Adaptive Learning in Continuous Games: Optimal Regret Bounds and Convergence to Nash Equilibrium

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

Online Learning in Continuous Games

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

- 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
- Nash equilibrium $\mathbf{x}_\star$: $\forall i \in \mathcal{N}, \forall x^i \in \mathcal{X}^i, \ell^i(x^i_\star, \mathbf{x}^{-i}_\star) \le \ell^i(x^i, \mathbf{x}^{-i}_\star)$
- Individual regret of agent i:

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

Limitations of Existing Methods

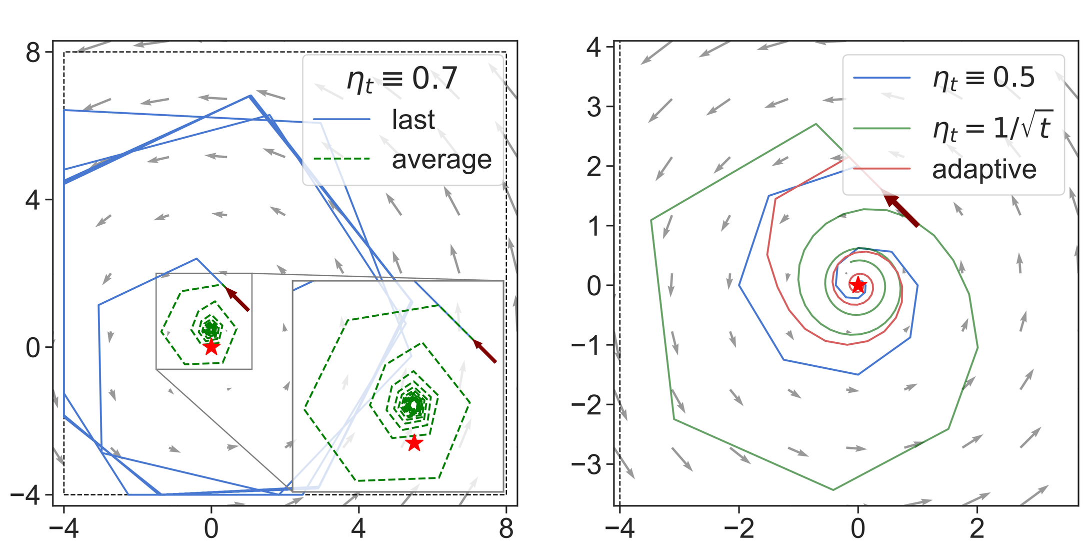
Need for adaptive stepsize

Two players run optimistic gradient with learning rates (η_t) for

$$\ell^1(\mathbf{x}) = -\ell^2(\mathbf{x}) = x^1x^2; \quad \mathcal{X}^1 = \mathcal{X}^2 = [-4, 8]; \quad x_\star = (0, 0)$$

Left: too large stepsize leads to divergence

Right: adaptive method converges fast without need for prior tuning



Need for Dual Averaging update

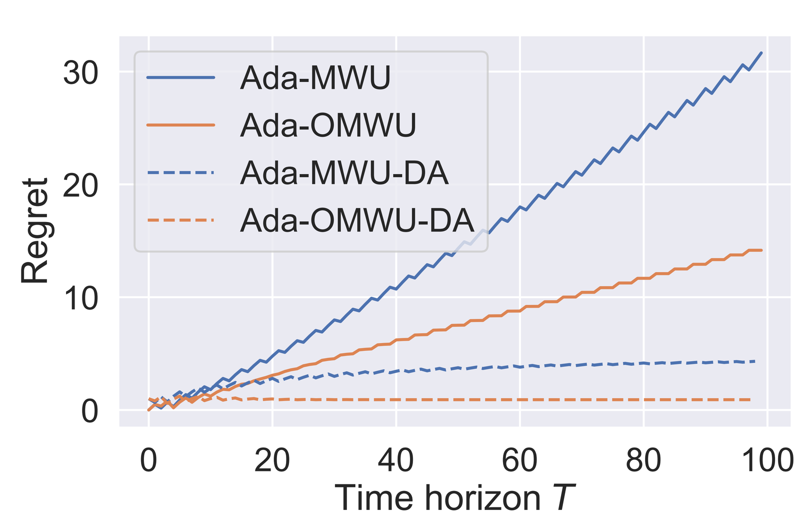
Feedback should enter the algorithm with equal weight

Assume linear loss and simplex-constrained action set.

- $\mathcal{X} = \{(w_1, w_2) \in \mathbb{R}^2_+, w_1 + w_2 = 1\}$
- Feedback sequence:

$$[\underbrace{-e_1, \ldots, -e_1}_{\lceil T/3 \rceil}, \underbrace{-e_2, \ldots, -e_2}_{\lfloor 2T/3 \rfloor}]$$

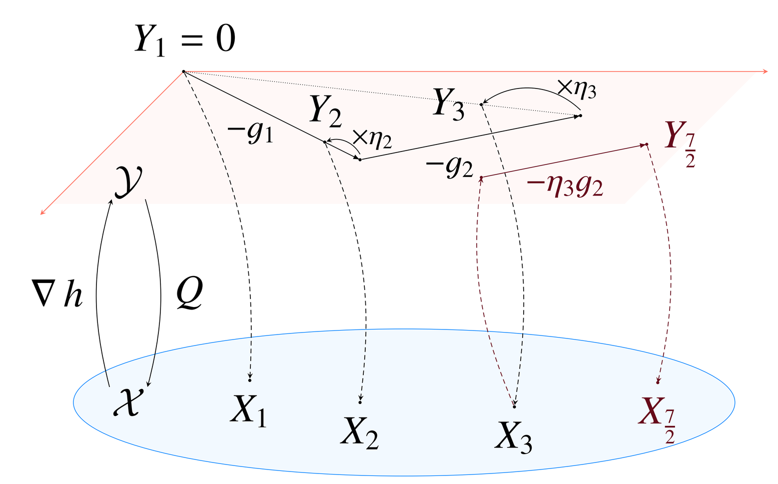
- Algorithm: Adaptive (Optimistic) Multiplicative Weight Update



TL;DR

We introduce for learning in continuous games a family of algorithms that are Adaptive, No-regret, Consistent, and Convergent.

Optimistic Dual Averaging



$$Y^i_t = -\eta^i_t \sum_{s=1}^{t-1} g^s_t$$

$$X^i_t = Q(Y^i_t) = \arg\min_{x\in\mathcal{X}^i} \sum_{s=1}^{t-1} \langle g^i_s, x\rangle + \frac{h^i(x)}{\eta^i_t}$$

$$X^i_{t+\frac{1}{2}} = \arg\min_{x\in\mathcal{X}^i} \langle g^i_{t-1}, x\rangle + \frac{D^i(x, X^i_t)}{\eta^i_t}$$

Regularizer h^i: 1-strongly convex and C^1
Mirror map: $Q^i(y) = \arg\max_{x\in\mathcal{X}^i} \langle y, x\rangle - h^i(x)$
Bregman divergence: $D^i(p, x) = h^i(p) - h^i(x) - \langle \nabla h^i(x), p - x\rangle$

Energy inequality and adaptive learning rate

Fenchel coupling. $F^i(p, y) = h^i(p) + (h^i)^*(y) - \langle y, p\rangle$

Let $\lambda^i_t = 1/\eta^i_t$, $\psi^i_t(p^i) = F^i(p^i, Y^i_t)$, $\varphi^i(p^i) = h^i(p^i) - \min h^i$. Then, for any $p^i \in \mathcal{X}^i$, we have

$$\lambda^i_{t+1} \underbrace{\psi^i_{t+1}(p^i)}_{\text{convergence measure}} \le \lambda^i_t \underbrace{\psi^i_t(p^i)} - \underbrace{\langle g^i_t, X^i_{t+\frac{1}{2}} - p^i\rangle}_{\text{linearized regret}} + (\lambda^i_{t+1} - \lambda^i_t)\varphi^i(p^i)$$
$$+ \langle \underbrace{g^i_t - g^i_{t-1}}_{\text{gradeint variation}}, X^i_{t+\frac{1}{2}} - X^i_{t+1}\rangle - \underbrace{\lambda^i_t D^i(X^i_{t+1}, X^i_{t+\frac{1}{2}}) - \lambda^i_t D^i(X^i_{t+\frac{1}{2}}, X^i_t)}_{\text{distance between successive iterates}}$$

Rearranging, we get

$$\textstyle\sum_{t=1}^T \langle g^i_t, X^i_{t+\frac{1}{2}} - p^i\rangle \le \lambda^i_{T+1}\varphi^i(p^i) + \sum_{t=1}^T \frac{\|g^i_t - g^i_{t-1}\|^2_{(i),*}}{\lambda^i_t} - \sum_{t=2}^T \frac{\lambda^i_{t-1}}{8}\|X^i_{t+\frac{1}{2}} - X^i_{t-\frac{1}{2}}\|^2_{(i)}$$

Take the adaptive learning rate

$$\eta^i_t = \frac{1}{\sqrt{\tau^i + \sum_{s=1}^{t-1}\|g^i_t - g^i_{t-1}\|^2_{(i),*}}} \qquad \text{(Adapt)}$$

- $\tau^i > 0$ can be chosen freely by the player
- η^i_t is thus computed solely based on local information available to each player

Theoretical Guarantees

No-regret: Playing the algorithm against any bounded feedback sequence incurs $\mathcal{O}(\sqrt{T})$ regret.

Consistent: If $\mathcal{X}^i$ is compact and the action profile $\mathbf{x}^{-i}_t$ of all other players converges to some limit profile $\mathbf{x}^{-i}_\infty$, the algorithm itself i converges to best response $\arg\min_{x^i\in\mathcal{X}^i} \ell^i(x^i, \mathbf{x}^{-i}_\infty)$.

Convergent: When employed by all players, the induced sequence of play converges to a Nash equlibrium if either

a. The game is strictly variationally stable.
b. The game is variationally stable and h^i is (sub)differentiable on all $\mathcal{X}^i$.

Variational Stability. A convex game is variationally stable if the set $\mathcal{X}_\star$ of Nash equilibria of the game is nonempty and

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

The game is strictly variationally stable if the inequality holds as a strict inequality whenever $\mathbf{x} \notin \mathcal{X}_\star$.

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

Proof sketch for convergence to NE

1. Show that λ^i_t converges to a finite constant when $t \to +\infty$.
2. Under a suitable divergence metric, establish the quasi-Fejér monotonicity of the iterates with respect to any NE.
3. Derive that $\|\mathbf{X}_{t+\frac{1}{2}} - \mathbf{X}_t\| \to 0$ and $\|\mathbf{X}_t - \mathbf{X}_{t-\frac{1}{2}}\| \to 0$ as $t \to +\infty$.
4. Prove that every cluster point of the sequence of play is a NE and conclude.

Regret minimization does not imply convergence

Each player has two pure strategies. P1 (resp. P2) wants to match the pure strategy of P2 (resp. P3); wheareas P3 wants to match the opposite of the pure strategy of P1.