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

Yu-Guan Hsieh, Kimon Antonakopoulos, and Panayotis Mertikopoulos

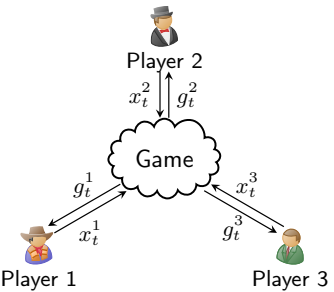
COLT 2021

Online learning in games: Setup

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

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

- Each player i has a convex closed action set $\mathcal{X}^i$ and a loss function $\ell^i \colon \mathcal{X}^1 \times \ldots \times \mathcal{X}^N \to \mathbb{R}$

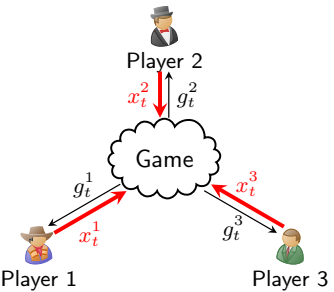


Online learning in games: Setup

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

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

- Each player i has a convex closed action set $\mathcal{X}^i$ and a loss function $\ell^i \colon \mathcal{X}^1 \times \ldots \times \mathcal{X}^N \to \mathbb{R}$

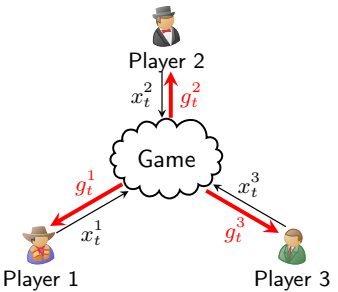


Online learning in games: Setup

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

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

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

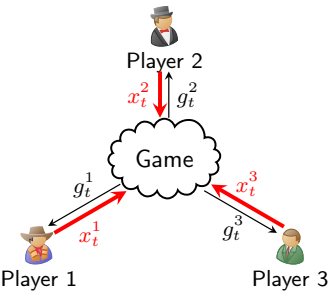


Online learning in games: Setup

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

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

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

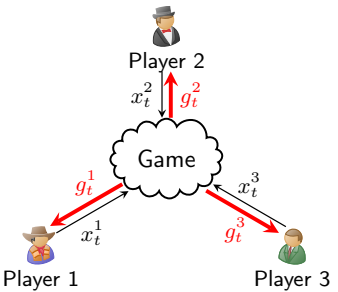


Online learning in games: Setup

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

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

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

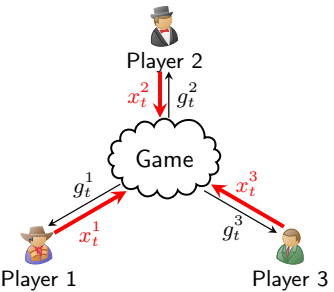


Online learning in games: Setup

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

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

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

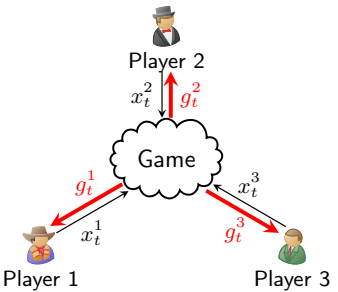


Online learning in games: Setup

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

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

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

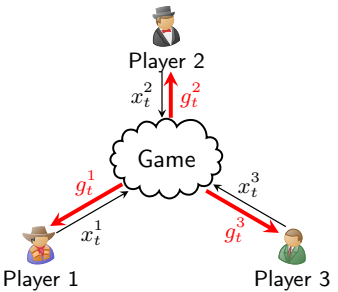


Online learning in games: Setup

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

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

- Each player i has a convex closed action set $\mathcal{X}^i$ and a loss function $\ell^i \colon \mathcal{X}^1 \times \ldots \times \mathcal{X}^N \to \mathbb{R}$
- 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 games: Nash equilibrium and Regret

- Nash equilibrium $\mathbf{x}_\star$: for all $i \in \mathcal{N}$ and all $x^i \in \mathcal{X}^i$, $\ell^i(x^i_\star, \mathbf{x}^{-i}_\star) \leq \ell^i(x^i, \mathbf{x}^{-i}_\star)$
  - Hard to compute in general
  - The players only knows the game via gradient feedback

Online learning in games: Nash equilibrium and Regret

- Nash equilibrium $\mathbf{x}_\star$: for all $i \in \mathcal{N}$ and all $x^i \in \mathcal{X}^i$, $\ell^i(x^i_\star, \mathbf{x}^{-i}_\star) \le \ell^i(x^i, \mathbf{x}^{-i}_\star)$
  - Hard to compute in general
  - The players only knows the game via gradient feedback
- 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 $\mathrm{Reg}^i_T(\mathcal{P}^i) = o(T)$.

Online learning in games: Nash equilibrium and Regret

- Nash equilibrium $\mathbf{x}_\star$: for all $i \in \mathcal{N}$ and all $x^i \in \mathcal{X}^i$, $\ell^i(x^i_\star, \mathbf{x}^{-i}_\star) \le \ell^i(x^i, \mathbf{x}^{-i}_\star)$
  - Hard to compute in general
  - The players only knows the game via gradient feedback
- 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 $\mathrm{Reg}^i_T(\mathcal{P}^i) = o(T)$.
- **Nash equilibrium leads to no regret but the converse is more delicate**

Limitations of existing (optimistic) methods

Fast convergence of sequence of play is mostly proved for **suitably tuned** learning rates

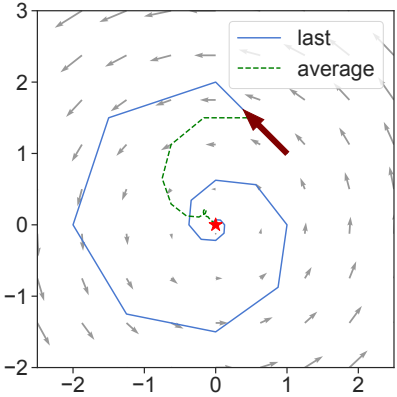
- Two-player planar bilinear zero-sum game

  $\ell^1(\mathbf{x}) = -\ell^2(\mathbf{x}) = x^1 x^2$ where $\mathcal{X}^1 = \mathcal{X}^2 = [-4, 8]$

- The two players play **optimistic** gradient (OG) with constant stepsize $\eta = 0.5$ and $T = 100$

Property

OG converges in bilinear zero-sum games



Limitations of existing (optimistic) methods

Fast convergence of sequence of play is mostly proved for **suitably tuned** learning rates

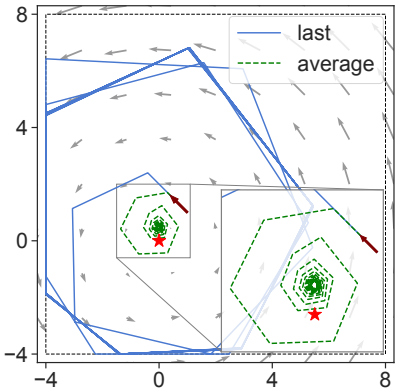
- Two-player planar bilinear zero-sum game

  $\ell^1(\mathbf{x}) = -\ell^2(\mathbf{x}) = x^1 x^2$ where $\mathcal{X}^1 = \mathcal{X}^2 = [-4, 8]$

- The two players play **optimistic** gradient (OG) with constant stepsize $\eta = 0.7$ and $T = 100$

Problem

This only holds when η is small enough



Limitations of existing (optimistic) methods

Fast convergence of sequence of play is mostly proved for **suitably tuned** learning rates

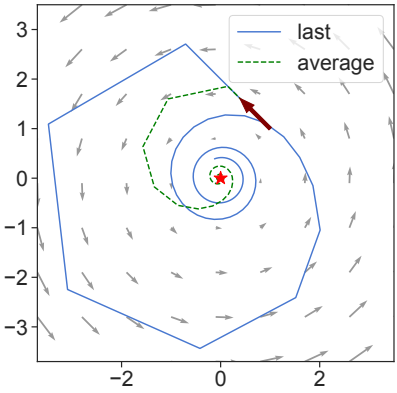
- Two-player planar bilinear zero-sum game

  $\ell^1(\mathbf{x}) = -\ell^2(\mathbf{x}) = x^1 x^2$ where $\mathcal{X}^1 = \mathcal{X}^2 = [-4, 8]$

- The two players play **optimistic** gradient (OG) with decreasing stepsize $\eta_t = 1/\sqrt{t}$ and $T = 100$

Solution?

$\eta_t \propto 1/\sqrt{t} \rightarrow$ slow convergence



Limitations of existing (optimistic) methods

Fast convergence of sequence of play is mostly proved for **suitably tuned** learning rates

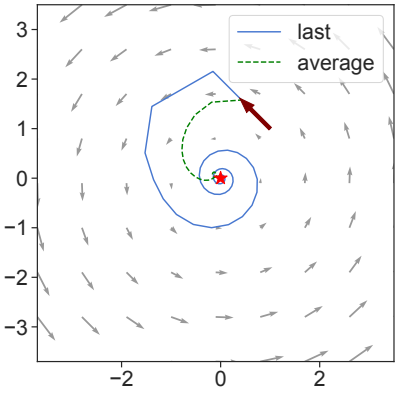
- Two-player planar bilinear zero-sum game

  $\ell^1(\mathbf{x}) = -\ell^2(\mathbf{x}) = x^1 x^2$ where $\mathcal{X}^1 = \mathcal{X}^2 = [-4, 8]$

- The two players play **optimistic** gradient (OG) with adaptive stepsize and $T = 100$

Solution

Adaptive learning ← focus of the work



Limitations of existing (optimistic) methods

Mirror descent type methods with dynamic learning rates **may incur regret**

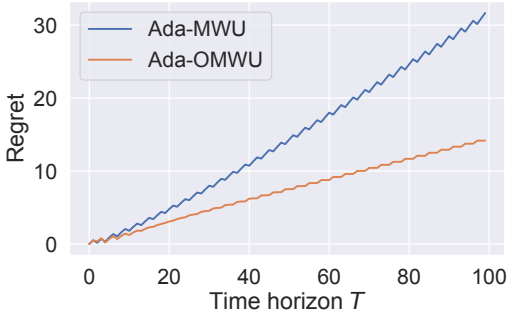
Assume that player 1 has a linear loss and simplex-constrained action set.

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

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

- Adaptive (Optimistic) Multiplicative Weight Update

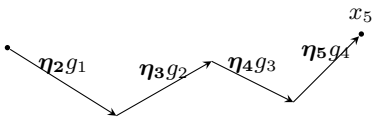
(Example from [Orabona and Pal 16])



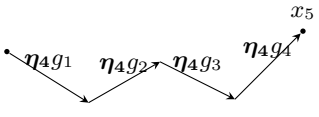
Limitations of existing (optimistic) methods

Mirror descent type methods with dynamic learning rates **may incur regret**

- Cause: new information enters MD with a decreasing weight
- Solution: enter each feedback with equal weight
  E.g. Dual averaging or stabilization technique



MD-type update



DA-type update

Limitations of existing (optimistic) methods

Mirror descent type methods with dynamic learning rates **may incur regret**

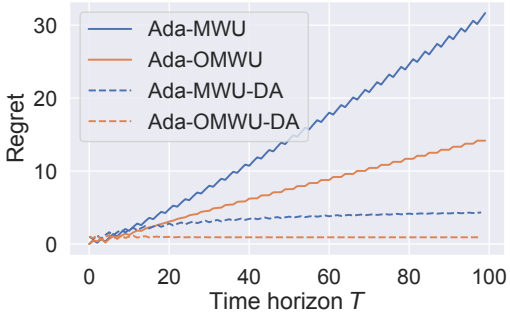
Assume that player 1 has a linear loss and simplex-constrained action set.

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

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

- Adaptive (Optimistic) Multiplicative Weight Update with Dual Averaging

(Example from [Orabona and Pal 16])



Contributions

We introduce a family of algorithms that are

- Adaptive: they do not require any prior tuning or knowledge of the game.

Contributions

We introduce a family of algorithms that are

- Adaptive: they do not require any prior tuning or knowledge of the game.
- No-regret: they achieve $\mathcal{O}(\sqrt{T})$ individual regret against arbitrary opponents.

Contributions

We introduce a family of algorithms that are

- Adaptive: they do not require any prior tuning or knowledge of the game.
- No-regret: they achieve $\mathcal{O}(\sqrt{T})$ individual regret against arbitrary opponents.
- Consistent: they converge to the best response against convergent opponents.

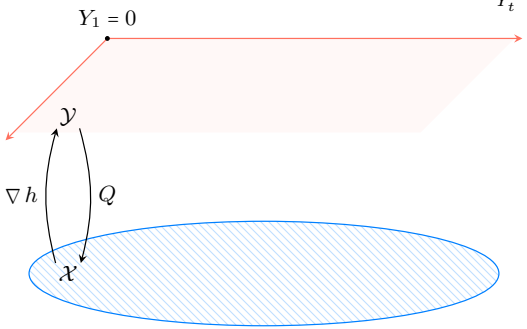
Contributions

We introduce a family of algorithms that are

- Adaptive: they do not require any prior tuning or knowledge of the game.
- No-regret: they achieve $\mathcal{O}(\sqrt{T})$ individual regret against arbitrary opponents.
- Consistent: they converge to the best response against convergent opponents.
- Convergent: if employed by all players in a monotone/variationally stable game, the induced sequence of play converges to Nash equilibrium.

Optimistic Dual Averaging (OptDA)

$$X_t^i = \arg\min_{x \in \mathcal{X}^i} \sum_{s=1}^{t-1} \langle g_s^i, x \rangle + \frac{h^i(x)}{\eta_t^i} = Q\Big(\underbrace{-\eta_t^i \sum_{s=1}^{t-1} g_t^s}_{Y_t^i} \Big), \qquad X_{t+\frac{1}{2}}^i = \arg\min_{x \in \mathcal{X}^i} \langle g_{t-1}^i, x \rangle + \frac{D^i(x, X_t^i)}{\eta_t^i}$$



Regularizer h^i: 1-strongly convex and C^1

Mirror map: $Q^i(y) = \arg\max_{x \in \mathcal{X}^i} \langle y, x \rangle - h^i(x)$

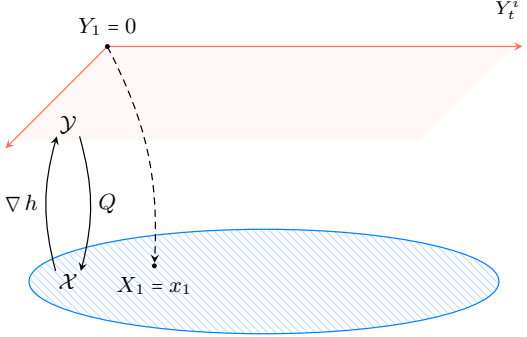
Bregman divergence:

$D^i(p, x) = h^i(p) - h^i(x) - \langle \nabla h^i(x), p - x \rangle$

- Play $x_t = X_{t+\frac{1}{2}}$ and receive feedback g_t
- Accumulate gradient and compute $X_{t+1}, X_{t+\frac{3}{2}}$

Optimistic Dual Averaging (OptDA)

$$X_t^i = \arg\min_{x \in \mathcal{X}^i} \sum_{s=1}^{t-1} \langle g_s^i, x \rangle + \frac{h^i(x)}{\eta_t^i} = Q\Big(\underbrace{-\eta_t^i \sum_{s=1}^{t-1} g_t^s}_{Y_t^i} \Big), \qquad X_{t+\frac{1}{2}}^i = \arg\min_{x \in \mathcal{X}^i} \langle g_{t-1}^i, x \rangle + \frac{D^i(x, X_t^i)}{\eta_t^i}$$



Regularizer h^i: 1-strongly convex and C^1

Mirror map: $Q^i(y) = \arg\max_{x \in \mathcal{X}^i} \langle y, x \rangle - h^i(x)$

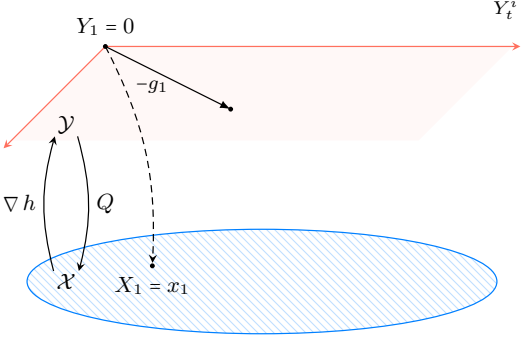
Bregman divergence:

$D^i(p, x) = h^i(p) - h^i(x) - \langle \nabla h^i(x), p - x \rangle$

- Play $x_1 = X_1$ and receive feedback g_1
- Accumulate gradient and compute $X_{t+1}, X_{t+\frac{3}{2}}$

Optimistic Dual Averaging (OptDA)

$$X_t^i = \arg\min_{x \in \mathcal{X}^i} \sum_{s=1}^{t-1} \langle g_s^i, x \rangle + \frac{h^i(x)}{\eta_t^i} = Q\Big(\underbrace{-\eta_t^i \sum_{s=1}^{t-1} g_t^s}_{Y_t^i} \Big), \qquad X_{t+\frac{1}{2}}^i = \arg\min_{x \in \mathcal{X}^i} \langle g_{t-1}^i, x \rangle + \frac{D^i(x, X_t^i)}{\eta_t^i}$$



Regularizer h^i: 1-strongly convex and C^1

Mirror map: $Q^i(y) = \arg\max_{x \in \mathcal{X}^i} \langle y, x \rangle - h^i(x)$

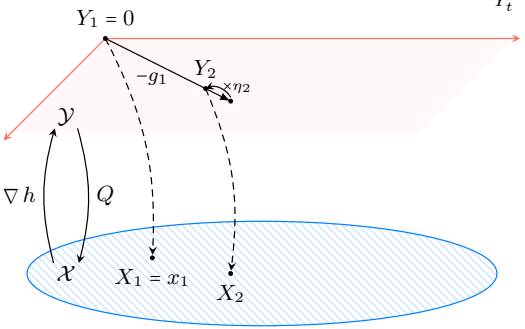
Bregman divergence:

$D^i(p, x) = h^i(p) - h^i(x) - \langle \nabla h^i(x), p - x \rangle$

- Play $x_1 = X_1$ and receive feedback g_1
- Accumulate gradient and compute X_2, $X_{\frac{5}{2}}$

Optimistic Dual Averaging (OptDA)

$$X_t^i = \arg\min_{x \in \mathcal{X}^i} \sum_{s=1}^{t-1} \langle g_s^i, x \rangle + \frac{h^i(x)}{\eta_t^i} = Q\Big(\underbrace{-\eta_t^i \sum_{s=1}^{t-1} g_t^s}_{Y_t^i} \Big), \qquad X_{t+\frac{1}{2}}^i = \arg\min_{x \in \mathcal{X}^i} \langle g_{t-1}^i, x \rangle + \frac{D^i(x, X_t^i)}{\eta_t^i}$$



Regularizer h^i: 1-strongly convex and C^1

Mirror map: $Q^i(y) = \arg\max_{x \in \mathcal{X}^i} \langle y, x \rangle - h^i(x)$

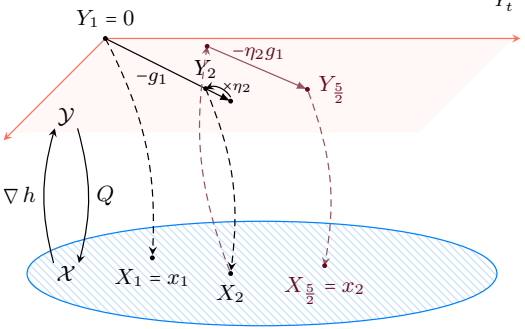
Bregman divergence:

$D^i(p, x) = h^i(p) - h^i(x) - \langle \nabla h^i(x), p - x \rangle$

- Play $x_1 = X_1$ and receive feedback g_1
- Accumulate gradient and compute X_2, $X_{\frac{5}{2}}$

Optimistic Dual Averaging (OptDA)

$$X_t^i = \arg\min_{x\in\mathcal{X}^i} \sum_{s=1}^{t-1} \langle g_s^i, x\rangle + \frac{h^i(x)}{\eta_t^i} = Q\Big(\underbrace{-\eta_t^i \sum_{s=1}^{t-1} g_t^s}_{Y_t^i}\Big), \qquad X_{t+\frac{1}{2}}^i = \arg\min_{x\in\mathcal{X}^i} \langle g_{t-1}^i, x\rangle + \frac{D^i(x, X_t^i)}{\eta_t^i}$$



Regularizer h^i: 1-strongly convex and C^1

Mirror map: $Q^i(y) = \arg\max_{x\in\mathcal{X}^i} \langle y, x\rangle - h^i(x)$

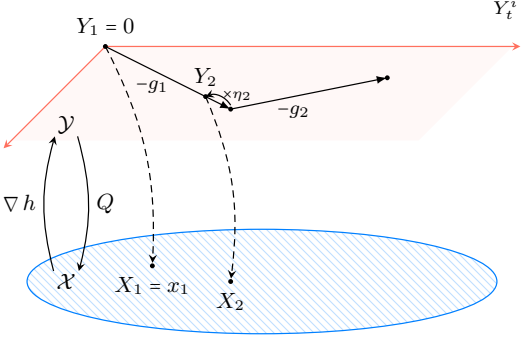
Bregman divergence:

$D^i(p, x) = h^i(p) - h^i(x) - \langle \nabla h^i(x), p - x\rangle$

- Play $x_2 = X_{\frac{5}{2}}$ and receive feedback g_2
- Accumulate gradient and compute X_2, $X_{\frac{5}{2}}$

Optimistic Dual Averaging (OptDA)

$$X_t^i = \arg\min_{x\in\mathcal{X}^i} \sum_{s=1}^{t-1} \langle g_s^i, x\rangle + \frac{h^i(x)}{\eta_t^i} = Q\Big(\underbrace{-\eta_t^i \sum_{s=1}^{t-1} g_t^s}_{Y_t^i}\Big), \qquad X_{t+\frac{1}{2}}^i = \arg\min_{x\in\mathcal{X}^i} \langle g_{t-1}^i, x\rangle + \frac{D^i(x, X_t^i)}{\eta_t^i}$$



Regularizer h^i: 1-strongly convex and C^1

Mirror map: $Q^i(y) = \arg\max_{x\in\mathcal{X}^i} \langle y, x\rangle - h^i(x)$

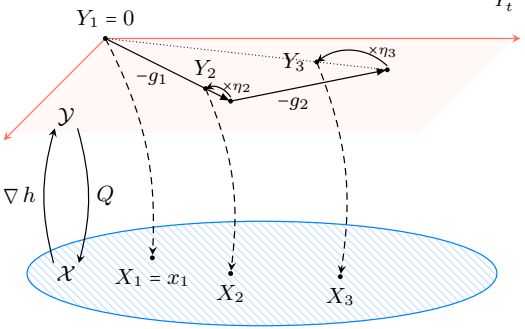
Bregman divergence:

$D^i(p, x) = h^i(p) - h^i(x) - \langle \nabla h^i(x), p - x\rangle$

- Play $x_2 = X_{\frac{5}{2}}$ and receive feedback g_2
- Accumulate gradient and compute X_3, $X_{\frac{7}{2}}$

Optimistic Dual Averaging (OptDA)

$$X_t^i = \arg\min_{x\in\mathcal{X}^i} \sum_{s=1}^{t-1} \langle g_s^i, x\rangle + \frac{h^i(x)}{\eta_t^i} = Q\Big(\underbrace{-\eta_t^i \sum_{s=1}^{t-1} g_t^s}_{Y_t^i}\Big), \qquad X_{t+\frac{1}{2}}^i = \arg\min_{x\in\mathcal{X}^i} \langle g_{t-1}^i, x\rangle + \frac{D^i(x, X_t^i)}{\eta_t^i}$$



Regularizer h^i: 1-strongly convex and C^1

Mirror map: $Q^i(y) = \arg\max_{x\in\mathcal{X}^i} \langle y, x\rangle - h^i(x)$

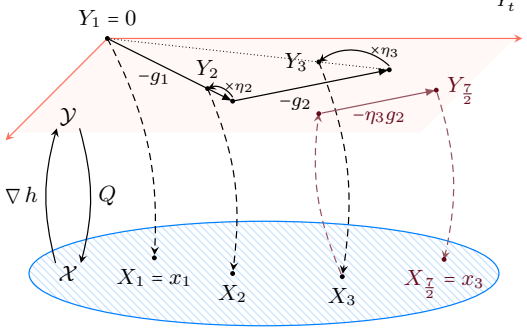
Bregman divergence:

$D^i(p, x) = h^i(p) - h^i(x) - \langle \nabla h^i(x), p - x\rangle$

- Play $x_2 = X_{\frac{5}{2}}$ and receive feedback g_2
- Accumulate gradient and compute X_3, $X_{\frac{7}{2}}$

Optimistic Dual Averaging (OptDA)

$$X_t^i = \arg\min_{x\in\mathcal{X}^i} \sum_{s=1}^{t-1} \langle g_s^i, x\rangle + \frac{h^i(x)}{\eta_t^i} = Q\Big(\underbrace{-\eta_t^i \sum_{s=1}^{t-1} g_t^s}_{Y_t^i}\Big), \qquad X_{t+\frac{1}{2}}^i = \arg\min_{x\in\mathcal{X}^i} \langle g_{t-1}^i, x\rangle + \frac{D^i(x, X_t^i)}{\eta_t^i}$$



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:

$D^i(p, x) = h^i(p) - h^i(x) - \langle \nabla h^i(x), p - x\rangle$

- Play $x_3 = X_{\frac{7}{2}}$ and receive feedback g_3
- Accumulate gradient and compute X_3, $X_{\frac{7}{2}}$

Optimistic Dual Averaging: Examples

- OG-OptDA ▸ $\mathcal{X}^i$ convex closed ▸ $h^i(x) = \dfrac{\|x\|_2^2}{2}$ ▸ Q: Euclidean projection $\Pi_{\mathcal{X}}$

$$X_t^i = \Pi_{\mathcal{X}}(-\eta_t^i \sum_{s=1}^{t-1} g_t^s), \qquad X_{t+\frac{1}{2}}^i = \Pi_{\mathcal{X}}(X_t^i - \eta_t^i g_{t-1}^i)$$

- Stabilized OMWU ▸ $\mathcal{X}^i = \Delta^{d^i-1}$ ▸ $h^i(x) = \sum_{k=1}^{d_i} x_{[k]} \log x_{[k]}$ ▸ Q: Softmax

$$X_{t+\frac{1}{2},[k]}^{(i)} = \frac{\exp(-\eta_t^i(\sum_{s=1}^{t-1} g_{s,[k]} + g_{t-1,[k]}))}{\sum_{l=1}^{d_i} \exp(-\eta_t^i(\sum_{s=1}^{t-1} g_{s,[l]} + g_{t-1,[l]}))}$$

Energy inequality

Suppose that player i runs OptDA or DS-OptMD. Then, for any $p^i \in \mathcal{X}^i$, we have

$$\lambda_{t+1}^i \psi_{t+1}^i(p^i) \leq \lambda_t^i \psi_t^i(p^i) - \langle g_t^i, X_{t+\frac{1}{2}}^i - p^i \rangle + (\lambda_{t+1}^i - \lambda_t^i)\varphi^i(p^i)$$
$$+ \langle g_t^i - g_{t-1}^i, X_{t+\frac{1}{2}}^i - X_{t+1}^i \rangle - \lambda_t^i D^i(X_{t+1}^i, X_{t+\frac{1}{2}}^i) - \lambda_t^i D^i(X_{t+\frac{1}{2}}^i, X_t^i)$$

where $(\psi_t^i)_{t\in\mathbb{N}}$ and φ are non-negative, and $\lambda_t^i = 1/\eta_t^i$.

ψ_t^i is a convergence measure (Bregman divergence or Fenchel coupling)

1. $\psi_t^i(p^i) \geq \frac{1}{2}\|X_t^i - p_t^i\|^2$
2. Reciprocity condition: if $X_t^i \to p^i$ then $\psi_t^i(p^i) \to 0$

Energy inequality

Suppose that player i runs OptDA or DS-OptMD. Then, for any $p^i \in \mathcal{X}^i$, we have

$$\lambda_{t+1}^i \psi_{t+1}^i(p^i) \le \lambda_t^i \psi_t^i(p^i) - \langle g_t^i, X_{t+\frac{1}{2}}^i - p^i \rangle + (\lambda_{t+1}^i - \lambda_t^i)\varphi^i(p^i)$$
$$+ \langle g_t^i - g_{t-1}^i, X_{t+\frac{1}{2}}^i - X_{t+1}^i \rangle - \lambda_t^i D^i(X_{t+1}^i, X_{t+\frac{1}{2}}^i) - \lambda_t^i D^i(X_{t+\frac{1}{2}}^i, X_t^i)$$

where $(\psi_t^i)_{t\in\mathbb{N}}$ and φ are non-negative, and $\lambda_t^i = 1/\eta_t^i$.

Sum the energy inequality from $t = 1$ to T gives

$$\sum_{t=1}^T \langle g_t^i, X_{t+\frac{1}{2}}^i - p^i \rangle \le \lambda_{T+1}^i \varphi^i(p^i) + \sum_{t=1}^T \frac{\|g_t^i - g_{t-1}^i\|_{(i),*}^2}{\lambda_t^i} - \sum_{t=2}^T \frac{\lambda_{t-1}^i}{8} \|X_{t+\frac{1}{2}}^i - X_{t-\frac{1}{2}}^i\|_{(i)}^2 \quad (1)$$

Adaptive learning rate

$$\sum_{t=1}^{T}\langle g_t^i, X_{t+\frac{1}{2}}^i - p^i\rangle \le \lambda_{T+1}^i \varphi^i(p^i) + \sum_{t=1}^{T} \frac{\|g_t^i - g_{t-1}^i\|_{(i),*}^2}{\lambda_t^i} - \sum_{t=2}^{T} \frac{\lambda_{t-1}^i}{8}\|X_{t+\frac{1}{2}}^i - X_{t-\frac{1}{2}}^i\|_{(i)}^2$$

Take the adaptive learning rate

$$\eta_t^i = \frac{1}{\sqrt{\tau^i + \sum_{s=1}^{t-1}\|g_t^i - g_{t-1}^i\|_{(i),*}^2}} \qquad \text{(Adapt)}$$

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

Theoretical guarantees for general convex games

Let player i plays OptDA or DS-OptMD with (Adapt):

- No-regret: If $\mathcal{P}^i \subseteq \mathcal{X}^i$ is bounded and $G = \sup_t \|g_t^i\|$, the regret incurred by the player is bounded as $\mathrm{Reg}_T^i(\mathcal{P}^i) = \mathcal{O}(G\sqrt{T} + G^2)$.

Theoretical guarantees for general convex games

Let player i plays OptDA or DS-OptMD with (Adapt):

- No-regret: If $\mathcal{P}^i \subseteq \mathcal{X}^i$ is bounded and $G = \sup_t \|g_t^i\|$, the regret incurred by the player is bounded as $\mathrm{Reg}_T^i(\mathcal{P}^i) = \mathcal{O}(G\sqrt{T} + G^2)$.
- Consistent: If $\mathcal{X}^i$ is compact and the action profile $\mathbf{x}_t^{-i}$ of all other players converges to some limit profile $\mathbf{x}_\infty^{-i}$, the trajectory of chosen actions of player i converges to the best response set $\arg\min_{x^i \in \mathcal{X}^i} \ell^i(x^i, \mathbf{x}_\infty^{-i})$.

Theoretical guarantees for general convex games: Proof sketch

If $\mathcal{P}^i \subseteq \mathcal{X}^i$ is bounded and $G = \sup_t \|g_t^i\|$, the regret incurred by the player i is bounded as $\mathrm{Reg}_T^i(\mathcal{P}^i) = \mathcal{O}(G\sqrt{T} + G^2)$.

Drop $-\sum_{t=2}^{T} \frac{\lambda_{t-1}^i}{8} \|X_{t+\frac{1}{2}}^i - X_{t-\frac{1}{2}}^i\|_{(i)}^2$ in (1) gives

$$\sum_{t=1}^{T} \langle g_t^i, X_{t+\frac{1}{2}}^i - p^i \rangle \leq \lambda_{T+1}^i \varphi^i(p^i) + \sum_{t=1}^{T} \frac{\|g_t^i - g_{t-1}^i\|_{(i),*}^2}{\lambda_t^i}$$

Applying the AdaGrad lemma shows $\mathrm{Reg}_T^i(\mathcal{P}^i) = \mathcal{O}\left(\sqrt{\sum_{t=1}^{T} \|g_t^i - g_{t-1}^i\|^2} + G^2 \right)$

Variational Stability

Definition [Variationally stable games]

Let $\mathbf{V} = (\nabla_1 \ell^1, \ldots, \nabla_M \ell^M)$. A continuous convex game is variationally stable if the set $\mathcal{X}_\star$ of Nash equilibria of the game is nonempty and

$$\langle \mathbf{V}(\mathbf{x}), \mathbf{x} - \mathbf{x}_\star \rangle = \sum_{i=1}^{N} \langle \nabla_i \ell^i(\mathbf{x}), x^i - x^i_\star \rangle \geq 0 \quad \text{for all } \mathbf{x} \in \mathcal{X},\ \mathbf{x}_\star \in \mathcal{X}_\star. \tag{2}$$

The game is strictly variationally stable if (2) holds as a strict inequality whenever $\mathbf{x} \notin \mathcal{X}_\star$.

Especially, a game is variationally stable if $\mathbf{V}$ is monotone

Examples: • Convex-concave zero-sum games • Zero-sum polymatrix games
• Cournot oligopolies • Kelly auctions

Theoretical guarantees for variationally stable games

If all players use OptDA or DS-OptMD with (Adapt) in a variationally stable game:

- Constant individual regret For all $i \in \mathcal{N}$ and every bounded comparator set $\mathcal{P}^i \subseteq \mathcal{X}^i$, the individual regret of player i is bounded as $\mathrm{Reg}_T^i(\mathcal{P}^i) = \mathcal{O}(1)$.

Theoretical guarantees for variationally stable games

If all players use OptDA or DS-OptMD with (Adapt) in a variationally stable game:

- Constant individual regret For all $i \in \mathcal{N}$ and every bounded comparator set $\mathcal{P}^i \subseteq \mathcal{X}^i$, the individual regret of player i is bounded as $\mathrm{Reg}_T^i(\mathcal{P}^i) = \mathcal{O}(1)$.
- Convergence to Nash equilibrium The induced trajectory of play converges to a Nash equilibrium provided that either of the following is satisfied:

Theoretical guarantees for variationally stable games

If all players use OptDA or DS-OptMD with (Adapt) in a variationally stable game:

- Constant individual regret For all $i \in \mathcal{N}$ and every bounded comparator set $\mathcal{P}^i \subseteq \mathcal{X}^i$, the individual regret of player i is bounded as $\mathrm{Reg}_T^i(\mathcal{P}^i) = \mathcal{O}(1)$.
- Convergence to Nash equilibrium The induced trajectory of play converges to a Nash equilibrium provided that either of the following is satisfied:
  a. The game is strictly variationally stable.

Theoretical guarantees for variationally stable games

If all players use OptDA or DS-OptMD with (Adapt) in a variationally stable game:

- Constant individual regret For all $i \in \mathcal{N}$ and every bounded comparator set $\mathcal{P}^i \subseteq \mathcal{X}^i$, the individual regret of player i is bounded as $\mathrm{Reg}_T^i(\mathcal{P}^i) = \mathcal{O}(1)$.
- Convergence to Nash equilibrium The induced trajectory of play converges to a Nash equilibrium provided that either of the following is satisfied:
  a. The game is strictly variationally stable.
  b. The game is variationally stable and h^i is (sub)differentiable on all $\mathcal{X}^i$.

Theoretical guarantees for variationally stable games

If all players use OptDA or DS-OptMD with (Adapt) in a variationally stable game:

- Constant individual regret For all $i \in \mathcal{N}$ and every bounded comparator set $\mathcal{P}^i \subseteq \mathcal{X}^i$, the individual regret of player i is bounded as $\mathrm{Reg}_T^i(\mathcal{P}^i) = \mathcal{O}(1)$.
- Convergence to Nash equilibrium The induced trajectory of play converges to a Nash equilibrium provided that either of the following is satisfied:
  a. The game is strictly variationally stable.
  b. The game is variationally stable and h^i is (sub)differentiable on all $\mathcal{X}^i$.
  c. The players of a two-player finite zero-sum game follow stabilized OMWU.

  [Highlight: we do not assume uniqueness of Nash equilibrium]

Theoretical guarantees for variationally stable games: Proof sketch

1. Show that λ_t^i converges to a finite constant when $t \to +\infty$.

Theoretical guarantees for variationally stable games: Proof sketch

1. Show that λ_t^i 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 Nash equilibrium $\mathbf{x}_\star$.

Theoretical guarantees for variationally stable games: Proof sketch

1. Show that λ_t^i 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 Nash equilibrium $\mathbf{x}_\star$.
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 $t \to +\infty$.

Theoretical guarantees for variationally stable games: Proof sketch

1. Show that λ_t^i 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 Nash equilibrium $\mathbf{x}_\star$.
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. For a and b (general): Prove that every cluster point of the sequence of play is a Nash equilibrium and conclude.

   For c (OWMU): Prove that the sequence of play has at most one cluster point and subsequently this cluster point must be a Nash equlibrium.

Conclusion and perspective

Adaptive optimistic algorithms

- Achieve no regret
- Converge to Nash equilibrium in many games

For future research:

- What happens when the algorithm does not converge?

Conclusion and perspective

Adaptive optimistic algorithms

- Achieve no regret
- Converge to Nash equilibrium in many games

For future research:

- What happens when the algorithm does not converge?

Thanks for your attention!