Corrupted Learning Dynamics in Games

https://arxiv.org/abs/2412.07120

Taira Tsuchiya[1,2], Shinji Ito[1,2], Haipeng Luo[3]

[1]The University of Tokyo, [2]RIKEN, [3]University of Southern California

June 30, 2025

38th Conference on Learning Theory (COLT 2025)

Outline

- Introduction
  - Two-player zero-sum games, Nash equilibrium
  - Nash equilibrium and no-regret dynamics
  - Fast convergence in games
  - Research questions
- Learning in corrupted two-player zero-sum games
- Lower bounds
- Learning in corrupted multi-player general-sum games
- Conclusion and discussion

Outline

- Introduction
  - Two-player zero-sum games, Nash equilibrium
  - Nash equilibrium and no-regret dynamics
  - Fast convergence in games
  - Research questions
- Learning in corrupted two-player zero-sum games
- Lower bounds
- Learning in corrupted multi-player general-sum games
- Conclusion and discussion

Two-player zero-sum games

Consider a two-player zero-sum game with a payoff matrix $A \in [-1, 1]^{m_x \times m_y}$
(m_x, m_y: the number of actions of x- and y-players)

1. x-player and y-player simultaneously choose *strategies* $x \in \Delta_{m_x}$ and $y \in \Delta_{m_y}$, respectively. ($\Delta_m = \{x \in [0, 1]^m : \|x\|_1 = 1\}$: the $(m-1)$-dimensional probability simplex)

Two-player zero-sum games

Consider a two-player zero-sum game with a payoff matrix $A \in [-1, 1]^{m_x \times m_y}$
(m_x, m_y: the number of actions of x- and y-players)

1. x-player and y-player simultaneously choose *strategies* $x \in \Delta_{m_x}$ and $y \in \Delta_{m_y}$, respectively. ($\Delta_m = \{x \in [0, 1]^m : \|x\|_1 = 1\}$: the $(m-1)$-dimensional probability simplex)
2. x-player gains a (expected) payoff of $x^\top Ay$, and y-player incurs a loss of $x^\top Ay$ (thus zero-sum).

Two-player zero-sum games

Consider a two-player zero-sum game with a payoff matrix $A \in [-1, 1]^{m_x \times m_y}$ (m_x, m_y: the number of actions of x- and y-players)

1. x-player and y-player simultaneously choose *strategies* $x \in \Delta_{m_x}$ and $y \in \Delta_{m_y}$, respectively. ($\Delta_m = \{x \in [0,1]^m \colon \|x\|_1 = 1\}$: the $(m-1)$-dimensional probability simplex)
2. x-player gains a (expected) payoff of $x^\top A y$, and y-player incurs a loss of $x^\top A y$ (thus zero-sum).

Two-player zero-sum games

Consider a two-player zero-sum game with a payoff matrix $A \in [-1, 1]^{m_x \times m_y}$
(m_x, m_y: the number of actions of x- and y-players)

1. x-player and y-player simultaneously choose *strategies* $x \in \Delta_{m_x}$ and $y \in \Delta_{m_y}$, respectively. ($\Delta_m = \{x \in [0, 1]^m \colon \|x\|_1 = 1\}$: the $(m-1)$-dimensional probability simplex)
2. x-player gains a (expected) payoff of $x^\top Ay$, and y-player incurs a loss of $x^\top Ay$ (thus zero-sum).

We say that a strategy x is *pure* if $x = \mathrm{e}_i$ for some $i \in [m_x] := \{1, \ldots, m_x\}$.

Two-player zero-sum games

Example 1. 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

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

Nash equilibrium

Nash equilibrium: a pair of (possibly randomized) actions in which no player has an incentive to deviate

Nash equilibrium

Nash equilibrium: a pair of (possibly randomized) actions in which no player has an incentive to deviate

Definition (Nash equilibrium)

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

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

Nash equilibrium

Nash equilibrium: a pair of (possibly randomized) actions in which no player has an incentive to deviate

Definition (Nash equilibrium)

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

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

In the Rock-Paper-Scissors example, the pair $x = (1/3, 1/3, 1/3), y = (1/3, 1/3, 1/3)$ is a Nash equilibrium.

A Nash equilibrium can be computed by linear programming, but ...

- ☹ Can be computationally infeasible for large-scale payoff matrices
- ☹ Cannot be used when the payoff matrix is unknown

Nash equilibrium

Nash equilibrium: a pair of (possibly randomized) actions in which no player has an incentive to deviate

Definition (Nash equilibrium)

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

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

In the Rock-Paper-Scissors example, the pair $x = (1/3, 1/3, 1/3), y = (1/3, 1/3, 1/3)$ is a Nash equilibrium.

A Nash equilibrium can be computed by linear programming, but ...

- ☹ Can be computationally infeasible for large-scale payoff matrices
- ☹ Cannot be used when the payoff matrix is unknown

$\rightarrow$ solution: learning in games

Learning in games

Learning in games

Multiple players interact in a shared environment,
each aiming to maximize their total rewards (= minimize their *regret*)
by iteratively adapting their strategies based on repeated interactions

Learning in games

Learning in games

Multiple players interact in a shared environment,
each aiming to maximize their total rewards (= minimize their *regret*)
by iteratively adapting their strategies based on repeated interactions

Broader applications

- Minimax optimization (*e.g.*, $\min_x \max_y x^\top A y$)
- Multi-agent reinforcement learning
- Superhuman AI for poker, human-level AI for Stratego
- Alignment of LLMs
- ...

Learning in two-player zero-sum games

Learning in two-player zero-sum games with 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 $t = 1, \ldots, T$:

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

Learning in two-player zero-sum games with 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 $t = 1, \ldots, T$:

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 observes a expected reward vector $g^{(t)} = Ay^{(t)}$ and
   y-player observes a expected loss vector $\ell^{(t)} = A^\top x^{(t)}$;

Learning in two-player zero-sum games

Learning in two-player zero-sum games with 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 $t = 1, \ldots, T$:

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 observes a expected reward vector $g^{(t)} = Ay^{(t)}$ and
   y-player observes a expected loss vector $\ell^{(t)} = A^\top x^{(t)}$;
3. x-player gains a payoff of $\langle x^{(t)}, Ay^{(t)} \rangle = \langle x^{(t)}, g^{(t)} \rangle$ and
   y-player incurs a loss of $\langle x^{(t)}, Ay^{(t)} \rangle = \langle y^{(t)}, \ell^{(t)} \rangle$; **(thus zero-sum)**

Learning in two-player zero-sum games

Learning in two-player zero-sum games with 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 $t = 1, \ldots, T$:

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 observes a expected reward vector $g^{(t)} = Ay^{(t)}$ and
   y-player observes a expected loss vector $\ell^{(t)} = A^\top x^{(t)}$;
3. x-player gains a payoff of $\langle x^{(t)}, Ay^{(t)} \rangle = \langle x^{(t)}, g^{(t)} \rangle$ and
   y-player incurs a loss of $\langle x^{(t)}, Ay^{(t)} \rangle = \langle y^{(t)}, \ell^{(t)} \rangle$; **(thus zero-sum)**

The goal of x-/y- players is to minimize the **regret** (without knowing A):

- $\mathrm{Reg}^T_{x,g} = \max_{x^* \in \Delta_{m_x}} \left\{ \sum_{t=1}^T \langle x^*, g^{(t)} \rangle - \sum_{t=1}^T \langle x^{(t)}, g^{(t)} \rangle \right\}$,
- $\mathrm{Reg}^T_{y,\ell} = \max_{y^* \in \Delta_{m_y}} \left\{ \sum_{t=1}^T \langle y^{(t)}, \ell^{(t)} \rangle - \sum_{t=1}^T \langle y^*, \ell^{(t)} \rangle \right\}$.

No-regret learning dynamics and Nash equilibrium

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

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

No-regret learning dynamics and Nash equilibrium

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

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

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. Then its product distribution $(\bar{x}_T, \bar{y}_T)$ is a $((\mathrm{Reg}_{x,g}^T + \mathrm{Reg}_{y,\ell}^T)/T)$-approximate Nash equilibrium.

No-regret learning dynamics and Nash equilibrium

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

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

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. Then its product distribution $(\bar{x}_T, \bar{y}_T)$ is a $((\mathrm{Reg}^T_{x,g} + \mathrm{Reg}^T_{y,\ell})/T)$-approximate Nash equilibrium.

$\rightarrow$ When the x- and y-players use standard online convex optimization algorithms, we can guarantee $O(1/\sqrt{T})$ convergence to a Nash eq! (w/ uncoupled dynamics)

No-regret learning dynamics and Nash equilibrium

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

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

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. Then its product distribution $(\bar{x}_T, \bar{y}_T)$ is a $((\mathrm{Reg}_{x,g}^T + \mathrm{Reg}_{y,\ell}^T)/T)$-approximate Nash equilibrium.

$\rightarrow$ When the x- and y-players use standard online convex optimization algorithms, we can guarantee $O(1/\sqrt{T})$ convergence to a Nash eq! (w/ uncoupled dynamics)

e.g., Hedge algorithm guarantees $\mathrm{Reg}_{x,g}^T = \widetilde{O}(\sqrt{T})$ and $\mathrm{Reg}_{y,\ell}^T = \widetilde{O}(\sqrt{T})$.

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

No-regret learning dynamics and Nash equilibrium

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

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

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. Then its product distribution $(\bar{x}_T, \bar{y}_T)$ is a $((\mathrm{Reg}_{x,g}^T + \mathrm{Reg}_{y,\ell}^T)/T)$-approximate Nash equilibrium.

$\rightarrow$ When the x- and y-players use standard online convex optimization algorithms, we can guarantee $O(1/\sqrt{T})$ convergence to a Nash eq! (w/ uncoupled dynamics)

e.g., Hedge algorithm guarantees $\mathrm{Reg}_{x,g}^T = \widetilde{O}(\sqrt{T})$ and $\mathrm{Reg}_{y,\ell}^T = \widetilde{O}(\sqrt{T})$.

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

Q. Is this optimal rate in learning in games?

Fast convergence in games

Hedge algorithm (recall $g^{(t)} = Ay^{(t)}$ and $\ell^{(t)} = A^\top x^{(t)}$):

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

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

Fast convergence in games

Hedge algorithm (recall $g^{(t)} = Ay^{(t)}$ and $\ell^{(t)} = A^\top x^{(t)}$):

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

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

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

If x- and y-players **fully** *follow optimistic Hedge with* **constant** *learning rates $\eta_x, \eta_y \simeq 1$, then $\mathrm{Reg}_{x,g}^T = \widetilde{O}(1)$ and $\mathrm{Reg}_{y,\ell}^T = \widetilde{O}(1)$, which implies an $\widetilde{O}(1/T)$ conv. rate to Nash.*

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 $g^{(t+1)} = Ay^{(t+1)}$).

Fast convergence in games

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

If x- and y-players **fully** *follow optimistic Hedge with* **constant** *learning rates $\eta_x, \eta_y \simeq 1$, then $\mathrm{Reg}_{x,g}^T = \widetilde{O}(1)$ and $\mathrm{Reg}_{y,\ell}^T = \widetilde{O}(1)$, which implies an $\widetilde{O}(1/T)$ conv. rate to Nash.*

Q. What if the opponent does not follow optimistic Hedge with a constant learning rate?

Fast convergence in games

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

If x- and y-players **fully** *follow optimistic Hedge with* **constant** *learning rates $\eta_x, \eta_y \simeq 1$, then $\mathrm{Reg}_{x,g}^T = \widetilde{O}(1)$ and $\mathrm{Reg}_{y,\ell}^T = \widetilde{O}(1)$, which implies an $\widetilde{O}(1/T)$ conv. rate to Nash.*

Q. What if the opponent does not follow optimistic Hedge with a constant learning rate?

Continuing with the algorithm may lead to a linear regret: $\mathrm{Reg}_x^T = \Omega(T)$.

Fast convergence in games

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

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

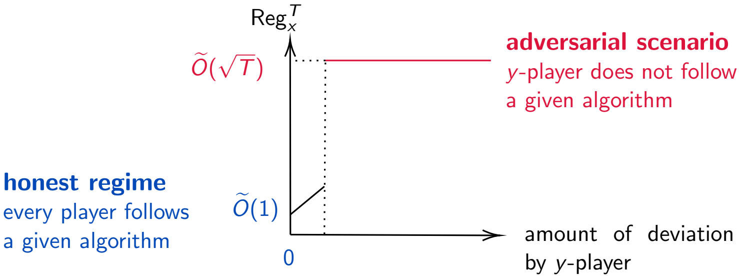
Theorem (Syrgkanis et al. 2015)

If x- and y-players **fully** *follow optimistic Hedge with* **constant** *learning rates $\eta_x, \eta_y \simeq 1$, then $\mathrm{Reg}_{x,g}^T = \widetilde{O}(1)$ and $\mathrm{Reg}_{y,\ell}^T = \widetilde{O}(1)$, which implies an $\widetilde{O}(1/T)$ conv. rate to Nash.*

Q. What if the opponent does not follow optimistic Hedge with a constant learning rate?
Continuing with the algorithm may lead to a linear regret: $\mathrm{Reg}_x^T = \Omega(T)$.
$\rightarrow$ 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 $\widetilde{O}(\sqrt{T})$
(*e.g.*, Hedge with learning rate of $\Theta(1/\sqrt{T})$)

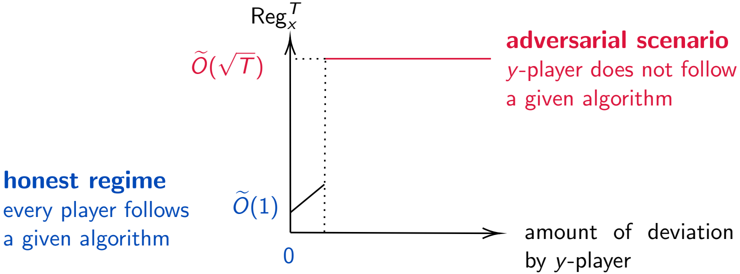
Research questions

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



Research questions

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



Research questions

- Can we adapt to deviations of the opponent from a given algorithm?
- Can we characterize regret and convergence rates to an equilibrium in such a corrupted game?

Research questions

Research questions

- Can we adapt to deviations of the opponent from a given algorithm?
- Can we characterize regret and convergence rates to an equilibrium in such a corrupted game?

Our contributions

- Establish a framework of corrupted games, in which each player may deviate from a prescribed algorithm
- Give a nearly complete characterization of learning dynamics in corrupted games, by deriving regret upper and lower bounds in (normal-form) two-player zero-sum and multi-player general-sum games

Outline

- Introduction
  - Two-player zero-sum games, Nash equilibrium
  - Nash equilibrium and no-regret dynamics
  - Fast convergence in games
  - Research questions
- Learning in corrupted two-player zero-sum games
- Lower bounds
- Learning in corrupted multi-player general-sum games
- Conclusion and discussion

Corrupted regime in two-player zero-sum games

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

1. A prescribed algorithm suggests strategies $\widehat{x}^{(t)} \in \Delta_{m_x}$ and $\widehat{y}^{(t)} \in \Delta_{m_y}$;

Corrupted regime in two-player zero-sum games

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

1. A prescribed algorithm suggests strategies $\widehat{x}^{(t)} \in \Delta_{m_x}$ and $\widehat{y}^{(t)} \in \Delta_{m_y}$;
2. **(corruption of strategies)**
   x-player selects a strategy $x^{(t)} \leftarrow \widehat{x}^{(t)} + \widehat{c}_x^{(t)}$ and
   y-player selects $y^{(t)} \leftarrow \widehat{y}^{(t)} + \widehat{c}_y^{(t)}$;
   Note: The corruption is allowed to depend arbitrarily on the past observations.

Corrupted regime in two-player zero-sum games

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

1. A prescribed algorithm suggests strategies $\widehat{x}^{(t)} \in \Delta_{m_x}$ and $\widehat{y}^{(t)} \in \Delta_{m_y}$;
2. **(corruption of strategies)**
   x-player selects a strategy $x^{(t)} \leftarrow \widehat{x}^{(t)} + \widehat{c}_x^{(t)}$ and
   y-player selects $y^{(t)} \leftarrow \widehat{y}^{(t)} + \widehat{c}_y^{(t)}$;
   Note: The corruption is allowed to depend arbitrarily on the past observations.
3. x-player observes a expected reward vector $g^{(t)} = Ay^{(t)}$ and
   y-player observes a expected loss vector $\ell^{(t)} = A^\top x^{(t)}$;

Corrupted regime in two-player zero-sum games

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

1. A prescribed algorithm suggests strategies $\widehat{x}^{(t)} \in \Delta_{m_x}$ and $\widehat{y}^{(t)} \in \Delta_{m_y}$;
2. **(corruption of strategies)**
   x-player selects a strategy $x^{(t)} \leftarrow \widehat{x}^{(t)} + \widehat{c}_x^{(t)}$ and
   y-player selects $y^{(t)} \leftarrow \widehat{y}^{(t)} + \widehat{c}_y^{(t)}$;
   Note: The corruption is allowed to depend arbitrarily on the past observations.
3. x-player observes a expected reward vector $g^{(t)} = Ay^{(t)}$ and
   y-player observes a expected loss vector $\ell^{(t)} = A^\top x^{(t)}$;
4. x-player gains a payoff of $\langle x^{(t)}, g^{(t)} \rangle$ and y-player incurs a loss of $\langle y^{(t)}, \ell^{(t)} \rangle$

Corrupted regime in two-player zero-sum games

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

1. A prescribed algorithm suggests strategies $\widehat{x}^{(t)} \in \Delta_{m_x}$ and $\widehat{y}^{(t)} \in \Delta_{m_y}$;
2. **(corruption of strategies)**
   x-player selects a strategy $x^{(t)} \leftarrow \widehat{x}^{(t)} + \widehat{c}_x^{(t)}$ and
   y-player selects $y^{(t)} \leftarrow \widehat{y}^{(t)} + \widehat{c}_y^{(t)}$;
   Note: The corruption is allowed to depend arbitrarily on the past observations.
3. x-player observes a expected reward vector $g^{(t)} = Ay^{(t)}$ and
   y-player observes a expected loss vector $\ell^{(t)} = A^\top x^{(t)}$;
4. x-player gains a payoff of $\langle x^{(t)}, g^{(t)} \rangle$ and y-player incurs a loss of $\langle y^{(t)}, \ell^{(t)} \rangle$

Cumulative corruption of strategies: $\widehat{C}_x = \sum_{t=1}^T \|\widehat{c}_x^{(t)}\|_1$, $\widehat{C}_y = \sum_{t=1}^T \|\widehat{c}_y^{(t)}\|_1$

Corrupted regime in two-player zero-sum games

We investigate a scenario where **the observed utilities may also be corrupted**.

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

1. A prescribed algorithm suggests strategies $\widehat{x}^{(t)} \in \Delta_{m_x}$ and $\widehat{y}^{(t)} \in \Delta_{m_y}$;
2. x-player selects a strategy $x^{(t)} \leftarrow \widehat{x}^{(t)} + \widehat{c}_x^{(t)}$ and
   y-player selects $y^{(t)} \leftarrow \widehat{y}^{(t)} + \widehat{c}_y^{(t)}$;
3. **(corruption of utilities)**
   x-player observes a corrupted reward vector $\widetilde{g}^{(t)} = g^{(t)} + \widetilde{c}_x^{(t)}$ for $g^{(t)} = Ay^{(t)}$,
   y-player observes a corrupted loss vector $\widetilde{\ell}^{(t)} = \ell^{(t)} + \widetilde{c}_y^{(t)}$ for $\ell^{(t)} = A^\top x^{(t)}$;

Corrupted regime in two-player zero-sum games

We investigate a scenario where **the observed utilities may also be corrupted**.

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

1. A prescribed algorithm suggests strategies $\widehat{x}^{(t)} \in \Delta_{m_x}$ and $\widehat{y}^{(t)} \in \Delta_{m_y}$;
2. x-player selects a strategy $x^{(t)} \leftarrow \widehat{x}^{(t)} + \widehat{c}_x^{(t)}$ and
   y-player selects $y^{(t)} \leftarrow \widehat{y}^{(t)} + \widehat{c}_y^{(t)}$;
3. **(corruption of utilities)**
   x-player observes a corrupted reward vector $\widetilde{g}^{(t)} = g^{(t)} + \widetilde{c}_x^{(t)}$ for $g^{(t)} = Ay^{(t)}$,
   y-player observes a corrupted loss vector $\widetilde{\ell}^{(t)} = \ell^{(t)} + \widetilde{c}_y^{(t)}$ for $\ell^{(t)} = A^\top x^{(t)}$;
4. x-player gains a payoff of $\langle x^{(t)}, g^{(t)} \rangle$ or $\langle x^{(t)}, \widetilde{g}^{(t)} \rangle$ and y-player incurs a loss of $\langle y^{(t)}, \ell^{(t)} \rangle$ or $\langle y^{(t)}, \widetilde{\ell}^{(t)} \rangle$

Corrupted regime in two-player zero-sum games

We investigate a scenario where **the observed utilities may also be corrupted**.

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

1. A prescribed algorithm suggests strategies $\widehat{x}^{(t)} \in \Delta_{m_x}$ and $\widehat{y}^{(t)} \in \Delta_{m_y}$;
2. x-player selects a strategy $x^{(t)} \leftarrow \widehat{x}^{(t)} + \widehat{c}_x^{(t)}$ and
   y-player selects $y^{(t)} \leftarrow \widehat{y}^{(t)} + \widehat{c}_y^{(t)}$;
3. **(corruption of utilities)**
   x-player observes a corrupted reward vector $\widetilde{g}^{(t)} = g^{(t)} + \widetilde{c}_x^{(t)}$ for $g^{(t)} = Ay^{(t)}$,
   y-player observes a corrupted loss vector $\widetilde{\ell}^{(t)} = \ell^{(t)} + \widetilde{c}_y^{(t)}$ for $\ell^{(t)} = A^\top x^{(t)}$;
4. x-player gains a payoff of $\langle x^{(t)}, g^{(t)} \rangle$ or $\langle x^{(t)}, \widetilde{g}^{(t)} \rangle$ and y-player incurs a loss of $\langle y^{(t)}, \ell^{(t)} \rangle$ or $\langle y^{(t)}, \widetilde{\ell}^{(t)} \rangle$

Cumulative corruption of strategies and utilities:

- $\widehat{C}_x = \sum_{t=1}^T \|\widehat{c}_x^{(t)}\|_1$, $\widetilde{C}_x = \sum_{t=1}^T \|\widetilde{c}_x^{(t)}\|_\infty$, and $C_x = \widehat{C}_x + 2\widetilde{C}_x$.
- $\widehat{C}_y = \sum_{t=1}^T \|\widehat{c}_y^{(t)}\|_1$, $\widetilde{C}_y = \sum_{t=1}^T \|\widetilde{c}_y^{(t)}\|_\infty$, and $C_y = \widehat{C}_y + 2\widetilde{C}_y$.

Corrupted regime in two-player zero-sum games

We investigate a scenario where **the observed utilities may also be corrupted**.

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

1. A prescribed algorithm suggests strategies $\widehat{x}^{(t)} \in \Delta_{m_x}$ and $\widehat{y}^{(t)} \in \Delta_{m_y}$;
2. x-player selects a strategy $x^{(t)} \leftarrow \widehat{x}^{(t)} + \widehat{c}_x^{(t)}$ and
   y-player selects $y^{(t)} \leftarrow \widehat{y}^{(t)} + \widehat{c}_y^{(t)}$;
3. **(corruption of utilities)**
   x-player observes a corrupted reward vector $\widetilde{g}^{(t)} = g^{(t)} + \widetilde{c}_x^{(t)}$ for $g^{(t)} = Ay^{(t)}$,
   y-player observes a corrupted loss vector $\widetilde{\ell}^{(t)} = \ell^{(t)} + \widetilde{c}_y^{(t)}$ for $\ell^{(t)} = A^\top x^{(t)}$;
4. x-player gains a payoff of $\langle x^{(t)}, g^{(t)} \rangle$ or $\langle x^{(t)}, \widetilde{g}^{(t)} \rangle$ and y-player incurs a loss of $\langle y^{(t)}, \ell^{(t)} \rangle$ or $\langle y^{(t)}, \widetilde{\ell}^{(t)} \rangle$

- corrupted regime with no corruptions = the honest regime
- corrupted regime w/ $\widetilde{C}_y = \Omega(T)$ = adversarial scenario for x-player

Our algorithm: Optimistic Hedge with adaptive learning rate

Syrgkanis et al. (2015): Optimistic Hedge with constant learning rate for the honest regime

$$x^{(t)}(i) \propto \exp\left(\eta_x\left(\sum_{s=1}^{t-1} g_s(i) + g_{t-1}(i)\right)\right), \; \eta_x \simeq 1, \quad \forall i \in [m_x]$$

Our algorithm: Optimistic Hedge with adaptive learning rate

Syrgkanis et al. (2015): Optimistic Hedge with constant learning rate for the honest regime

$$x^{(t)}(i) \propto \exp\left(\eta_x\left(\sum_{s=1}^{t-1} g_s(i) + g_{t-1}(i)\right)\right), \ \eta_x \simeq 1, \quad \forall i \in [m_x]$$

Ours: Optimistic Hedge with adaptive learning rate for the corrupted regime (not formally defined)

$$x^{(t)}(i) \propto \exp\left(\eta_x^{(t)}\left(\sum_{s=1}^{t-1} \widetilde{g}_s(i) + \widetilde{g}_{t-1}(i)\right)\right), \ \eta_x^{(t)} = \sqrt{\frac{\log_+(m_x)/2}{\log_+(m_x) + \sum_{s=1}^{t-1} \|\widetilde{g}^{(s)} - \widetilde{g}^{(s-1)}\|_\infty^2}}$$

with $\log_+(z) = \max\{\log z, 4\}$.

This is a very standard choice of learning rate (recall AdaGrad), but adjusted to satisfy $\eta_x^{(t)} \leq 1/\sqrt{2}$.

Main result (1): Regret upper bound in the corrupted regime

Cumulative corruption of strategies and utilities

- $\widehat{C}_x = \sum_{t=1}^{T} \|\widehat{c}_x^{(t)}\|_1$, $\widetilde{C}_x = \sum_{t=1}^{T} \|\widetilde{c}_x^{(t)}\|_\infty$, and $C_x = \widehat{C}_x + 2\widetilde{C}_x$.
- $\widehat{C}_y = \sum_{t=1}^{T} \|\widehat{c}_y^{(t)}\|_1$, $\widetilde{C}_y = \sum_{t=1}^{T} \|\widetilde{c}_y^{(t)}\|_\infty$, and $C_y = \widehat{C}_y + 2\widetilde{C}_y$.

Regret upper bounds of the x-player:

	Honest regime	Corrupted regime
Syrgkanis et al. (2015)	$\log(m_x m_y)$	$\log(m_x m_y) + \sqrt{T \log m_x} + C_x$

Main result (1): Regret upper bound in the corrupted regime

Cumulative corruption of strategies and utilities

- $\widehat{C}_x = \sum_{t=1}^{T} \|\widehat{c}_x^{(t)}\|_1$, $\widetilde{C}_x = \sum_{t=1}^{T} \|\widetilde{c}_x^{(t)}\|_\infty$, and $C_x = \widehat{C}_x + 2\widetilde{C}_x$.
- $\widehat{C}_y = \sum_{t=1}^{T} \|\widehat{c}_y^{(t)}\|_1$, $\widetilde{C}_y = \sum_{t=1}^{T} \|\widetilde{c}_y^{(t)}\|_\infty$, and $C_y = \widehat{C}_y + 2\widetilde{C}_y$.

Regret upper bounds of the x-player:

	Honest regime	Corrupted regime
Syrgkanis et al. (2015)	$\log(m_x m_y)$	$\log(m_x m_y) + \sqrt{T \log m_x} + C_x$
Ours	$\sqrt{\log(m_x m_y) \log m_x}$	$\min\left\{\sqrt{(\log(m_x m_y) + C_x + C_y) \log m_x}, \sqrt{T \log m_x}\right\} + C_x$

Main result (1): Regret upper bound in the corrupted regime

Cumulative corruption of strategies and utilities

- $\widehat{C}_x = \sum_{t=1}^T \|\widehat{c}_x^{(t)}\|_1$, $\widetilde{C}_x = \sum_{t=1}^T \|\widetilde{c}_x^{(t)}\|_\infty$, and $C_x = \widehat{C}_x + 2\widetilde{C}_x$.
- $\widehat{C}_y = \sum_{t=1}^T \|\widehat{c}_y^{(t)}\|_1$, $\widetilde{C}_y = \sum_{t=1}^T \|\widetilde{c}_y^{(t)}\|_\infty$, and $C_y = \widehat{C}_y + 2\widetilde{C}_y$.

Regret upper bounds of the x-player:

	Honest regime	Corrupted regime
Syrgkanis et al. (2015)	$\log(m_x m_y)$	$\log(m_x m_y) + \sqrt{T \log m_x} + C_x$
Ours	$\sqrt{\log(m_x m_y) \log m_x}$	$\min\left\{\sqrt{(\log(m_x m_y) + C_x + C_y) \log m_x}, \sqrt{T \log m_x}\right\} + C_x$

The bound $\mathrm{Reg}_{x,g}^T \lesssim \sqrt{\widehat{C}_y} + \widehat{C}_x$ in the corrupted regime ...

- smoothly interpolates between the $\widetilde{O}(1)$ regret in the honest regime and the $\widetilde{O}(\sqrt{T})$ regret in the adversarial scenario (noting $C_y \in [0, 3T]$).

Main result (1): Regret upper bound in the corrupted regime

Cumulative corruption of strategies and utilities

- $\widehat{C}_x = \sum_{t=1}^{T} \|\widehat{c}_x^{(t)}\|_1$, $\widetilde{C}_x = \sum_{t=1}^{T} \|\widetilde{c}_x^{(t)}\|_\infty$, and $C_x = \widehat{C}_x + 2\widetilde{C}_x$.
- $\widehat{C}_y = \sum_{t=1}^{T} \|\widehat{c}_y^{(t)}\|_1$, $\widetilde{C}_y = \sum_{t=1}^{T} \|\widetilde{c}_y^{(t)}\|_\infty$, and $C_y = \widehat{C}_y + 2\widetilde{C}_y$.

Regret upper bounds of the x-player:

	Honest regime	Corrupted regime
Syrgkanis et al. (2015)	$\log(m_x m_y)$	$\log(m_x m_y) + \sqrt{T \log m_x} + C_x$
Ours	$\sqrt{\log(m_x m_y) \log m_x}$	$\min\left\{\sqrt{(\log(m_x m_y) + C_x + C_y) \log m_x}, \sqrt{T \log m_x}\right\} + C_x$

The bound $\mathrm{Reg}_{x,g}^{T} \lesssim \sqrt{\widehat{C}_y} + \widehat{C}_x$ in the corrupted regime ...

- smoothly interpolates between the $\widetilde{O}(1)$ regret in the honest regime and the $\widetilde{O}(\sqrt{T})$ regret in the adversarial scenario (noting $C_y \in [0, 3T]$).
- incentivizes players to follow the given algorithm:
  - any deviation by an opponent incurs only a square-root penalty $\sqrt{\widehat{C}_y}$,
  - whereas a deviation by a player from the given algorithm incurs a linear penalty $\widehat{C}_x$.

Proof sketch of $\mathrm{Reg}_{x,g}^{T} \lesssim \sqrt{\widehat{C}_y} + \widehat{C}_x$

(omitting corruption of utilities, log and const factors)

Use the standard analysis of Optimistic Hedge:

$$\mathrm{Reg}_{\widehat{x},g}^{T} = \max_{x^* \in \Delta_{m_x}} \left\{ \sum_{t=1}^{T} \langle x^*, g^{(t)} \rangle - \sum_{t=1}^{T} \langle \widehat{x}^{(t)}, g^{(t)} \rangle \right\}$$

$\widehat{x}^{(t)}$: suggested strategy

Proof sketch of $\mathrm{Reg}_{x,g}^{T} \lesssim \sqrt{\widehat{C}_y} + \widehat{C}_x$

(omitting corruption of utilities, log and const factors)

Use the standard analysis of Optimistic Hedge:

$$\mathrm{Reg}_{\widehat{x},g}^{T} \lesssim \frac{1}{\eta_x^{(T+1)}} + \sum_{t=1}^{T} \eta_x^{(t)} \|g^{(t)} - g^{(t-1)}\|_\infty^2 - \sum_{t=1}^{T} \frac{1}{4\eta_x^{(t)}} \|\widehat{x}^{(t+1)} - \widehat{x}^{(t)}\|_1^2$$

Proof sketch of $\mathrm{Reg}_{x,g}^{T} \lesssim \sqrt{\widehat{C}_y} + \widehat{C}_x$

(omitting corruption of utilities, log and const factors)

Use the standard analysis of Optimistic Hedge:

$$\mathrm{Reg}_{\widehat{x},g}^{T} \lesssim \frac{1}{\eta_x^{(T+1)}} + \sum_{t=1}^{T} \eta_x^{(t)} \|g^{(t)} - g^{(t-1)}\|_\infty^2 - \sum_{t=1}^{T} \frac{1}{4\eta_x^{(t)}} \|\widehat{x}^{(t+1)} - \widehat{x}^{(t)}\|_1^2$$

$$\lesssim \sqrt{\sum_{t=1}^{T} \|g^{(t)} - g^{(t-1)}\|_\infty^2} - \sum_{t=1}^{T} \|\widehat{x}^{(t)} - \widehat{x}^{(t-1)}\|_1^2 \qquad (\text{def of } \eta_x^{(t)} \text{ \& } \eta_x^{(t)} \le 1/\sqrt{2})$$

Proof sketch of $\mathrm{Reg}^T_{x,g} \lesssim \sqrt{\widehat{C}_y} + \widehat{C}_x$

(omitting corruption of utilities, log and const factors)

Use the standard analysis of Optimistic Hedge:

$$\mathrm{Reg}^T_{\widehat{x},g} \lesssim \frac{1}{\eta_x^{(T+1)}} + \sum_{t=1}^T \eta_x^{(t)} \|g^{(t)} - g^{(t-1)}\|_\infty^2 - \sum_{t=1}^T \frac{1}{4\eta_x^{(t)}} \|\widehat{x}^{(t+1)} - \widehat{x}^{(t)}\|_1^2$$

$$\lesssim \sqrt{\sum_{t=1}^T \|g^{(t)} - g^{(t-1)}\|_\infty^2} - \sum_{t=1}^T \|\widehat{x}^{(t)} - \widehat{x}^{(t-1)}\|_1^2 \qquad (\text{def of } \eta_x^{(t)} \ \& \ \eta_x^{(t)} \le 1/\sqrt{2})$$

The first term is evaluated as (recalling $\widehat{C}_y = \sum_{t=1}^T \|y^{(t)} - \widehat{y}^{(t)}\|_1$)

$$\sum_{t=1}^T \|g^{(t)} - g^{(t-1)}\|_\infty^2 = \sum_{t=1}^T \|A(y^{(t)} - y^{(t-1)})\|_\infty^2 \le \sum_{t=1}^T \|y^{(t)} - y^{(t-1)}\|_1^2$$

$$\le 4\sum_{t=1}^T \|y^{(t)} - \widehat{y}^{(t)}\|_1^2 + 4\sum_{t=1}^T \|\widehat{y}^{(t)} - \widehat{y}^{(t-1)}\|_1^2 \lesssim \widehat{C}_y + \sum_{t=1}^T \|\widehat{y}^{(t)} - \widehat{y}^{(t-1)}\|_1^2 .$$

Proof sketch of $\mathrm{Reg}^T_{x,g} \lesssim \sqrt{\widehat{C}_y} + \widehat{C}_x$ (cont'd)

Previous slide:

$$\mathrm{Reg}^T_{\widehat{x},g} \lesssim \sqrt{\sum_{t=1}^{T} \|g^{(t)} - g^{(t-1)}\|_\infty^2} - \sum_{t=1}^{T} \|\widehat{x}^{(t)} - \widehat{x}^{(t-1)}\|_1^2 \,,$$

$$\sum_{t=1}^{T} \|g^{(t)} - g^{(t-1)}\|_\infty^2 \lesssim \widehat{C}_y + \sum_{t=1}^{T} \|\widehat{y}^{(t)} - \widehat{y}^{(t-1)}\|_1^2 \,.$$

Proof sketch of $\mathrm{Reg}_{x,g}^T \lesssim \sqrt{\widehat{C}_y} + \widehat{C}_x$ (cont'd)

Previous slide:

$$\mathrm{Reg}_{\widehat{x},g}^T \lesssim \sqrt{\sum_{t=1}^T \|g^{(t)} - g^{(t-1)}\|_\infty^2 - \sum_{t=1}^T \|\widehat{x}^{(t)} - \widehat{x}^{(t-1)}\|_1^2}\,,$$

$$\sum_{t=1}^T \|g^{(t)} - g^{(t-1)}\|_\infty^2 \lesssim \widehat{C}_y + \sum_{t=1}^T \|\widehat{y}^{(t)} - \widehat{y}^{(t-1)}\|_1^2\,.$$

Combining these two and using $|\mathrm{Reg}_{\widehat{x},g}^T - \mathrm{Reg}_{x,g}^T| \le \widehat{C}_x$ give

$$\begin{aligned}\mathrm{Reg}_{x,g}^T &\le \mathrm{Reg}_{\widehat{x},g}^T + \widehat{C}_x \\ &\lesssim \sqrt{\widehat{C}_y + \sum_{t=1}^T \|\widehat{y}^{(t)} - \widehat{y}^{(t-1)}\|_1^2 - \sum_{t=1}^T \|\widehat{x}^{(t)} - \widehat{x}^{(t-1)}\|_1^2} + \widehat{C}_x\,.\end{aligned}$$

Proof sketch of $\mathrm{Reg}_{x,g}^T \lesssim \sqrt{\widehat{C}_y} + \widehat{C}_x$ (cont'd)

Combining these two and using $|\mathrm{Reg}_{\widehat{x},g}^T - \mathrm{Reg}_{x,g}^T| \le \widehat{C}_x$ give

$$\begin{aligned}\mathrm{Reg}_{x,g}^T &\le \mathrm{Reg}_{\widehat{x},g}^T + \widehat{C}_x \\ &\lesssim \sqrt{\widehat{C}_y + \sum_{t=1}^{T} \|\widehat{y}^{(t)} - \widehat{y}^{(t-1)}\|_1^2} - \sum_{t=1}^{T} \|\widehat{x}^{(t)} - \widehat{x}^{(t-1)}\|_1^2 + \widehat{C}_x\,.\end{aligned}$$

Similarly, we have

$$\mathrm{Reg}_{y,\ell}^T \lesssim \sqrt{\widehat{C}_x + \sum_{t=1}^{T} \|\widehat{x}^{(t)} - \widehat{x}^{(t-1)}\|_1^2} - \sum_{t=1}^{T} \|\widehat{y}^{(t)} - \widehat{y}^{(t-1)}\|_1^2 + \widehat{C}_y\,.$$

Summing up these two inequalities gives ...

Proof sketch of $\mathrm{Reg}_{x,g}^T \lesssim \sqrt{\widehat{C}_y} + \widehat{C}_x$ (cont'd)

Summing up these two inequalities gives

$$\mathrm{Reg}_{x,g}^T + \mathrm{Reg}_{y,\ell}^T \lesssim \sqrt{\widehat{C}_y + \sum_{t=1}^{T} \|\widehat{y}^{(t)} - \widehat{y}^{(t-1)}\|_1^2} + \sqrt{\widehat{C}_x + \sum_{t=1}^{T} \|\widehat{x}^{(t)} - \widehat{x}^{(t-1)}\|_1^2}$$
$$- \sum_{t=1}^{T} \left(\|\widehat{x}^{(t)} - \widehat{x}^{(t-1)}\|_1^2 + \|\widehat{y}^{(t)} - \widehat{y}^{(t-1)}\|_1^2 \right) + \widehat{C}_x + \widehat{C}_y$$

Proof sketch of $\mathrm{Reg}_{x,g}^T \lesssim \sqrt{\widehat{C}_y} + \widehat{C}_x$ (cont'd)

Summing up these two inequalities gives

$$\begin{aligned}
\mathrm{Reg}_{x,g}^T + \mathrm{Reg}_{y,\ell}^T &\lesssim \sqrt{\widehat{C}_y + \sum_{t=1}^{T} \|\widehat{y}^{(t)} - \widehat{y}^{(t-1)}\|_1^2} + \sqrt{\widehat{C}_x + \sum_{t=1}^{T} \|\widehat{x}^{(t)} - \widehat{x}^{(t-1)}\|_1^2} \\
&\quad - \sum_{t=1}^{T} \Big(\|\widehat{x}^{(t)} - \widehat{x}^{(t-1)}\|_1^2 + \|\widehat{y}^{(t)} - \widehat{y}^{(t-1)}\|_1^2 \Big) + \widehat{C}_x + \widehat{C}_y \\
\text{(Cauchy–Schwarz)} \quad &\lesssim \sqrt{\widehat{C}_x + \widehat{C}_y} + \widehat{C}_x + \widehat{C}_y - \frac{1}{2} \sum_{t=1}^{T} \Big(\|\widehat{x}^{(t)} - \widehat{x}^{(t-1)}\|_1^2 + \|\widehat{y}^{(t)} - \widehat{y}^{(t-1)}\|_1^2 \Big).
\end{aligned}$$

Proof sketch of $\mathrm{Reg}^T_{x,g} \lesssim \sqrt{\widehat{C}_y} + \widehat{C}_x$ (cont'd)

Summing up these two inequalities gives

$$\begin{aligned}
\mathrm{Reg}^T_{x,g} + \mathrm{Reg}^T_{y,\ell} &\lesssim \sqrt{\widehat{C}_y + \sum_{t=1}^{T} \|\widehat{y}^{(t)} - \widehat{y}^{(t-1)}\|_1^2} + \sqrt{\widehat{C}_x + \sum_{t=1}^{T} \|\widehat{x}^{(t)} - \widehat{x}^{(t-1)}\|_1^2} \\
&\quad - \sum_{t=1}^{T} \left(\|\widehat{x}^{(t)} - \widehat{x}^{(t-1)}\|_1^2 + \|\widehat{y}^{(t)} - \widehat{y}^{(t-1)}\|_1^2 \right) + \widehat{C}_x + \widehat{C}_y \\
\text{(Cauchy–Schwarz)} \quad &\lesssim \sqrt{\widehat{C}_x + \widehat{C}_y} + \widehat{C}_x + \widehat{C}_y - \frac{1}{2} \sum_{t=1}^{T} \left(\|\widehat{x}^{(t)} - \widehat{x}^{(t-1)}\|_1^2 + \|\widehat{y}^{(t)} - \widehat{y}^{(t-1)}\|_1^2 \right).
\end{aligned}$$

Since $\mathrm{Reg}^T_{x,g} + \mathrm{Reg}^T_{y,\ell} \geq 0$ (from the definition of the Nash eq), ...

Proof sketch of $\mathrm{Reg}_{x,g}^T \lesssim \sqrt{\widehat{C}_y} + \widehat{C}_x$ (cont'd)

Since $\mathrm{Reg}_{x,g}^T + \mathrm{Reg}_{y,\ell}^T \geq 0$ (from the definition of the Nash eq),

$$\sum_{t=1}^{T} \left(\|\widehat{x}^{(t)} - \widehat{x}^{(t-1)}\|_1^2 + \|\widehat{y}^{(t)} - \widehat{y}^{(t-1)}\|_1^2 \right) \lesssim \sqrt{\widehat{C}_x + \widehat{C}_y} + \widehat{C}_x + \widehat{C}_y \,.$$

Proof sketch of $\mathrm{Reg}_{x,g}^T \lesssim \sqrt{\widehat{C}_y} + \widehat{C}_x$ (cont'd)

Since $\mathrm{Reg}_{x,g}^T + \mathrm{Reg}_{y,\ell}^T \geq 0$ (from the definition of the Nash eq),

$$\sum_{t=1}^{T} \left(\|\widehat{x}^{(t)} - \widehat{x}^{(t-1)}\|_1^2 + \|\widehat{y}^{(t)} - \widehat{y}^{(t-1)}\|_1^2 \right) \lesssim \sqrt{\widehat{C}_x + \widehat{C}_y} + \widehat{C}_x + \widehat{C}_y \,.$$

Recalling that $\mathrm{Reg}_{x,g}^T$ is upper bouned as

$$\mathrm{Reg}_{x,g}^T \lesssim \sqrt{\widehat{C}_y + \sum_{t=1}^{T} \|\widehat{y}^{(t)} - \widehat{y}^{(t-1)}\|_1^2} - \sum_{t=1}^{T} \|\widehat{x}^{(t)} - \widehat{x}^{(t-1)}\|_1^2 + \widehat{C}_x \,,$$

Proof sketch of $\mathrm{Reg}_{x,g}^T \lesssim \sqrt{\widehat{C}_y} + \widehat{C}_x$ (cont'd)

Since $\mathrm{Reg}_{x,g}^T + \mathrm{Reg}_{y,\ell}^T \geq 0$ (from the definition of the Nash eq),

$$\sum_{t=1}^{T}\left(\|\widehat{x}^{(t)} - \widehat{x}^{(t-1)}\|_1^2 + \|\widehat{y}^{(t)} - \widehat{y}^{(t-1)}\|_1^2\right) \lesssim \sqrt{\widehat{C}_x + \widehat{C}_y} + \widehat{C}_x + \widehat{C}_y\,.$$

Recalling that $\mathrm{Reg}_{x,g}^T$ is upper bouned as

$$\mathrm{Reg}_{x,g}^T \lesssim \sqrt{\widehat{C}_y + \textstyle\sum_{t=1}^T \|\widehat{y}^{(t)} - \widehat{y}^{(t-1)}\|_1^2} - \textstyle\sum_{t=1}^T \|\widehat{x}^{(t)} - \widehat{x}^{(t-1)}\|_1^2 + \widehat{C}_x\,,$$

we obtain

$$\mathrm{Reg}_{x,g}^T \lesssim \sqrt{\sqrt{\widehat{C}_x + \widehat{C}_y} + \widehat{C}_x + \widehat{C}_y} + \widehat{C}_x \lesssim \sqrt{\widehat{C}_y} + \widehat{C}_x\,,$$

as desired.

Proof sketch of $\mathrm{Reg}_{x,g}^T \lesssim \sqrt{\widehat{C}_y} + \widehat{C}_x$ (cont'd)

Since $\mathrm{Reg}_{x,g}^T + \mathrm{Reg}_{y,\ell}^T \geq 0$ (from the definition of the Nash eq),

$$\sum_{t=1}^{T}\left(\|\widehat{x}^{(t)} - \widehat{x}^{(t-1)}\|_1^2 + \|\widehat{y}^{(t)} - \widehat{y}^{(t-1)}\|_1^2\right) \lesssim \sqrt{\widehat{C}_x + \widehat{C}_y} + \widehat{C}_x + \widehat{C}_y\,.$$

Recalling that $\mathrm{Reg}_{x,g}^T$ is upper bouned as

$$\mathrm{Reg}_{x,g}^T \lesssim \sqrt{\widehat{C}_y + \sum_{t=1}^T \|\widehat{y}^{(t)} - \widehat{y}^{(t-1)}\|_1^2} - \sum_{t=1}^T \|\widehat{x}^{(t)} - \widehat{x}^{(t-1)}\|_1^2 + \widehat{C}_x\,,$$

we obtain

$$\mathrm{Reg}_{x,g}^T \lesssim \sqrt{\sqrt{\widehat{C}_x + \widehat{C}_y} + \widehat{C}_x + \widehat{C}_y} + \widehat{C}_x \lesssim \sqrt{\widehat{C}_y} + \widehat{C}_x\,,$$

as desired.

Deriving the regret upper bounds for two-player zero-sum games is straightforward!

Outline

- Introduction
  - Two-player zero-sum games, Nash equilibrium
  - Nash equilibrium and no-regret dynamics
  - Fast convergence in games
  - Research questions
- Learning in corrupted two-player zero-sum games
- Lower bounds
- Learning in corrupted multi-player general-sum games
- Conclusion and discussion

Main result (2)-(i): Lower bound in terms of $\widetilde{C}_x$ and $\widetilde{C}_y$

Defining the regret $\mathrm{Reg}_{x,\widetilde{g}}^T$ w.r.t. the corrupted gradients $\widetilde{g}_1, \ldots, \widetilde{g}_T$, we can show the following upper bound (omitted in the theorem above):

$$\mathrm{Reg}_{x,\widetilde{g}}^T \lesssim \min\left\{\sqrt{(\log(m_x m_y) + C_x + C_y) \log m_x}, \sqrt{T \log m_x}\right\} + \widehat{C}_x .$$

If corruption occurs only in x-player's observed utilities (*i.e.*, $\widehat{C}_x = \widehat{C}_y = \widetilde{C}_y = 0$),

$$\mathrm{Reg}_{x,\widetilde{g}}^T = O(\sqrt{\widetilde{C}_x \log m_x}) ,$$

which matches the following lower bound:

Main result (2)-(i): Lower bound in terms of $\widetilde{C}_x$ and $\widetilde{C}_y$

Defining the regret $\mathrm{Reg}^T_{x,\widetilde{g}}$ w.r.t. the corrupted gradients $\widetilde{g}_1, \dots, \widetilde{g}_T$, we can show the following upper bound (omitted in the theorem above):

$$\mathrm{Reg}^T_{x,\widetilde{g}} \lesssim \min\left\{\sqrt{(\log(m_x m_y) + C_x + C_y) \log m_x}, \sqrt{T \log m_x}\right\} + \widehat{C}_x \,.$$

If corruption occurs only in x-player's observed utilities (*i.e.*, $\widehat{C}_x = \widehat{C}_y = \widetilde{C}_y = 0$),

$$\mathrm{Reg}^T_{x,\widetilde{g}} = O(\sqrt{\widetilde{C}_x \log m_x}) \,,$$

which matches the following lower bound:

Theorem (Lower bounds in the corrupted regime)

For any learning dynamics,

(i) *there exists a corrupted game with* $\sum_{t=1}^T \|g^{(t)} - \widetilde{g}^{(t)}\|_\infty \le \widetilde{C}_x$ *such that* $\mathrm{Reg}^T_{x,\widetilde{g}} = \mathrm{Reg}^T_{\widehat{x},\widetilde{g}} = \Omega(\sqrt{\widetilde{C}_x \log m_x})$;

(there exists a corrupted game with $\sum_{t=1}^T \|\ell^{(t)} - \widetilde{\ell}^{(t)}\|_\infty \le \widetilde{C}_y$ *such that* $\mathrm{Reg}^T_{y,\widetilde{\ell}} = \mathrm{Reg}^T_{\widehat{y},\widetilde{\ell}} = \Omega(\sqrt{\widetilde{C}_y \log m_y})$.*)*

Main result (2)-(i): Lower bound in terms of $\widetilde{C}_x$ and $\widetilde{C}_y$

Construct a corrupted game with $\sum_{t=1}^{T}\|g^{(t)} - \widetilde{g}^{(t)}\|_\infty \le \widetilde{C}_x$ such that $\mathrm{Reg}_{x,\widetilde{g}}^{T} = \mathrm{Reg}_{\widehat{x},\widetilde{g}}^{T} = \Omega(\sqrt{\widetilde{C}_x \log m_x})$.

Idea. Let $A = 0$ and use the following lower bound for online linear optimization over simplex:

$$\forall\, \mathrm{Alg},\ \exists\, \widetilde{g}^{(1)}, \dots, \widetilde{g}^{(T_0)} \in [0,1]^{m_x},\ \max_{x^* \in \Delta_{m_x}} \sum_{t=1}^{T_0} \langle x^* - x^{(t)}, \widetilde{g}^{(t)} \rangle = \Omega(\sqrt{T_0 \log m_x})\,.$$

Main result (2)-(i): Lower bound in terms of $\widetilde{C}_x$ and $\widetilde{C}_y$

Construct a corrupted game with $\sum_{t=1}^{T} \|g^{(t)} - \widetilde{g}^{(t)}\|_\infty \le \widetilde{C}_x$ such that $\mathrm{Reg}_{x,\widetilde{g}}^{T} = \mathrm{Reg}_{\widehat{x},\widetilde{g}}^{T} = \Omega(\sqrt{\widetilde{C}_x \log m_x})$.

Idea. Let $A = 0$ and use the following lower bound for online linear optimization over simplex:

$$\forall\, \mathrm{Alg},\ \exists\, \widetilde{g}^{(1)}, \ldots, \widetilde{g}^{(T_0)} \in [0,1]^{m_x},\ \max_{x^* \in \Delta_{m_x}} \sum_{t=1}^{T_0} \langle x^* - x^{(t)}, \widetilde{g}^{(t)} \rangle = \Omega(\sqrt{T_0 \log m_x})\,.$$

Proof. For rounds $t = 1, \ldots, \widetilde{C}_x/2$, the expected reward vectors $g^{(t)}$ are corrupted so that $\sum_{t=1}^{\widetilde{C}_x/2} \|g^{(t)} - \widetilde{g}^{(t)}\|_\infty \le \widetilde{C}_x$, and no corruption occurs beyond this. Then, since $A = 0$,

$$\mathrm{Reg}_{x,\widetilde{g}}^{T} = \max_{x^* \in \Delta_{m_x}} \sum_{t=1}^{T} \langle x^* - x^{(t)}, \widetilde{g}^{(t)} \rangle = \max_{x^* \in \Delta_{m_x}} \left\{ \sum_{t=1}^{\widetilde{C}_x/2} \langle x^* - x^{(t)}, \widetilde{g}^{(t)} \rangle + \sum_{t=\widetilde{C}_x/2+1}^{T} \langle x^* - x^{(t)}, A y^{(t)} \rangle \right\}$$

$$= \max_{x^* \in \Delta_{m_x}} \sum_{t=1}^{\widetilde{C}_x/2} \langle x^* - x^{(t)}, \widetilde{g}^{(t)} \rangle \ge \Omega\left(\sqrt{\widetilde{C}_x \log m_x} \right). \quad \square$$

Main result (2)-(ii): Lower bound for the player's own strategy deviation

Regret caused by the player's own deviation from the suggested strategies $\widehat{x}^{(t)}, \widehat{y}^{(t)}$

$$\mathrm{Reg}_{x,g}^{T} \lesssim \min\left\{\sqrt{(\log(m_x m_y) + C_x + C_y) \log m_x}, \sqrt{T \log m_x}\right\} + C_x$$

If corruption occurs only in x-player's strategies (*i.e.*, $\widehat{C}_y = \widetilde{C}_x = \widetilde{C}_y = 0$),

$$\mathrm{Reg}_{x,g}^{T} = \widetilde{O}(\widehat{C}_x),$$

which matches the following lower bound:

Main result (2)-(ii): Lower bound for the player's own strategy deviation

Regret caused by the player's own deviation from the suggested strategies $\widehat{x}^{(t)}, \widehat{y}^{(t)}$

$$\mathrm{Reg}_{x,g}^{T} \lesssim \min\left\{\sqrt{(\log(m_x m_y) + C_x + C_y) \log m_x}, \sqrt{T \log m_x}\right\} + C_x$$

If corruption occurs only in x-player's strategies (*i.e.*, $\widehat{C}_y = \widetilde{C}_x = \widetilde{C}_y = 0$),

$$\mathrm{Reg}_{x,g}^{T} = \widetilde{O}(\widehat{C}_x)\,,$$

which matches the following lower bound:

Theorem (Lower bounds in the corrupted regime)

For any learning dynamics,

(ii) *there exists a corrupted game with $\sum_{t=1}^{T} \|x^{(t)} - \widehat{x}^{(t)}\|_1 \leq \widehat{C}_x$ such that $\mathrm{Reg}_{x,g}^{T} = \mathrm{Reg}_{x,\widetilde{g}}^{T} = \Omega(\widehat{C}_x)$;*

(there exists a corrupted game with $\sum_{t=1}^{T} \|y^{(t)} - \widehat{y}^{(t)}\|_1 \leq \widehat{C}_y$ such that $\mathrm{Reg}_{y,\ell}^{T} = \mathrm{Reg}_{y,\widetilde{\ell}}^{T} = \Omega(\widehat{C}_y)$.)

Main result (2)-(ii): Lower bound for the player's own strategy deviation

Construct a corrupted game with $\sum_{t=1}^{T} \|x^{(t)} - \widehat{x}^{(t)}\|_1 \leq \widehat{C}_x$ such that $\mathrm{Reg}_{x,g}^{T} = \mathrm{Reg}_{x,\widetilde{g}}^{T} = \Omega(\widehat{C}_x)$.

Idea. Construct a payoff matrix with an action with a low reward, and then design corrupted strategies that select the action. In particular, consider

$$A = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ & & \vdots & \\ 1 & 1 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \end{pmatrix} \quad \text{and} \quad x^{(t)} = \begin{cases} \widehat{x}^{(t)} + \widehat{c}_x^{(t)} = e_{m_x} & t = 1, \ldots, \widehat{C}_x/2 \\ \widehat{x}^{(t)} & t = \widehat{C}_x/2 + 1, \ldots, T \end{cases}$$

Main result (2)-(ii): Lower bound for the player's own strategy deviation

Construct a corrupted game with $\sum_{t=1}^{T} \|x^{(t)} - \widehat{x}^{(t)}\|_1 \leq \widehat{C}_x$ such that $\mathrm{Reg}_{x,g}^{T} = \mathrm{Reg}_{x,\widetilde{g}}^{T} = \Omega(\widehat{C}_x)$.

Idea. Construct a payoff matrix with an action with a low reward, and then design corrupted strategies that select the action. In particular, consider

$$A = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ & & \vdots & \\ 1 & 1 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \end{pmatrix} \quad \text{and} \quad x^{(t)} = \begin{cases} \widehat{x}^{(t)} + \widehat{c}_x^{(t)} = e_{m_x} & t = 1, \ldots, \widehat{C}_x/2 \\ \widehat{x}^{(t)} & t = \widehat{C}_x/2 + 1, \ldots, T \end{cases}$$

Proof. For each $t = 1, \ldots, \widehat{C}_x/2$, we also have $Ay^{(t)} = \mathbf{1} - e_{m_x}$ and $\langle x^{(t)}, Ay^{(t)} \rangle = 0$ since $A^\top x^{(t)} = 0$.

Hence, for any $x^* \in \Delta_{m_x}$,

$$\sum_{t=1}^{\widehat{C}_x/2} \langle x^* - x^{(t)}, g^{(t)} \rangle = \sum_{t=1}^{\widehat{C}_x/2} \langle x^*, Ay^{(t)} \rangle = \sum_{t=1}^{\widehat{C}_x/2} \langle x^*, \mathbf{1} - e_{m_x} \rangle = \frac{\widehat{C}_x}{2}(1 - x^*(m_x)), \tag{1}$$

where we used $\langle x^{(t)}, Ay^{(t)} \rangle = 0$, $Ay^{(t)} = \mathbf{1} - e_{m_x}$, and $x^* \in \Delta_{m_x}$. Therefore,

$$\begin{aligned}
\mathrm{Reg}_{x,g}^T &= \max_{x^* \in \Delta_{m_x}} \sum_{t=1}^{T} \langle x^* - x^{(t)}, g^{(t)} \rangle \\
&= \max_{x^* \in \Delta_{m_x}} \left\{ \sum_{t=1}^{\widehat{C}_x/2} \langle x^* - x^{(t)}, g^{(t)} \rangle + \sum_{t=\widehat{C}_x/2+1}^{T} \langle x^* - x^{(t)}, Ay^{(t)} \rangle \right\} \\
\text{(by (1))} &= \max_{x^* \in \Delta_{m_x}} \left\{ \frac{\widehat{C}_x}{2}(1 - x^*(m_x)) + \sum_{t=\widehat{C}_x/2+1}^{T} \langle x^* - x^{(t)}, \mathbf{1} - e_{m_x} \rangle \right\} \\
&= \max_{x^* \in \Delta_{m_x}} \left\{ \frac{\widehat{C}_x}{2}(1 - x^*(m_x)) + \sum_{t=\widehat{C}_x/2+1}^{T} (x^{(t)}(m_x) - x^*(m_x)) \right\} \geq \frac{\widehat{C}_x}{2}. \quad \square
\end{aligned}$$

Main result (2)-(iii): Lower bound for the opponent's strategy deviation

Our upper bound: $\mathrm{Reg}_{x,g}^{T} \lesssim \min\left\{\sqrt{(\log(m_x m_y) + C_x + C_y)\log m_x}, \sqrt{T \log m_x}\right\} + C_x$

If corruption occurs only in y-player's strategies (*i.e.*, $\widehat{C}_x = \widetilde{C}_x = \widetilde{C}_y = 0$),

$$\mathrm{Reg}_{\widehat{x},g}^{T} = \widetilde{O}\left(\sqrt{\widehat{C}_y}\right), \quad \mathrm{Reg}_{\widehat{y},\ell}^{T} = \widetilde{O}\left(\sqrt{\widehat{C}_y}\right),$$

which matches the following lower bound:

Main result (2)-(iii): Lower bound for the opponent's strategy deviation

Our upper bound: $\mathrm{Reg}_{x,g}^{T} \lesssim \min\left\{\sqrt{(\log(m_x m_y) + C_x + C_y) \log m_x}, \sqrt{T \log m_x}\right\} + C_x$

If corruption occurs only in y-player's strategies (*i.e.*, $\widehat{C}_x = \widetilde{C}_x = \widetilde{C}_y = 0$),

$$\mathrm{Reg}_{\widehat{x},g}^{T} = \widetilde{O}\left(\sqrt{\widehat{C}_y}\right), \quad \mathrm{Reg}_{\widehat{y},\ell}^{T} = \widetilde{O}\left(\sqrt{\widehat{C}_y}\right),$$

which matches the following lower bound:

Theorem (Lower bounds in the corrupted regime)

For any learning dynamics,

(iii) *there exists a corrupted game with* $\sum_{t=1}^{T} \|y^{(t)} - \widehat{y}^{(t)}\|_1 \leq \widehat{C}_y$ *such that* $\max\left\{\mathrm{Reg}_{\widehat{x},g}^{T}, \mathrm{Reg}_{\widehat{y},\ell}^{T}\right\} = \Omega\left(\sqrt{\widehat{C}_y}\right)$;

(there exists a corrupted game with $\sum_{t=1}^{T} \|x^{(t)} - \widehat{x}^{(t)}\|_1 \leq \widehat{C}_x$ *such that* $\max\left\{\mathrm{Reg}_{\widehat{x},g}^{T}, \mathrm{Reg}_{\widehat{y},\ell}^{T}\right\} = \Omega(\sqrt{\widehat{C}_x})$*.)*

Similar to the lower bounds of Syrgkanis et al. (2015) and Chen and Peng (2020), but their bounds are for Hedge and are not for corrupted games

Main result (2)-(iii): Lower bound for the opponent's strategy deviation

Construct a corrupted game with $\sum_{t=1}^{T} \|y^{(t)} - \widehat{y}^{(t)}\|_1 \leq \widehat{C}_y$ such that $\max\{\mathrm{Reg}_{\widehat{x},g}^{T}, \mathrm{Reg}_{\widehat{y},\ell}^{T}\} = \Omega(\sqrt{\widehat{C}_y})$.

Proof sketch. It suffices to prove

$$\exists \text{ absolute const } \kappa > 0,\ \mathrm{Reg}_{\widehat{y},\ell}^{T} < \kappa\sqrt{\widehat{C}_y} \implies \mathrm{Reg}_{\widehat{x},g}^{T} \geq \kappa\sqrt{\widehat{C}_x}\,.$$

Consider $A = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}$ optimal action of y-player

For $t = 1, \ldots, \widehat{C}_y/2$, use the lower bound for online linear optimization over simplex:

$$\max_{x \in \Delta_{m_x}} \sum_{t=1}^{\widehat{C}_y/2} \langle x - \widehat{x}^{(t)}, \check{g}^{(t)} \rangle \geq \frac{1}{2}\sqrt{\widehat{C}_y}\,. \quad (\text{choose } y^{(t)} \text{ such that } \check{g}^{(t)} = Ay^{(t)})$$

For $t = \widehat{C}_y/2 + 1, \ldots, T$, y-player can select actions 1 and 2 at most $\kappa\sqrt{\widehat{C}_y}$ times.

x-player's regret after round $t = \widehat{C}_y/2 + 1$ is lower bounded by $-\kappa\sqrt{\widehat{C}_y}$.

Choosing $\kappa = 1/4$ gives $\mathrm{Reg}_{\widehat{x},g}^{T} \geq \frac{1}{2}\sqrt{\widehat{C}_y} - \kappa\sqrt{\widehat{C}_y} = \kappa\sqrt{\widehat{C}_y}\,.$ □

Outline

- Introduction
  - Two-player zero-sum games, Nash equilibrium
  - Nash equilibrium and no-regret dynamics
  - Fast convergence in games
  - Research questions
- Learning in corrupted two-player zero-sum games
- Lower bounds
- Learning in corrupted multi-player general-sum games
- Conclusion and discussion

Multi-player general-sum games

Zero-sum games: Players have exactly opposed interests
General-sum games: Players may have partially aligned interests

Multi-player general-sum games

Zero-sum games: Players have exactly opposed interests
General-sum games: Players may have partially aligned interests

- $n \geq 2$: the number of players
- Each player $i \in [n]$ has an *action set* $\mathcal{A}_i$ with $|\mathcal{A}_i| = m_i$ and a *utility function* $u_i : \mathcal{A}_1 \times \cdots \times \mathcal{A}_n \to [-1, 1]$

Multi-player general-sum games

Zero-sum games: Players have exactly opposed interests
General-sum games: Players may have partially aligned interests

- $n \geq 2$: the number of players
- Each player $i \in [n]$ has an *action set* $\mathcal{A}_i$ with $|\mathcal{A}_i| = m_i$ and a *utility function* $u_i \colon \mathcal{A}_1 \times \cdots \times \mathcal{A}_n \to [-1, 1]$

1. Each player $i \in [n]$ selects a strategy $x_i \in \Delta_{m_i}$;

Multi-player general-sum games

Zero-sum games: Players have exactly opposed interests
General-sum games: Players may have partially aligned interests

- $n \geq 2$: the number of players
- Each player $i \in [n]$ has an *action set* $\mathcal{A}_i$ with $|\mathcal{A}_i| = m_i$ and a *utility function* $u_i \colon \mathcal{A}_1 \times \cdots \times \mathcal{A}_n \to [-1, 1]$

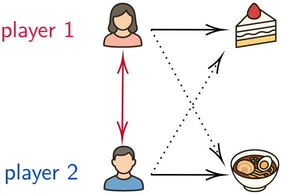
1. Each player $i \in [n]$ selects a strategy $x_i \in \Delta_{m_i}$;
2. Each player i gains a reward of $\mathbb{E}_{a_1 \sim x_1, \ldots, a_n \sim x_n}[u_i(a_1, \ldots, a_n)]$.

Example of general-sum games

Example 1. Lunch dilemma (known as battle of the sexes, Bach or Stravinsky)
Players 1 and 2 want to have lunch together,
but have a choice between two restaurants (Cake or Ramen) to go

Example of general-sum games

Example 1. Lunch dilemma (known as battle of the sexes, Bach or Stravinsky)
Players 1 and 2 want to have lunch together,
but have a choice between two restaurants (Cake or Ramen) to go



player 2

player 1

	3, 2	0, 0
	0, 0	2, 3

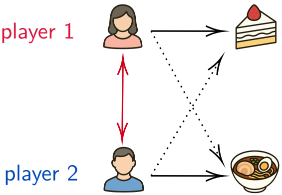
utility functions

$$u_1 = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}$$

$$u_2 = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$$

Example of general-sum games

Example 1. Lunch dilemma (known as battle of the sexes, Bach or Stravinsky)
Players 1 and 2 want to have lunch together,
but have a choice between two restaurants (Cake or Ramen) to go



player 2

player 1		
	3, 2	0, 0
	0, 0	2, 3

utility functions

$$u_1 = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}$$

$$u_2 = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$$

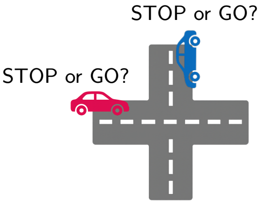
Nash eq: (pure) (Cake,Cake), (Ramen,Ramen),
(mixed) Cake with prob. 3/5 and Ramen with prob. 2/5 (for player 1)

Example of general-sum games (cont'd)

Example 2. Game of chicken: two-player "STOP-or-GO" at intersection game

Example of general-sum games (cont'd)

Example 2. Game of chicken: two-player "STOP-or-GO" at intersection game



player 1 \ player 2	STOP	GO
STOP	0, 0	0, 1
GO	1, 0	−100, −100

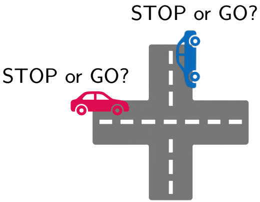
utility functions

$$u_1 = \begin{pmatrix} 0 & 0 \\ 1 & -100 \end{pmatrix}$$

$$u_2 = \begin{pmatrix} 0 & 1 \\ 0 & -100 \end{pmatrix}$$

Example of general-sum games (cont'd)

Example 2. Game of chicken: two-player "STOP-or-GO" at intersection game



player 1 \ player 2	STOP	GO
STOP	0, 0	0, 1
GO	1, 0	−100, −100

utility functions

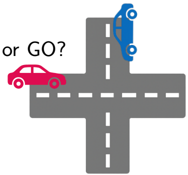
$$u_1 = \begin{pmatrix} 0 & 0 \\ 1 & -100 \end{pmatrix}$$

$$u_2 = \begin{pmatrix} 0 & 1 \\ 0 & -100 \end{pmatrix}$$

Nash eq: (pure) (STOP,GO), (GO,STOP),
(mixed) STOP with prob. $100/101$ and GO with prob. $1/101$

Example of general-sum games (cont'd)

Example 2. Game of chicken: two-player "STOP-or-GO" at intersection game



player 1 \ player 2	STOP	GO
STOP	$0, 0$	$0, 1$
GO	$1, 0$	$-100, -100$

utility functions

$$u_1 = \begin{pmatrix} 0 & 0 \\ 1 & -100 \end{pmatrix}$$

$$u_2 = \begin{pmatrix} 0 & 1 \\ 0 & -100 \end{pmatrix}$$

Nash eq: (pure) (STOP,GO), (GO,STOP),
(mixed) STOP with prob. $100/101$ and GO with prob. $1/101$

- ☹ In the mixed strategy, both players compromise too much and get low payoffs.
- ☹ (Believed that) Nash eq cannot be computed in polynomial time w.r.t. the action size

Correlated equilibrium

Correlated equilibrium: A probability distribution over $\times_{i=1}^{n}\mathcal{A}_i$ such that following the "signal" is always the best, no matter how a player considers modifying their response.

Correlated equilibrium

Correlated equilibrium: A probability distribution over $\times_{i=1}^{n}\mathcal{A}_i$ such that following the "signal" is always the best, no matter how a player considers modifying their response.

Definition (Correlated equilibrium, Aumann 1974)

A probability distribution σ over action sets $\times_{i=1}^{n}\mathcal{A}_i$ is an ε-*approximate correlated equilibrium* if for any player $i \in [n]$ and any (swap) function $\phi_i \colon \mathcal{A}_i \to \mathcal{A}_i$,

$$\mathbb{E}_{a\sim\sigma}[u_i(a)] \geq \mathbb{E}_{a\sim\sigma}[u_i(\phi_i(a_i), a_{-i})] - \varepsilon\,,$$

where $a_{-i} = (a_1, \ldots, a_{i-1}, a_{i+1}, \ldots a_n)$.

Correlated equilibrium

Correlated equilibrium: A probability distribution over $\times_{i=1}^{n}\mathcal{A}_i$ such that following the "signal" is always the best, no matter how a player considers modifying their response.

Definition (Correlated equilibrium, Aumann 1974)

A probability distribution σ over action sets $\times_{i=1}^{n}\mathcal{A}_i$ is an ε-*approximate correlated equilibrium* if for any player $i \in [n]$ and any (swap) function $\phi_i \colon \mathcal{A}_i \to \mathcal{A}_i$,

$$\mathbb{E}_{a\sim\sigma}[u_i(a)] \geq \mathbb{E}_{a\sim\sigma}[u_i(\phi_i(a_i), a_{-i})] - \varepsilon\,,$$

where $a_{-i} = (a_1, \ldots, a_{i-1}, a_{i+1}, \ldots a_n)$.

Examples of correlated equilibrium

- Lunch dilemma: flip a coin (observable to both Players 1 and 2); if heads, choose (Cake, Cake), if tails, choose (Ramen, Ramen).
- Game of chicken: a traffic light that outputs (STOP, GO) or (GO, STOP) with equal probability.

Learning in multi-player general-sum games

- $n \geq 2$: the number of players
- Each player $i \in [n]$ has an *action set* $\mathcal{A}_i$ with $|\mathcal{A}_i| = m_i$ and a *utility function* $u_i \colon \mathcal{A}_1 \times \cdots \times \mathcal{A}_n \to [-1, 1]$

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

1. Each player $i \in [n]$ selects a strategy $x_i^{(t)} \in \Delta_{m_i}$;

Learning in multi-player general-sum games

- $n \geq 2$: the number of players
- Each player $i \in [n]$ has an *action set* $\mathcal{A}_i$ with $|\mathcal{A}_i| = m_i$ and a *utility function* $u_i \colon \mathcal{A}_1 \times \cdots \times \mathcal{A}_n \to [-1, 1]$

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

1. Each player $i \in [n]$ selects a strategy $x_i^{(t)} \in \Delta_{m_i}$;
2. Each player i observes a utility vector $u_i^{(t)} \in [-1, 1]^{m_i}$, where $u_i^{(t)}(a_i) = \mathbb{E}_{a_{-i} \sim x_{-i}^{(t)}}[u_i(a_i, a_{-i})]$. (recall $Ay^{(t)}$ in zero-sum games);

Learning in multi-player general-sum games

- $n \geq 2$: the number of players
- Each player $i \in [n]$ has an *action set* $\mathcal{A}_i$ with $|\mathcal{A}_i| = m_i$ and a *utility function* $u_i \colon \mathcal{A}_1 \times \cdots \times \mathcal{A}_n \to [-1, 1]$

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

1. Each player $i \in [n]$ selects a strategy $x_i^{(t)} \in \Delta_{m_i}$;
2. Each player i observes a utility vector $u_i^{(t)} \in [-1, 1]^{m_i}$, where $u_i^{(t)}(a_i) = \mathbb{E}_{a_{-i} \sim x_{-i}^{(t)}}[u_i(a_i, a_{-i})]$. (recall $Ay^{(t)}$ in zero-sum games);
3. Each player i gains a reward of $\langle x_i^{(t)}, u_i^{(t)} \rangle$.

Learning in multi-player general-sum games

- $n \geq 2$: the number of players
- Each player $i \in [n]$ has an *action set* $\mathcal{A}_i$ with $|\mathcal{A}_i| = m_i$ and a *utility function* $u_i : \mathcal{A}_1 \times \cdots \times \mathcal{A}_n \to [-1, 1]$

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

1. Each player $i \in [n]$ selects a strategy $x_i^{(t)} \in \Delta_{m_i}$;
2. Each player i observes a utility vector $u_i^{(t)} \in [-1, 1]^{m_i}$, where $u_i^{(t)}(a_i) = \mathbb{E}_{a_{-i} \sim x_{-i}^{(t)}}[u_i(a_i, a_{-i})]$. (recall $Ay^{(t)}$ in zero-sum games);
3. Each player i gains a reward of $\langle x_i^{(t)}, u_i^{(t)} \rangle$.

The goal of each player i is to minimize the regret (w/o knowing utilities $\{u_i\}_{i \in [n]}$):

$$\mathrm{Reg}^T_{x_i, u_i} = \max_{x^* \in \Delta_{m_x}} \left\{ \sum_{t=1}^{T} \langle x^*, u_i^{(t)} \rangle - \sum_{t=1}^{T} \langle x_i^{(t)}, u_i^{(t)} \rangle \right\}$$

Learning in multi-player general-sum games

- $n \geq 2$: the number of players
- Each player $i \in [n]$ has an *action set* $\mathcal{A}_i$ with $|\mathcal{A}_i| = m_i$ and a *utility function* $u_i \colon \mathcal{A}_1 \times \cdots \times \mathcal{A}_n \to [-1, 1]$

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

1. Each player $i \in [n]$ selects a strategy $x_i^{(t)} \in \Delta_{m_i}$;
2. Each player i observes a utility vector $u_i^{(t)} \in [-1, 1]^{m_i}$, where $u_i^{(t)}(a_i) = \mathbb{E}_{a_{-i} \sim x_{-i}^{(t)}}[u_i(a_i, a_{-i})]$. (recall $Ay^{(t)}$ in zero-sum games);
3. Each player i gains a reward of $\langle x_i^{(t)}, u_i^{(t)} \rangle$.

The goal of each player i is to minimize the **swap regret** (w/o knowing utilities $\{u_i\}_{i \in [n]}$):

$$\mathrm{SwapReg}_{x_i, u_i}^T = \max_{M \in \mathcal{M}_{m_i}} \sum_{t=1}^T \langle x_i^{(t)}, M u_i^{(t)} - u_i^{(t)} \rangle ,$$

where $\mathcal{M}_m = \{M \in [0,1]^{m \times m} \colon M(k, \cdot) \in \Delta_m \text{ for } k \in [m]\}$. "I should've played strategy $M^\top x_i^{(t)}$ instead of $x_i^{(t)}$ ..."

No-swap-regret learning dynamics and correlated equilibrium

A probability distribution σ over action sets $\times_{i=1}^{n} \mathcal{A}_i$ is an ε*-approximate correlated equilibrium* if for any player $i \in [n]$ and any (swap) function $\phi_i \colon \mathcal{A}_i \to \mathcal{A}_i$,

$$\mathbb{E}_{a \sim \sigma}[u_i(a)] \geq \mathbb{E}_{a \sim \sigma}[u_i(\phi_i(a_i), a_{-i})] - \varepsilon.$$

Theorem (Foster and Vohra 1997)

Let $\sigma^{(t)} = \otimes_{i \in [n]} x_i^{(t)} \in \Delta(\times_{i=1}^{n} \mathcal{A}_i)$ *given by* $\sigma^{(t)}(a_1, \ldots, a_n) = \prod_{i \in [n]} x_i^{(t)}(a_i)$ *for each* $a_i \in \mathcal{A}_i$ *be the joint distribution at round* t*. Then, its time-averaged distribution* $\sigma = \frac{1}{T} \sum_{t=1}^{T} \sigma^{(t)}$ *is a* $(\max_{i \in [n]} \mathrm{SwapReg}_{x_i, u_i}^{T} / T)$*-approximate correlated equilibrium.*

Corrupted regime in multi-player general-sum games

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

1. A prescribed algorithm suggests a strategy $\widehat{x}_i^{(t)} \in \Delta_{m_i}$ for each player $i \in [n]$;
2. **(corruption of strategies)**
   Each player $i \in [n]$ selects a strategy $x_i^{(t)} \leftarrow \widehat{x}_i^{(t)} + \widehat{c}_i^{(t)}$;
3. **(corruption of utilities)**
   Each player i observes a corrupted utility vector $\widetilde{u}_i^{(t)} \leftarrow u_i^{(t)} + \widetilde{c}_i^{(t)}$;
4. Each player i gains a reward of $\langle x_i^{(t)}, u_i^{(t)} \rangle$ or $\langle x^{(t)}, \widetilde{u}_i^{(t)} \rangle$

Corrupted regime in multi-player general-sum games

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

1. A prescribed algorithm suggests a strategy $\widehat{x}_i^{(t)} \in \Delta_{m_i}$ for each player $i \in [n]$;
2. **(corruption of strategies)**
   Each player $i \in [n]$ selects a strategy $x_i^{(t)} \leftarrow \widehat{x}_i^{(t)} + \widehat{c}_i^{(t)}$;
3. **(corruption of utilities)**
   Each player i observes a corrupted utility vector $\widetilde{u}_i^{(t)} \leftarrow u_i^{(t)} + \widetilde{c}_i^{(t)}$;
4. Each player i gains a reward of $\langle x_i^{(t)}, u_i^{(t)} \rangle$ or $\langle x^{(t)}, \widetilde{u}_i^{(t)} \rangle$

- the corrupted regime with no corruptions = the honest regime
- the corrupted regime with arbitrary strategies by the opponent players $j \neq [n] \setminus \{i\}$ = the adversarial scenario for player i

Swap regret minimization to regret minimization

A well-known reduction due to Blum and Mansour (2007), informal:

1. Run m_i external regret minimizers (one for each action) for each player i (with utility vector $\widetilde{u}_{i,a}^{(s)} = \widehat{x}_i^{(s)}(a)\widetilde{u}_i^{(s)}$ ($s < t$) for a-th minimizer output $\widehat{x}_i^{(t)}(a)$);

Swap regret minimization to regret minimization

A well-known reduction due to Blum and Mansour (2007), informal:

1. Run m_i external regret minimizers (one for each action) for each player i (with utility vector $\widetilde{u}_{i,a}^{(s)} = \widehat{x}_i^{(s)}(a)\widetilde{u}_i^{(s)}$ $(s < t)$ for a-th minimizer output $\widehat{x}_i^{(t)}(a)$);
2. Define a Markov chain $Q_i^{(t)}$ from the outputs $y_{i,a}^{(t)} \in \Delta_{m_i}$ for each $a \in \mathcal{A}_i$ (Use $Q_i^{(t)} \in [0,1]^{m_i \times m_i}$ whose a-th row is $y_{i,a}^{(t)}$ for each $a \in \mathcal{A}_i$);

Swap regret minimization to regret minimization

A well-known reduction due to Blum and Mansour (2007), informal:

1. Run m_i external regret minimizers (one for each action) for each player i (with utility vector $\widetilde{u}_{i,a}^{(s)} = \widehat{x}_i^{(s)}(a)\widetilde{u}_i^{(s)}$ $(s < t)$ for a-th minimizer output $\widehat{x}_i^{(t)}(a)$);
2. Define a Markov chain $Q_i^{(t)}$ from the outputs $y_{i,a}^{(t)} \in \Delta_{m_i}$ for each $a \in \mathcal{A}_i$ (Use $Q_i^{(t)} \in [0,1]^{m_i \times m_i}$ whose a-th row is $y_{i,a}^{(t)}$ for each $a \in \mathcal{A}_i$);
3. Use a stationary distribution of the Markov chain induced by $Q_i^{(t)}$ as $\widehat{x}_i^{(t)}$.

Swap regret minimization to regret minimization

A well-known reduction due to Blum and Mansour (2007), informal:

1. Run m_i external regret minimizers (one for each action) for each player i (with utility vector $\widetilde{u}_{i,a}^{(s)} = \widehat{x}_i^{(s)}(a)\widetilde{u}_i^{(s)}$ $(s < t)$ for a-th minimizer output $\widehat{x}_i^{(t)}(a)$);
2. Define a Markov chain $Q_i^{(t)}$ from the outputs $y_{i,a}^{(t)} \in \Delta_{m_i}$ for each $a \in \mathcal{A}_i$ (Use $Q_i^{(t)} \in [0,1]^{m_i \times m_i}$ whose a-th row is $y_{i,a}^{(t)}$ for each $a \in \mathcal{A}_i$);
3. Use a stationary distribution of the Markov chain induced by $Q_i^{(t)}$ as $\widehat{x}_i^{(t)}$.

The swap regret is the sum of the external regret of external regret minimizer $a \in \mathcal{A}_i$.

Lemma (Blum and Mansour 2007)

Define $\widetilde{u}_{i,a}^{(t)} = \widehat{x}_i^{(t)}(a)\widetilde{u}_i^{(t)}$ *and* $\widetilde{\mathrm{Reg}}_{i,a}^{T}(y^*) = \sum_{t=1}^{T} \langle y^* - y_{i,a}^{(t)}, \widetilde{u}_{i,a}^{(t)} \rangle$. *Then,*

$$\mathrm{SwapReg}_{\widehat{x}_i, \widetilde{u}_i}^{T}(M) = \sum_{a \in \mathcal{A}_i} \widetilde{\mathrm{Reg}}_{i,a}^{T}(M(a, \cdot)) \,.$$

Our algorithm: OFTRL with log-barrier and adaptive lr

Sufficient to construct a no-external-regret algorithm that aims to minimize the external regret $\widetilde{\mathrm{Reg}}_{i,a}^{T}(M(a,\cdot))$.

Our algorithm: OFTRL with log-barrier and adaptive lr

Sufficient to construct a no-external-regret algorithm that aims to minimize the external regret $\widetilde{\mathrm{Reg}}_{i,a}^{T}(M(a,\cdot))$.

OFTRL with log-barrier $\phi(x) = -\sum_k \log(x(k))$ and adaptive learning rate: compute $y_{i,a}^{(t)} \in \Delta_{m_i}$ for each $a \in \mathcal{A}_i$ by

$$y_{i,a}^{(t)} = \operatorname*{arg\,max}_{y \in \Delta_{m_i}} \left\{ \left\langle y, \widetilde{u}_{i,a}^{(t-1)} + \sum_{s=1}^{t-1} \widetilde{u}_{i,a}^{(s)} \right\rangle - \frac{\phi(y)}{\eta_{i,a}^{(t)}} \right\}, \quad \eta_{i,a}^{(t)} = \min\left\{ \sqrt{\frac{m_i \log T/8}{4 + \sum_{s=1}^{t-1} \|\widetilde{u}_{i,a}^{(s)} - \widetilde{u}_{i,a}^{(s-1)}\|_\infty^2}}, \bar{\eta}_i \right\},$$

where $\widetilde{u}_{i,a}^{(t)} = \widehat{x}_i^{(t)}(a)\widetilde{u}_i^{(t)}$ and $\bar{\eta}_i = \frac{1}{256 n \sqrt{m_i}}$.

Our algorithm: OFTRL with log-barrier and adaptive lr

Sufficient to construct a no-external-regret algorithm that aims to minimize the external regret $\widetilde{\mathrm{Reg}}^{T}_{i,a}(M(a,\cdot))$.

OFTRL with log-barrier $\phi(x) = -\sum_k \log(x(k))$ and adaptive learning rate: compute $y^{(t)}_{i,a} \in \Delta_{m_i}$ for each $a \in \mathcal{A}_i$ by

$$y^{(t)}_{i,a} = \operatorname*{arg\,max}_{y \in \Delta_{m_i}} \left\{ \left\langle y, \widetilde{u}^{(t-1)}_{i,a} + \sum_{s=1}^{t-1} \widetilde{u}^{(s)}_{i,a} \right\rangle - \frac{\phi(y)}{\eta^{(t)}_{i,a}} \right\}, \ \eta^{(t)}_{i,a} = \min\left\{ \sqrt{\frac{m_i \log T/8}{4 + \sum_{s=1}^{t-1} \|\widetilde{u}^{(s)}_{i,a} - \widetilde{u}^{(s-1)}_{i,a}\|_\infty^2}}, \bar{\eta}_i \right\},$$

where $\widetilde{u}^{(t)}_{i,a} = \widehat{x}^{(t)}_i(a)\widetilde{u}^{(t)}_i$ and $\bar{\eta}_i = \frac{1}{256 n \sqrt{m_i}}$.

We use the *expert-wise* adaptive learning rate $\eta^{(t)}_{i,a}$ for each player $i \in [n]$ and $a \in \mathcal{A}_i$, while Anagnostides et al. (2022) uses a constant common learning rate, $\eta^{(t)}_{i,a} = \eta_i$.

Main result (3): Swap regret bound in the corrupted regime

Swap regret upper bounds of player i in multi-player general-sum games with n-players and m-actions after T rounds

$\widehat{C}_i \in [0, 2T]$: the cumulative amount of corruption in strategies for player i, $\widehat{S} = \sum_{i \in [n]} \widehat{C}_i$

References	Honest	Corrupted (if no corruption in observed utilities)
Chen and Peng (2020)	$\sqrt{n}(m \log m)^{3/4} T^{1/4}$	$\sqrt{mT \log m} + \widehat{C}_i$
Anagnostides et al. (2022)	$nm^{5/2} \log T$	$nm^{5/2} \log T + \sqrt{mT \log m} + \widehat{C}_i$

Main result (3): Swap regret bound in the corrupted regime

Swap regret upper bounds of player i in multi-player general-sum games with n-players and m-actions after T rounds

$\widehat{C}_i \in [0, 2T]$: the cumulative amount of corruption in strategies for player i, $\widehat{S} = \sum_{i \in [n]} \widehat{C}_i$

References	Honest	Corrupted (if no corruption in observed utilities)
Chen and Peng (2020)	$\sqrt{n}(m \log m)^{3/4} T^{1/4}$	$\sqrt{mT \log m} + \widehat{C}_i$
Anagnostides et al. (2022)	$nm^{5/2} \log T$	$nm^{5/2} \log T + \sqrt{mT \log m} + \widehat{C}_i$
Ours	$nm^{5/2} \log T$	$nm^{5/2} \log T + \min\left\{\sqrt{\widehat{S}(nm^2 + m^{5/2}) \log T}, m\sqrt{T \log T}\right\} + \widehat{C}_i$

Main result (3): Swap regret bound in the corrupted regime

Swap regret upper bounds of player i in multi-player general-sum games with n-players and m-actions after T rounds

$\widehat{C}_i \in [0, 2T]$: the cumulative amount of corruption in strategies for player i, $\widehat{S} = \sum_{i \in [n]} \widehat{C}_i$

References	Honest	Corrupted (if no corruption in observed utilities)
Chen and Peng (2020)	$\sqrt{n}(m \log m)^{3/4} T^{1/4}$	$\sqrt{mT \log m} + \widehat{C}_i$
Anagnostides et al. (2022)	$nm^{5/2} \log T$	$nm^{5/2} \log T + \sqrt{mT \log m} + \widehat{C}_i$
Ours	$nm^{5/2} \log T$	$nm^{5/2} \log T + \min\left\{\sqrt{\widehat{S}(nm^2 + m^{5/2}) \log T}, m\sqrt{T \log T}\right\} + \widehat{C}_i$

Compared to the best bounds by Anagnostides et al. (2022), our algorithm achieves

- the same bound in the honest regime
- a new adaptive bound in the corrupted regime in terms of $\widehat{S}$ and $\widetilde{S}$
- a worst-case bound that is $\sqrt{m}$-times worse than their bound of $O(nm^{5/2} \log T + \sqrt{Tm \log m})$.

Main result (3): Swap regret bound in the corrupted regime

Swap regret upper bounds of player i in multi-player general-sum games with n-players and m-actions after T rounds

$\widehat{C}_i \in [0, 2T]$: the cumulative amount of corruption in strategies for player i, $\widehat{S} = \sum_{i\in[n]} \widehat{C}_i$

References	Honest	Corrupted (if no corruption in observed utilities)
Chen and Peng (2020)	$\sqrt{n}(m \log m)^{3/4} T^{1/4}$	$\sqrt{mT \log m} + \widehat{C}_i$
Anagnostides et al. (2022)	$nm^{5/2} \log T$	$nm^{5/2} \log T + \sqrt{mT \log m} + \widehat{C}_i$
Ours	$nm^{5/2} \log T$	$nm^{5/2} \log T + \min\left\{\sqrt{\widehat{S}(nm^2 + m^{5/2}) \log T}, m\sqrt{T \log T}\right\} + \widehat{C}_i$

Compared to the best bounds by Anagnostides et al. (2022), our algorithm achieves

- the same bound in the honest regime
- a new adaptive bound in the corrupted regime in terms of $\widehat{S}$ and $\widetilde{S}$
- a worst-case bound that is $\sqrt{m}$-times worse than their bound of $O(nm^{5/2} \log T + \sqrt{Tm \log m})$.

Again, the bound in corrupted regime incentivizes players to follow the given algorithm.

Key lemma: Stability of stationary distributions

Lemma (Stability of stationary distributions of Markov chains)

Recall that OFTRL update is given by

$$y_{i,a}^{(t)} = \arg\max_{y \in \Delta_m} \left\{ -F_{i,a}^{(t)}(y) \right\}, \quad F_{i,a}^{(t)}(y) := -\left(\left\langle y, \widetilde{u}_{i,a}^{(t-1)} + \sum_{s=1}^{t-1} \widetilde{u}_{i,a}^{(s)} \right\rangle - \frac{1}{\eta_{i,a}^{(t)}} \phi(y) \right).$$

Then, the choice of the learning rate $\eta_{i,a}^{(t)}$ *guarantees*

$$\sum_{a \in \mathcal{A}} \| y_a^{(t+1)} - y_a^{(t)} \|_{y_a^{(t)}, F_a^{(t+1)}} = \frac{1}{\sqrt{\eta_a^{(t+1)}}} \sum_{a \in \mathcal{A}} \| y_a^{(t+1)} - y_a^{(t)} \|_{y_a^{(t)}, \phi} \leq \frac{1}{2}.$$

Recall that we define a Markov chain from $y_{i,a}^{(t)} \in \Delta_{m_i}$ for each $a \in \mathcal{A}_i$ ($Q_i^{(t)} \in [0,1]^{m_i \times m_i}$ whose a-th row is $y_{i,a}^{(t)}$ for each $a \in \mathcal{A}_i$) and use its stationary distribution as $y_{i,a}^{(t)}$.

Summary

Research questions

- Can we adapt to deviations of the opponent from a given algorithm?
- Can we characterize regret and convergence rates to an equilibrium in such a corrupted game?

Our contributions

- Establish a framework of corrupted games, in which each player may deviate from a prescribed algorithm
- Give a nearly complete characterization of learning dynamics in corrupted games, by deriving regret upper and lower bounds in (normal-form) two-player zero-sum and multi-player general-sum games:

$$\text{Roughly,} \quad \mathrm{Reg}_{x,g}^{T} = \widetilde{\Theta}(\sqrt{C_y} + C_x)\,, \quad \mathrm{SwapReg}_{x_i,u_i}^{T} = \widetilde{\Theta}\Big(\sqrt{\textstyle\sum_{j \neq i} C_j} + C_i\Big)\,.$$

References I

- Anagnostides, Ioannis et al. (2022). “Uncoupled Learning Dynamics with $O(\log T)$ Swap Regret in Multiplayer Games”. In: *Advances in Neural Information Processing Systems*. Vol. 35. Curran Associates, Inc., pp. 3292–3304.
- Aumann, Robert J. (1974). “Subjectivity and correlation in randomized strategies”. In: *Journal of Mathematical Economics* 1.1, pp. 67–96.
- Blum, Avrim and Yishay Mansour (2007). “From External to Internal Regret”. In: *Journal of Machine Learning Research* 8.47, pp. 1307–1324.
- Chen, Xi and Binghui Peng (2020). “Hedging in games: Faster convergence of external and swap regrets”. In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., pp. 18990–18999.
- Foster, Dean P. and Rakesh V. Vohra (1997). “Calibrated Learning and Correlated Equilibrium”. In: *Games and Economic Behavior* 21.1, pp. 40–55.
- Freund, Yoav and Robert E. Schapire (1999). “Adaptive Game Playing Using Multiplicative Weights”. In: *Games and Economic Behavior* 29.1, pp. 79–103.

References II

- Rakhlin, Alexander and Karthik Sridharan (2013). "Online Learning with Predictable Sequences". In: *Proceedings of the 26th Annual Conference on Learning Theory*. Vol. 30, pp. 993–1019.
- 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.