No-Regret Learning in Games with Noisy Feedback: Faster Rates and Adaptivity via Learning Rate Separation

Yu-Guan Hsieh, Kimon Antonakopoulos, Volkan Cevher, Panayotis Mertikopoulos

Université Grenoble Alpes, EPFL

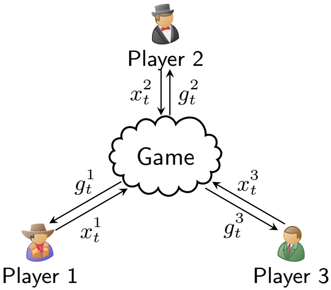
NeurIPS, December 2022

Online Learning in Continuous Games

At each round $t = 1, 2, \ldots$, each player $i \in \mathcal{N}$

- Plays an action $x_t^i \in \mathcal{X}^i$
- Suffers loss $\ell^i(\mathbf{x}_t)$ and receives (first order) feedback $g_t^i \approx \nabla_i \ell^i(\mathbf{x}_t)$

- Each player i has a convex closed action set $\mathcal{X}^i$ and a loss function $\ell^i : \mathcal{X}^1 \times \ldots \times \mathcal{X}^N \to \mathbb{R}$
- Joint action of all players $\mathbf{x} = (x^i)_{i \in \mathcal{N}} = (x^i, \mathbf{x}^{-i})$
- $\ell^i(\cdot, \mathbf{x}^{-i})$ is convex and $\nabla_i \ell^i(\mathbf{x}_t)$ is Lipschitz continuous

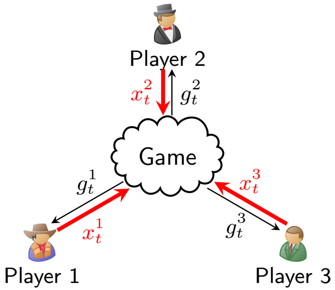


Online Learning in Continuous Games

At each round $t = 1, 2, \ldots$, each player $i \in \mathcal{N}$

- Plays an action $x_t^i \in \mathcal{X}^i$
- Suffers loss $\ell^i(\mathbf{x}_t)$ and receives (first order) feedback $g_t^i \approx \nabla_i \ell^i(\mathbf{x}_t)$

- Each player i has a convex closed action set $\mathcal{X}^i$ and a loss function $\ell^i : \mathcal{X}^1 \times \ldots \times \mathcal{X}^N \to \mathbb{R}$
- Joint action of all players $\mathbf{x} = (x^i)_{i \in \mathcal{N}} = (x^i, \mathbf{x}^{-i})$
- $\ell^i(\cdot, \mathbf{x}^{-i})$ is convex and $\nabla_i \ell^i(\mathbf{x}_t)$ is Lipschitz continuous

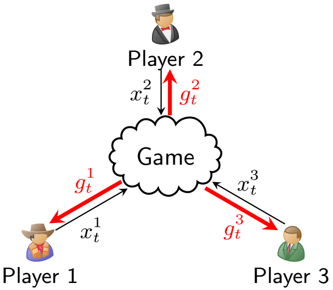


Online Learning in Continuous Games

At each round $t = 1, 2, \dots$, each player $i \in \mathcal{N}$

- Plays an action $x_t^i \in \mathcal{X}^i$
- Suffers loss $\ell^i(\mathbf{x}_t)$ and receives (first order) feedback $g_t^i \approx \nabla_i \ell^i(\mathbf{x}_t)$

- Each player i has a convex closed action set $\mathcal{X}^i$ and a loss function $\ell^i : \mathcal{X}^1 \times \ldots \times \mathcal{X}^N \to \mathbb{R}$
- Joint action of all players $\mathbf{x} = (x^i)_{i \in \mathcal{N}} = (x^i, \mathbf{x}^{-i})$
- $\ell^i(\cdot, \mathbf{x}^{-i})$ is convex and $\nabla_i \ell^i(\mathbf{x}_t)$ is Lipschitz continuous

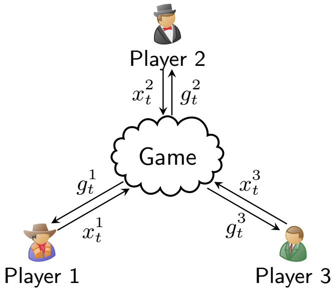


Online Learning in Continuous Games

At each round $t = 1, 2, \ldots$, each player $i \in \mathcal{N}$

- Plays an action $x_t^i \in \mathcal{X}^i$
- Suffers loss $\ell^i(\mathbf{x}_t)$ and receives (first order) feedback $g_t^i \approx \nabla_i \ell^i(\mathbf{x}_t)$

- Each player i has a convex closed action set $\mathcal{X}^i$ and a loss function $\ell^i : \mathcal{X}^1 \times \ldots \times \mathcal{X}^N \to \mathbb{R}$
- Joint action of all players $\mathbf{x} = (x^i)_{i \in \mathcal{N}} = (x^i, \mathbf{x}^{-i})$
- $\ell^i(\cdot, \mathbf{x}^{-i})$ is convex and $\nabla_i \ell^i(\mathbf{x}_t)$ is Lipschitz continuous
- Players can be adversarial or optimizing their own benefit



Regret and Oracle

- Individual regret of player i:

$$\mathrm{Reg}_T^i(\mathcal{P}^i) = \max_{p^i \in \mathcal{P}^i} \sum_{t=1}^{T} \Big(\underbrace{\ell^i(x_t^i, \mathbf{x}_t^{-i}) - \ell^i(p^i, \mathbf{x}_t^{-i})}_{\text{cost of not playing } p^i \text{ in round } t} \Big).$$

No regret if $\mathrm{Reg}_T^i(\mathcal{P}^i) = o(T)$

Nearly constant regret is possible under perfect feedback if all players play some prescribed algorithm

Regret and Oracle

- Individual regret of player i:

$$\mathrm{Reg}^i_T(\mathcal{P}^i) = \max_{p^i \in \mathcal{P}^i} \sum_{t=1}^{T} \Big(\underbrace{\ell^i(x^i_t, \mathbf{x}^{-i}_t) - \ell^i(p^i, \mathbf{x}^{-i}_t)}_{\text{cost of not playing } p^i \text{ in round } t} \Big).$$

  No regret if $\mathrm{Reg}^i_T(\mathcal{P}^i) = o(T)$

  Nearly constant regret is possible under perfect feedback if all players play some prescribed algorithm

- Stochastic oracle $\mathbb{E}[g^i_t] = \nabla_i \ell^i(\mathbf{x}_t)$
  - Additive noise: $g^i_t = \nabla_i \ell^i(\mathbf{x}_t) + \xi^i_t$
  - Multiplicative noise: $g^i_t = \nabla_i \ell^i(\mathbf{x}_t)(1 + \xi^i_t)$

  where ξ^i_t is zero-mean and has finite variance

Constant Regret under Noisy Feedback

Question: Existing algorithms achieve constant regret in a broad family of games under perfect feedback. Can we achieve the same with noisy feedback?

Constant Regret under Noisy Feedback

Question: Existing algorithms achieve constant regret in a broad family of games under perfect feedback. Can we achieve the same with noisy feedback?

Answer: Yes if the noise is multiplicative. Just run optimistic algorithms with scale separation!
E.g., OG+ [$\mathbf{x}_t = \mathbf{X}_{t+\frac{1}{2}}$, $\eta_t \leq \gamma_t$]

$$\mathbf{X}_{t+\frac{1}{2}} = \mathbf{X}_t - \gamma_t \hat{\mathbf{V}}_{t-\frac{1}{2}}, \qquad \mathbf{X}_{t+1} = \mathbf{X}_t - \eta_t \hat{\mathbf{V}}_{t+\frac{1}{2}}$$

Additional assumption: variational stability of the game (include monotone games and especially zero-sum polymatrix games)

Constant Regret under Noisy Feedback

Question: Existing algorithms achieve constant regret in a broad family of games under perfect feedback. Can we achieve the same with noisy feedback?

- Bonus 1: Adaptivity to bypass the need for knowing constants and to cope with adversarial opponents

Constant Regret under Noisy Feedback

Question: Existing algorithms achieve constant regret in a broad family of games under perfect feedback. Can we achieve the same with noisy feedback?

- Bonus 1: Adaptivity to bypass the need for knowing constants and to cope with adversarial opponents
- Bonus 2: Last-iterate convergence to Nash Equilibrium

Illustrating Example: Failure of Existing Algorithm

- Draw $\mathcal{L}_1(\mathbf{x}) = 3\theta\phi$ or $\mathcal{L}_2(\mathbf{x}) = -\theta\phi$ with equal probability so

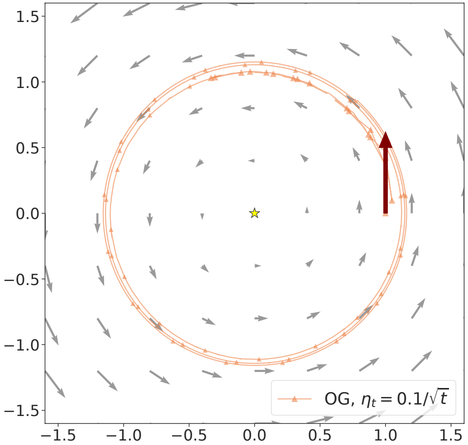
$$\ell^1 = -\ell^2 = (\mathcal{L}_1 + \mathcal{L}_2)/2 = \theta\phi$$

- Stochastic estimate $\mathbb{E}[\hat{\mathbf{V}}_{t+\frac{1}{2}}] = \mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})$

$$\hat{\mathbf{V}}_{t+\frac{1}{2}} = \begin{cases} (3\phi_{t+\frac{1}{2}}, -3\theta_{t+\frac{1}{2}}) & \text{with prob. } 1/2 \\ (-\phi_{t+\frac{1}{2}}, \theta_{t+\frac{1}{2}}) & \text{with prob. } 1/2 \end{cases}$$

- Optimistic gradient $[\mathbf{x}_t = \mathbf{X}_{t+\frac{1}{2}}]$

$$\mathbf{X}_{t+\frac{1}{2}} = \mathbf{X}_t - \eta_t \hat{\mathbf{V}}_{t-\frac{1}{2}}, \quad \mathbf{X}_{t+1} = \mathbf{X}_t - \eta_{t+1} \hat{\mathbf{V}}_{t+\frac{1}{2}}$$



Illustrating Example: Scale Separation

- Draw $\mathcal{L}_1(\mathbf{x}) = 3\theta\phi$ or $\mathcal{L}_2(\mathbf{x}) = -\theta\phi$ with equal probability so

$$\ell^1 = -\ell^2 = (\mathcal{L}_1 + \mathcal{L}_2)/2$$

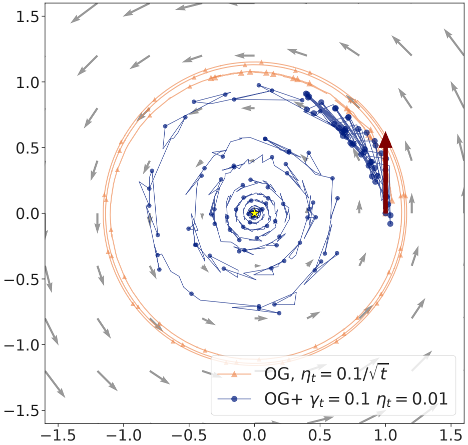
- OG+ $[\mathbf{x}_t = \mathbf{X}_{t+\frac{1}{2}}]$

$$\mathbf{X}_{t+\frac{1}{2}} = \mathbf{X}_t - \gamma_t \hat{\mathbf{V}}_{t-\frac{1}{2}}$$

$$\mathbf{X}_{t+1} = \mathbf{X}_t - \eta_t \hat{\mathbf{V}}_{t+\frac{1}{2}}$$

With $\gamma_t \geq \eta_t$

- This makes the noise an order smaller than the negative shift in the analysis



Illustrating Example: Adaptivity

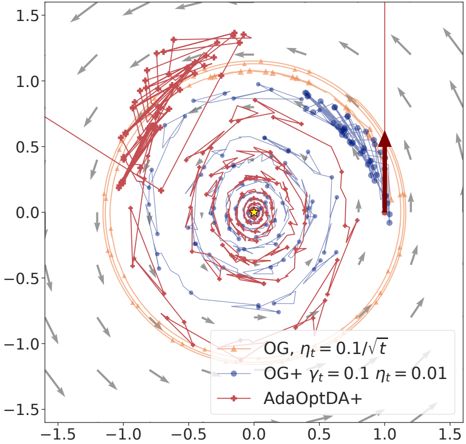
- OptDA+ $[\gamma_t^i \geq \eta_t^i]$

$$X_{t+\frac{1}{2}}^i = X_t^i - \gamma_t^i g_{t-1}^i \quad X_{t+1}^i = X_1^i - \eta_{t+1}^i \sum_{s=1}^{t} g_s^i$$

- AdaOptDA+ uses learning rate

$$\gamma_t^i = \frac{1}{\left(1 + \sum_{s=1}^{t-2} \|g_s^i\|^2\right)^{\frac{1}{4}}}$$

$$\eta_t^i = \frac{1}{\sqrt{1 + \sum_{s=1}^{t-2} \left(\|g_s^i\|^2 + \|X_s^i - X_{s+1}^i\|^2\right)}}$$



Summary of Results

	Adversarial	All players run the same algorithm			
	Bounded feedback	Additive noise		Multiplicative noise	
	Regret	Regret	Convergence	Regret	Convergence
OG	✗	✗	✗	✗	✗
OG+	✗	$\sqrt{t}\log t$	✓	cst	✓
OptDA+	$\sqrt{t}$	$\sqrt{t}$	–	cst	✓
AdaOptDA+	$t^{3/4}$	$\sqrt{t}$	–	cst	✓