \;\geq\; \frac{\log n}{\eta'} - o(1).$$

- First direct and unconditional lower bound for optimistic Hedge.
- Complements the conditional bound of [Anagnostides et al., 2024] for multiplayer general-sum games under a complexity-theoretic assumption.

Q2. A matching lower bound

Theorem (main lower bound) For any $\eta, \eta' > 0$, there exists a game such that

$$\mathrm{Reg}_T^x \;\geq\; \frac{\log m}{\eta} - o(1)\,, \quad \mathrm{Reg}_T^y \;\geq\; \frac{\log n}{\eta'} - o(1).$$

- First direct and unconditional lower bound for optimistic Hedge.
- Complements the conditional bound of [Anagnostides et al., 2024] for multiplayer general-sum games under a complexity-theoretic assumption.

Consequence. Tight *including the leading constant* for social regret:

- strongly-uncoupled: $\mathrm{SocialReg}_T \geq 2(\log m + \log n) - o(1)$,
- cardinality-aware: $\mathrm{SocialReg}_T \geq 4\sqrt{\log m \log n} - o(1)$.

Summary

Takeaway

- Cardinality-awareness turns $\log(mn)$ into $\sqrt{\log m \log n}$.
- Optimistic Hedge is tight in the numbers of actions.

Summary

Takeaway

- Cardinality-awareness turns $\log(mn)$ into $\sqrt{\log m \log n}$.
- Optimistic Hedge is tight in the numbers of actions.

Also in the paper

- Refined upper bound via convex reformulation over (η, η', c, c').
- Dynamic regret: near-matching $\sqrt{\log m \log n} \cdot \log T$ dependence.
- 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.

Appendix

Q2. Proof sketch

Game construction. For $\Delta_x, \Delta_y \in (0, 1]$, consider

$$A = \begin{pmatrix} 0 & \Delta_y & \cdots & \Delta_y \\ -\Delta_x & \Delta_y - \Delta_x & \cdots & \Delta_y - \Delta_x \\ \vdots & \vdots & \ddots & \vdots \\ -\Delta_x & \Delta_y - \Delta_x & \cdots & \Delta_y - \Delta_x \end{pmatrix}.$$

Action 1 is optimal with a round-independent gap: $g_t(1) - g_t(i) = \Delta_x$ and $\ell_t(j) - \ell_t(1) = \Delta_y$ for all t, $i, j \neq 1$.

Q2. Proof sketch

Game construction. For $\Delta_x, \Delta_y \in (0, 1]$, consider

$$A = \begin{pmatrix} 0 & \Delta_y & \cdots & \Delta_y \\ -\Delta_x & \Delta_y - \Delta_x & \cdots & \Delta_y - \Delta_x \\ \vdots & \vdots & \ddots & \vdots \\ -\Delta_x & \Delta_y - \Delta_x & \cdots & \Delta_y - \Delta_x \end{pmatrix}.$$

Action 1 is optimal with a round-independent gap: $g_t(1) - g_t(i) = \Delta_x$ and $\ell_t(j) - \ell_t(1) = \Delta_y$ for all t, $i, j \neq 1$.

Since the gap is constant, the one-step optimistic correction telescopes away:

$$\frac{w_t(i)}{w_t(1)} = \mathrm{e}^{-\eta \Delta_x t}, \; x_t(1) = \frac{1}{1 + (m-1)\,\mathrm{e}^{-\eta \Delta_x t}} \to 1 \text{ only exponentially fast.}$$

Q2. Proof sketch

Game construction. For $\Delta_x, \Delta_y \in (0, 1]$, consider

$$A = \begin{pmatrix} 0 & \Delta_y & \cdots & \Delta_y \\ -\Delta_x & \Delta_y - \Delta_x & \cdots & \Delta_y - \Delta_x \\ \vdots & \vdots & \ddots & \vdots \\ -\Delta_x & \Delta_y - \Delta_x & \cdots & \Delta_y - \Delta_x \end{pmatrix}.$$

Action 1 is optimal with a round-independent gap: $g_t(1) - g_t(i) = \Delta_x$ and $\ell_t(j) - \ell_t(1) = \Delta_y$ for all t, $i, j \neq 1$.

Since the gap is constant, the one-step optimistic correction telescopes away:

$$\frac{w_t(i)}{w_t(1)} = \mathrm{e}^{-\eta \Delta_x t}, \; x_t(1) = \frac{1}{1 + (m-1)\,\mathrm{e}^{-\eta \Delta_x t}} \to 1 \text{ only exponentially fast.}$$

Summing the per-round regret and optimizing $\Delta_x \in (0, 1]$ (telescoping lemma):

$$\mathrm{Reg}_T^{\mathrm{x}} \geq \Delta_x \sum_{t=1}^{T} (1 - x_t(1)) \geq \frac{\log m}{\eta} - o(1).$$