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

Yu-Guan Hsieh[1] Kimon Antonakopoulos[2] Volkan Cevher[2] Panayotis Mertikopoulos[1,3,4] ([1]UGA, Inria [2]EPFL [3]CNRS [4]Criteo AI Lab)

Online Learning in Continuous Games

At each round $t = 1, 2, \dots$, each player $i \in \mathcal{N} := \{1, \dots, N\}$
- Plays an action $x_t^i \in \mathcal{X}^i$ (closed convex)
- Suffers loss $\ell^i(\mathbf{x}_t)$ and receives estimate g_t^i of $\nabla_i \ell^i(\mathbf{x}_t)$

- $\ell^i(\cdot, \mathbf{x}^{-i})$ is convex and $\nabla_i \ell^i(\mathbf{x}_t)$ is Lipschitz continuous
- Nash equilibrium $\mathbf{x}_\star$: $\forall i \in \mathcal{N}, \forall x^i \in \mathcal{X}^i, \ell^i(x_\star^i, \mathbf{x}_\star^{-i}) \le \ell^i(x^i, \mathbf{x}_\star^{-i})$
- Individual regret of agent i:

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

- Opponents can be adversarial or optimizing their own objectives

The Callenge: Noisy Feedback

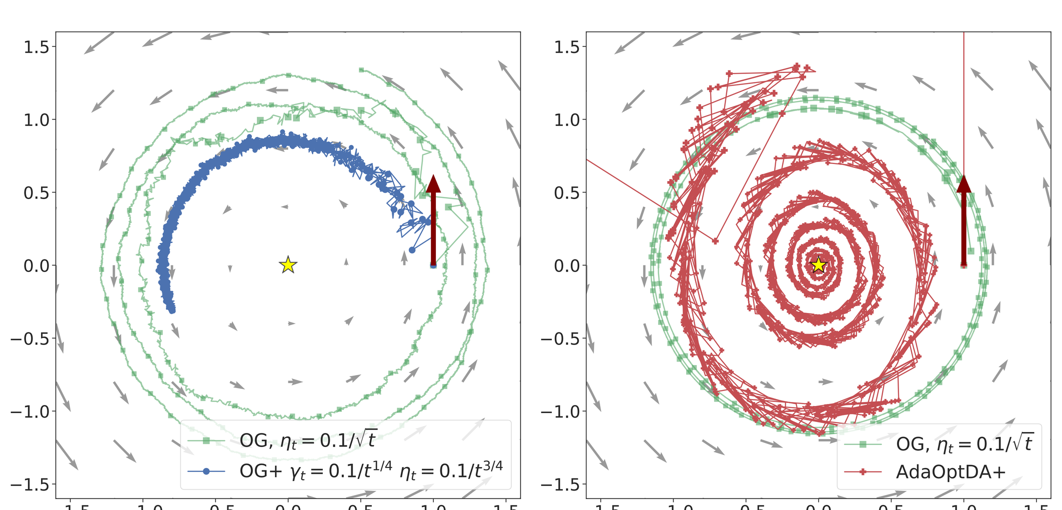
We consider
- **Additive noise**: $g_t^i = \nabla_i \ell^i(\mathbf{x}_t) + \xi_t^i$
- **Multiplicative noise**: $g_t^i = \nabla_i \ell^i(\mathbf{x}_t)(1 + \xi_t^i)$

Example. unconstrained two-player zero-sum bilinear games

$$\ell^1(\mathbf{x}) = -\ell^2(\mathbf{x}) = x^1 x^2; \quad \mathcal{X}^1 = \mathcal{X}^2 = \mathbb{R}; \quad x_\star = (0, 0)$$

Left: Additive Gaussian noise $\xi_t^1, \xi_t^2 \sim \mathcal{N}(0, I)$

Right: Multiplicative noise (ξ_t^1, ξ_t^2) is $(2, -2)$ or $(-2, 2)$ with prob $1/2$ for each



TL;DR

We show that **optimistic** gradient methods with **learning rate separation** achieve **constant regret** and **last-iterate convergence** in variationally stable games under **multiplicative noise**, and devise **adaptive** methods that achieve this automatically.

Optimistic Methods with Learning Rate Separation

- Optimistic gradient: $x_{t+1}^i = x_t^i - 2\eta_{t+1}^i g_t^i + \eta_t^i g_{t-1}^i$
- Rewrite with $X_{t+\frac{1}{2}}^i = x_t^i$ and separate the optimistic learning rate from the update learning rate

$$X_{t+\frac{1}{2}}^i = X_t^i - \gamma_t^i\, g_{t-1}^i, \qquad X_{t+1}^i = X_t^i - \eta_{t+1}^i\, g_t^i \qquad \text{(OG+)}$$

$$X_{t+\frac{1}{2}}^i = X_t^i - \gamma_t^i\, g_{t-1}^i, \qquad X_{t+1}^i = X_1^i - \eta_{t+1}^i \sum_{s=1}^{t} g_s^i \qquad \text{(OptDA+)}$$

Energy inequality

If all players play run OG+ or OptDA+, for any $p^i \in \mathcal{X}^i$, it holds

$$\begin{aligned}
\mathbb{E}_{t-1}\left[\frac{\|X_{t+1}^i - p^i\|^2}{\eta_{t+1}^i}\right] \le\ & \mathbb{E}_{t-1}\Bigg[\frac{\|X_t^i - p^i\|^2}{\eta_t^i} + \left(\frac{1}{\eta_{t+1}^i} - \frac{1}{\eta_t^i}\right)\|u_t^i - p^i\|^2 \\
\text{(linearized regret)}\quad & - 2\langle V^i(\mathbf{X}_{t+\frac{1}{2}}), X_{t+\frac{1}{2}}^i - p^i\rangle \\
\text{(negative drift)}\quad & - \gamma_t^i\,(\|V^i(\mathbf{X}_{t+\frac{1}{2}})\|^2 + \|V^i(\mathbf{X}_{t-\frac{1}{2}})\|^2) \\
\text{(use smoothness)}\quad & - \frac{\|X_t^i - X_{t+1}^i\|^2}{2\eta_t^i} + \gamma_t^i \|V^i(\mathbf{X}_{t+\frac{1}{2}}) - V^i(\mathbf{X}_{t-\frac{1}{2}})\|^2 \\
\text{(noise)}\quad & + (\gamma_t^i)^2 L \|\xi_{t-\frac{1}{2}}^i\|^2 + L\|\xi_{t-\frac{1}{2}}\|^2_{(\eta_t+\gamma_t)^2} + 2\,\eta_t^i\,\|g_t^i\|^2\Bigg]
\end{aligned}$$

- $V^i = \nabla_i \ell^i$ and $\|\xi_{t-\frac{1}{2}}\|^2_{(\eta_t+\gamma_t)^2} := \sum_{j=1}^N (\eta_t^j + \gamma_t^j)^2 \|\xi_{t-\frac{1}{2}}^j\|^2$
- $u_t^i = X_t^i$ if player i runs OG+ and $u_t^i = X_1^i$ if player i runs OptDA+

Results

	Adversarial	All players run the same algorithm			
		Additive noise		Multiplicative noise	
	Regret	Regret	Convergence	Regret	Convergence
OG	✗	✗	✗	✗	✗
OG+	✗	$\sqrt{t}\log t$	✓	cst	✓
OptDA+	$\sqrt{t}$	$\sqrt{t}$	–	cst	✓
Adapt	$t^{1/2+q}$	$\sqrt{t}$	–	cst	✓

Assumptions
- Unconstrained action sets
- For the adversarial setup we assume bounded feedback
- For the game-theoretic setup (i.e., when all players play the same algorithm) we assume variational stability, that is, the set $\mathcal{X}_\star$ of Nash equilibria of the game is nonempty and

$$\langle \mathbf{V}(\mathbf{x}), \mathbf{x} - \mathbf{x}_\star\rangle := \sum_{i=1}^N \langle \nabla_i \ell^i(\mathbf{x}), x^i - x_\star^i\rangle \ge 0 \quad \text{for all } \mathbf{x} \in \mathcal{X}, \mathbf{x}_\star \in \mathcal{X}_\star$$

Examples: • Convex-concave zero-sum • Zero-sum polymatrix

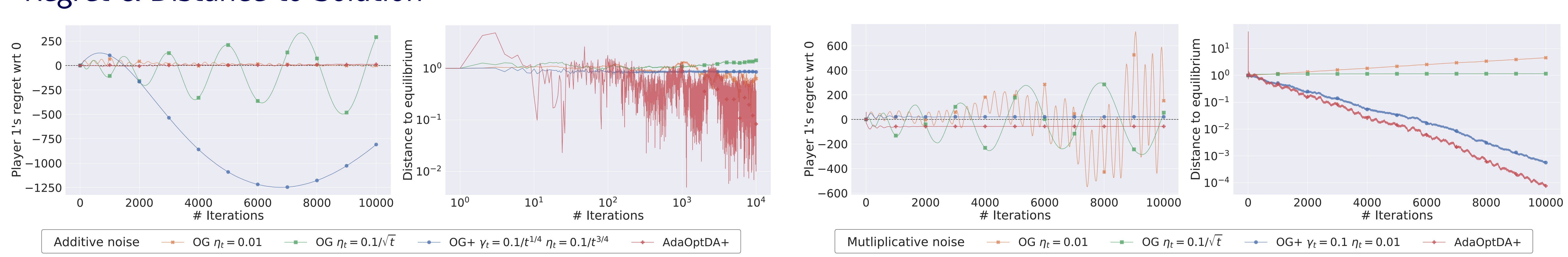
Adaptive Learning Rate

For some fixed, $q \in (0, 1/4]$ we consider the learning rates

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

- The method is adaptive in the following sense
  - Implementable by individual player using only local information and without any prior knowledge of the setting's parameters
  - Guarantee sublinear regret in the adversarial setup
  - Retain the same $\mathcal{O}(\sqrt{T})$ and $\mathcal{O}(1)$ regrets respectively under additive and multiplicative noise when employed by all players.
- Small q provides better fallback guarantee against arbitrary bounded sequence while larger q is more favorable in the game-theoretic setup (e.g., $\mathcal{O}(\exp(1/2q))$ regret)

Regret & Distance to Solution



Future directions. • Trajectory convergence for dual averaging under additive noise • Constraints • Bandit feedback • Partial adherence to the algorithm • Policy regret