Corrupted Learning Dynamics in Games

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

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

July 2, 2025

38th Conference on Learning Theory (COLT 2025), Lyon

Learning in two-player zero-sum normal-form 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 $t = 1, \ldots, T$: ($\Delta_m = \{x \in [0, 1]^m : \|x\|_1 = 1\}$: the $(m-1)$-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 normal-form 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 $t = 1, \ldots, T$: ($\Delta_m = \{x \in [0, 1]^m : \|x\|_1 = 1\}$: the $(m-1)$-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 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 normal-form 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 $t = 1, \dots, T$: ($\Delta_m = \{x \in [0,1]^m : \|x\|_1 = 1\}$: the $(m-1)$-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 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 normal-form 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 $t = 1, \ldots, T$: ($\Delta_m = \{x \in [0, 1]^m : \|x\|_1 = 1\}$: the $(m-1)$-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 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

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

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.*

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

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

Q. Is this optimal rate in learning in games?

Fast convergence in games

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

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

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.*

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 optimistic Hedge with constant lr may lead to a linear regret

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.*

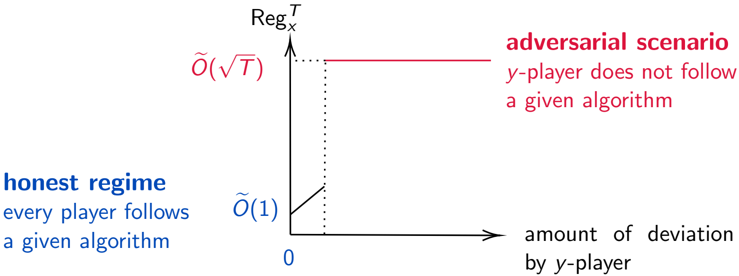
Q. What if the opponent does not follow optimistic Hedge with a constant learning rate?

Continuing with optimistic Hedge with constant lr may lead to a linear regret

$\rightarrow$ 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 $\widetilde{O}(\sqrt{T})$

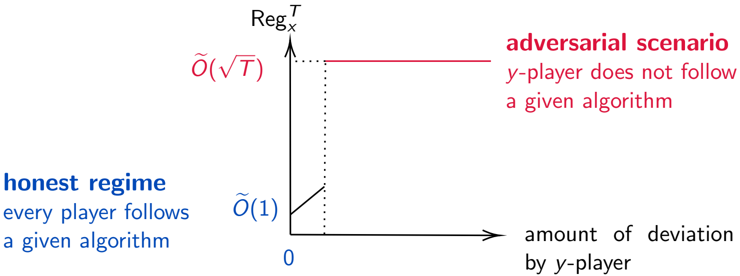
Research questions

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
- Derive regret upper and lower bounds in **two-player zero-sum and multiplayer general-sum games**

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 $t = 1, \ldots, T$:

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 $\ell^{(t)} = A^\top x^{(t)}$;

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 $\langle y^{(t)}, \widetilde{\ell}^{(t)} \rangle$

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 $t = 1, \dots, T$:

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 $\langle y^{(t)}, \widetilde{\ell}^{(t)} \rangle$

- corrupted regime with no corruptions = honest regime
- corrupted regime with arbitrary $\widetilde{C}_y$ = adversarial scenario for x-player

Our algorithm: Optimistic Hedge with adaptive learning rate

Syrgkanis et al. (2015): Optimistic Hedge with constant learning rate
(fast rates in 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 (fast rates in 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

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

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 $\widehat{C}_x$.

Main result (2): Lower bounds

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

$$\mathrm{Reg}^T_{x,\widetilde{g}} := \max_{x^* \in \Delta_{m_x}} \left\{ \sum_{t=1}^T \langle x^*, \widetilde{g}^{(t)} \rangle - \sum_{t=1}^T \langle x^{(t)}, \widetilde{g}^{(t)} \rangle \right\} = O(\sqrt{\widetilde{C}_x \log m_x}),$$

Main result (2): Lower bounds

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

$$\mathrm{Reg}^T_{x,\widetilde{g}} := \max_{x^* \in \Delta_{m_x}} \left\{ \sum_{t=1}^T \langle x^*, \widetilde{g}^{(t)} \rangle - \sum_{t=1}^T \langle x^{(t)}, \widetilde{g}^{(t)} \rangle \right\} = O(\sqrt{\widetilde{C}_x \log m_x}) \, ,$$

Main result (2): Lower bounds

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

$$\mathrm{Reg}^T_{x,\widetilde{g}} := \max_{x^* \in \Delta_{m_x}} \left\{ \sum_{t=1}^T \langle x^*, \widetilde{g}^{(t)} \rangle - \sum_{t=1}^T \langle x^{(t)}, \widetilde{g}^{(t)} \rangle \right\} = O\left(\sqrt{\widetilde{C}_x \log m_x}\right),$$

Theorem: For any learning dynamics, there exists a corrupted game with

$$\sum_{t=1}^T \|g^{(t)} - \widetilde{g}^{(t)}\|_\infty \le \widetilde{C}_x \text{ such that } \mathrm{Reg}^T_{x,\widetilde{g}} = \Omega\left(\sqrt{\widetilde{C}_x \log m_x}\right).$$

Main result (2): Lower bounds

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

$$\mathrm{Reg}^T_{x,\widetilde{g}} := \max_{x^* \in \Delta_{m_x}} \left\{ \sum_{t=1}^T \langle x^*, \widetilde{g}^{(t)} \rangle - \sum_{t=1}^T \langle x^{(t)}, \widetilde{g}^{(t)} \rangle \right\} = O(\sqrt{\widetilde{C}_x \log m_x}),$$

   Theorem: For any learning dynamics, there exists a corrupted game with

$$\sum_{t=1}^T \|g^{(t)} - \widetilde{g}^{(t)}\|_\infty \le \widetilde{C}_x \text{ such that } \mathrm{Reg}^T_{x,\widetilde{g}} = \Omega\left(\sqrt{\widetilde{C}_x \log m_x}\right).$$

2. player's own strategy deviation:
   If corruption occurs only in x-player's strategies, $\mathrm{Reg}^T_{x,g} = \widetilde{O}(\widehat{C}_x)$.

Main result (2): Lower bounds

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

$$\mathrm{Reg}^T_{x,\widetilde{g}} := \max_{x^* \in \Delta_{m_x}} \left\{ \sum_{t=1}^T \langle x^*, \widetilde{g}^{(t)} \rangle - \sum_{t=1}^T \langle x^{(t)}, \widetilde{g}^{(t)} \rangle \right\} = O\left(\sqrt{\widetilde{C}_x \log m_x}\right),$$

   Theorem: For any learning dynamics, there exists a corrupted game with

$$\sum_{t=1}^T \|g^{(t)} - \widetilde{g}^{(t)}\|_\infty \leq \widetilde{C}_x \text{ such that } \mathrm{Reg}^T_{x,\widetilde{g}} = \Omega\left(\sqrt{\widetilde{C}_x \log m_x}\right).$$

2. player's own strategy deviation:

   If corruption occurs only in x-player's strategies, $\mathrm{Reg}^T_{x,g} = \widetilde{O}(\widehat{C}_x)$.

Main result (2): Lower bounds

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

$$\mathrm{Reg}_{x,\widetilde{g}}^T := \max_{x^* \in \Delta_{m_x}} \left\{ \sum_{t=1}^T \langle x^*, \widetilde{g}^{(t)} \rangle - \sum_{t=1}^T \langle x^{(t)}, \widetilde{g}^{(t)} \rangle \right\} = O\left(\sqrt{\widetilde{C}_x \log m_x}\right),$$

Theorem: For any learning dynamics, there exists a corrupted game with

$$\sum_{t=1}^T \| g^{(t)} - \widetilde{g}^{(t)} \|_\infty \le \widetilde{C}_x \text{ such that } \mathrm{Reg}_{x,\widetilde{g}}^T = \Omega\left(\sqrt{\widetilde{C}_x \log m_x}\right).$$

2. player's own strategy deviation:

If corruption occurs only in x-player's strategies, $\mathrm{Reg}_{x,g}^T = \widetilde{O}(\widehat{C}_x)$.

Theorem: $\forall$ dynamics, $\exists$ game w/ $\sum_{t=1}^T \| x^{(t)} - \widehat{x}^{(t)} \|_1 \le \widehat{C}_x$ such that $\mathrm{Reg}_{x,g}^T = \Omega(\widehat{C}_x)$.

Main result (2): Lower bounds

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

$$\mathrm{Reg}^T_{x,\widetilde{g}} := \max_{x^* \in \Delta_{m_x}} \left\{ \sum_{t=1}^T \langle x^*, \widetilde{g}^{(t)} \rangle - \sum_{t=1}^T \langle x^{(t)}, \widetilde{g}^{(t)} \rangle \right\} = O\left(\sqrt{\widetilde{C}_x \log m_x}\right),$$

   Theorem: For any learning dynamics, there exists a corrupted game with $\sum_{t=1}^T \|g^{(t)} - \widetilde{g}^{(t)}\|_\infty \leq \widetilde{C}_x$ such that $\mathrm{Reg}^T_{x,\widetilde{g}} = \Omega\left(\sqrt{\widetilde{C}_x \log m_x}\right)$.

2. player's own strategy deviation:

   If corruption occurs only in x-player's strategies, $\mathrm{Reg}^T_{x,g} = \widetilde{O}(\widehat{C}_x)$.

   Theorem: $\forall$ dynamics, $\exists$ game w/ $\sum_{t=1}^T \|x^{(t)} - \widehat{x}^{(t)}\|_1 \leq \widehat{C}_x$ such that $\mathrm{Reg}^T_{x,g} = \Omega(\widehat{C}_x)$.

3. opponent's strategy deviation:

   If corruption occurs only in y-player's strategies, $\mathrm{Reg}^T_{\widehat{x},g} = \widetilde{O}\left(\sqrt{\widehat{C}_y}\right)$, $\mathrm{Reg}^T_{\widehat{y},\ell} = \widetilde{O}\left(\sqrt{\widehat{C}_y}\right)$.

Main result (2): Lower bounds

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

$$\mathrm{Reg}^T_{x,\widetilde{g}} := \max_{x^* \in \Delta_{m_x}} \left\{ \sum_{t=1}^T \langle x^*, \widetilde{g}^{(t)} \rangle - \sum_{t=1}^T \langle x^{(t)}, \widetilde{g}^{(t)} \rangle \right\} = O(\sqrt{\widetilde{C}_x \log m_x}),$$

   Theorem: For any learning dynamics, there exists a corrupted game with
   $\sum_{t=1}^T \|g^{(t)} - \widetilde{g}^{(t)}\|_\infty \le \widetilde{C}_x$ such that $\mathrm{Reg}^T_{x,\widetilde{g}} = \Omega\left(\sqrt{\widetilde{C}_x \log m_x}\right)$.

2. player's own strategy deviation:
   If corruption occurs only in x-player's strategies, $\mathrm{Reg}^T_{x,g} = \widetilde{O}(\widehat{C}_x)$.
   Theorem: $\forall$ dynamics, $\exists$ game w/ $\sum_{t=1}^T \|x^{(t)} - \widehat{x}^{(t)}\|_1 \le \widehat{C}_x$ such that $\mathrm{Reg}^T_{x,g} = \Omega(\widehat{C}_x)$.

3. opponent's strategy deviation:
   If corruption occurs only in y-player's strategies, $\mathrm{Reg}^T_{\widehat{x},g} = \widetilde{O}\left(\sqrt{\widehat{C}_y}\right)$, $\mathrm{Reg}^T_{\widehat{y},\ell} = \widetilde{O}(\sqrt{\widehat{C}_y})$.

Main result (2): Lower bounds

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

$$\mathrm{Reg}^T_{x,\widetilde{g}} := \max_{x^* \in \Delta_{m_x}} \left\{ \sum_{t=1}^T \langle x^*, \widetilde{g}^{(t)} \rangle - \sum_{t=1}^T \langle x^{(t)}, \widetilde{g}^{(t)} \rangle \right\} = O(\sqrt{\widetilde{C}_x \log m_x}),$$

   Theorem: For any learning dynamics, there exists a corrupted game with

$$\sum_{t=1}^T \|g^{(t)} - \widetilde{g}^{(t)}\|_\infty \le \widetilde{C}_x \text{ such that } \mathrm{Reg}^T_{x,\widetilde{g}} = \Omega\left(\sqrt{\widetilde{C}_x \log m_x}\right).$$

2. player's own strategy deviation:

   If corruption occurs only in x-player's strategies, $\mathrm{Reg}^T_{x,g} = \widetilde{O}(\widehat{C}_x)$.

   Theorem: $\forall$ dynamics, $\exists$ game w/ $\sum_{t=1}^T \|x^{(t)} - \widehat{x}^{(t)}\|_1 \le \widehat{C}_x$ such that $\mathrm{Reg}^T_{x,g} = \Omega(\widehat{C}_x)$.

3. opponent's strategy deviation:

   If corruption occurs only in y-player's strategies, $\mathrm{Reg}^T_{\widehat{x},g} = \widetilde{O}\left(\sqrt{\widehat{C}_y}\right)$, $\mathrm{Reg}^T_{\widehat{y},\ell} = \widetilde{O}(\sqrt{\widehat{C}_y})$.

   Theorem: $\forall$ dynamics, $\exists$ game w/ $\sum_{t=1}^T \|y^{(t)} - \widehat{y}^{(t)}\|_1 \le \widehat{C}_y$ such that

$$\max\left\{\mathrm{Reg}^T_{\widehat{x},g}, \mathrm{Reg}^T_{\widehat{y},\ell}\right\} = \Omega\left(\sqrt{\widehat{C}_y}\right).$$

Main result (3): Extension to corrupted multiplayer general-sum games

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

Main result (3): Extension to corrupted multiplayer general-sum games

Swap regret upper bounds of player i in multiplayer 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): Extension to corrupted multiplayer general-sum games

Swap regret upper bounds of player i in multiplayer 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	$nm^{5/2} \log T$	$nm^{5/2} \log T + \min\left\{\sqrt{\widehat{S}(nm^2 + m^{5/2}) \log T}, m\sqrt{T \log T}\right\} + \widehat{C}_i$

Key techniques: stability of stationary distributions of Markov chains determined by optimistic FTRL with adaptive learning rate

Summary

Our contributions

- Established a framework of corrupted games, in which each player may deviate from a prescribed algorithm
- Derived regret upper and lower bounds in two-player zero-sum and multiplayer general-sum games:

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

Many directions for future work

- extensive-form games, Markov games, ...
- another regret measure such as Φ-regret
- last-iterate convergence

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.
- 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.
- Freund, Yoav and Robert E. Schapire (1999). "Adaptive Game Playing Using Multiplicative Weights". In: *Games and Economic Behavior* 29.1, pp. 79–103.
- 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.

References II

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.