Semi Bandit Dynamics in Congestion Games: Convergence to Nash Equilibrium and No-Regret Guarantees.

Ioannis Panageas [1] Stratis Skoulakis[2] Luca Viano [2] Xiao Wang [3] Volkan Cevher [2]

[1]University of California Irvine

[2]École Polytechnique Fédérale de Lausanne (EPFL)

[3]Shanghai University of Finance and Economics (SUFE)

ICML 2023

Congestion Games

Congestion Games

- n selfish agents and m resources E, $|E| = m$.
- For resource $e \in E$ admits a cost function $c_e : \mathbb{N} \mapsto \mathbb{R}_+$

Congestion Games

- n selfish agents and m resources E, $|E| = m$.
- For resource $e \in E$ admits a cost function $c_e : \mathbb{N} \mapsto \mathbb{R}_+$
  - If ℓ agents use resource $e \in E$ then cost $c_e(\ell)$ is its cost.

Congestion Games

Congestion Games

- n selfish agents and m resources E, $|E| = m$.
- For resource $e \in E$ admits a cost function $c_e : \mathbb{N} \mapsto \mathbb{R}_+$
  - If ℓ agents use resource $e \in E$ then cost $c_e(\ell)$ is its cost.

Pure and Mixed strategies

For agent $i \in [n]$,

- *pure strategy* $p_i \subseteq E$ (a path from $s_i \in V$ to $t_i \in V$)

Congestion Games

Congestion Games

- n selfish agents and m resources E, $|E| = m$.
- For resource $e \in E$ admits a cost function $c_e : \mathbb{N} \mapsto \mathbb{R}_+$
  - If ℓ agents use resource $e \in E$ then cost $c_e(\ell)$ is its cost.

Pure and Mixed strategies

For agent $i \in [n]$,

- *pure strategy* $p_i \subseteq E$ (a path from $s_i \in V$ to $t_i \in V$)
- $\mathcal{P}_i = \{$*all possible pure strategies* p_i for agent $i\}$ (all possible (s_i, t_i)-paths)

Congestion Games

Congestion Games

- n selfish agents and m resources E, $|E| = m$.
- For resource $e \in E$ admits a cost function $c_e : \mathbb{N} \mapsto \mathbb{R}_+$
  - If ℓ agents use resource $e \in E$ then cost $c_e(\ell)$ is its cost.

Pure and Mixed strategies

For agent $i \in [n]$,

- *pure strategy* $p_i \subseteq E$ (a path from $s_i \in V$ to $t_i \in V$)
- $\mathcal{P}_i = \{$*all possible pure strategies* p_i for agent $i\}$ (all possible (s_i, t_i)-paths)
- *mixed strategy* $\pi_i \in \Delta(\mathcal{P}_i)$, prob. distr. over $\mathcal{P}_i$.

Congestion Games

Congestion Games

- n selfish agents and m resources E, $|E| = m$.
- For resource $e \in E$ admits a cost function $c_e : \mathbb{N} \mapsto \mathbb{R}_+$
  - If ℓ agents use resource $e \in E$ then cost $c_e(\ell)$ is its cost.

Pure and Mixed strategies

For agent $i \in [n]$,

- *pure strategy* $p_i \subseteq E$ (a path from $s_i \in V$ to $t_i \in V$)
- $\mathcal{P}_i = \{$*all possible pure strategies* p_i for agent $i\}$ (all possible (s_i, t_i)-paths)
- *mixed strategy* $\pi_i \in \Delta(\mathcal{P}_i)$, prob. distr. over $\mathcal{P}_i$.

Loads andd Costs

Given a *pure strategy profile* $p = (p_1, \ldots, p_n)$

- The *load of resource* e, $\ell_e(p) := \sum_{i=1}^{n} \mathrm{I}[e \in p_i]$ (number of agents using e)
- Each agent $i \in [n]$ admits cost

$$C_i(p_i, p_{-i}) = \sum_{e \in p_i} \underbrace{c_e(\ell_e(p))}_{\text{cost of } e}$$

Semi-Bandit Dynamics in Congestion Games

Semi-Bandit Dynamics in Congestion Games

1: **for** each round $t = 1, \ldots, T$ **do**
2: Each agent $i \in [n]$ selects a mixed strategy π_i^t,

Semi-Bandit Dynamics in Congestion Games

Semi-Bandit Dynamics in Congestion Games

1: **for** each round $t = 1, \ldots, T$ **do**
2: Each agent $i \in [n]$ selects a mixed strategy π_i^t,samples $p_i^t \sim \pi_i^t \in \Delta(\mathcal{P}_i)$

Semi-Bandit Dynamics in Congestion Games

Semi-Bandit Dynamics in Congestion Games

1: **for** each round $t = 1, \dots, T$ **do**

2: Each agent $i \in [n]$ selects a mixed strategy π_i^t,samples $p_i^t \sim \pi_i^t \in \Delta(\mathcal{P}_i)$ and suffers cost

$$C_i(p_i^t, p_{-i}^t) := \sum_{e \in p_i^t} c_e(\ell_e(p_i^t, p_{-i}^t)).$$

2: Each agent $i \in [n]$ *only observes* the costs $c_e(\ell_e(p_i^t, p_{-i}^t))$ of its selected resources, $e \in p_i^t$ (*semi-bandit*).

3: **end for**

Semi-Bandit Dynamics in Congestion Games

Semi-Bandit Dynamics in Congestion Games

for each round $t = 1, \ldots, T$ **do**

  Each agent $i \in [n]$ selects a mixed strategy π_i^t,samples $p_i^t \sim \pi_i^t \in \Delta(\mathcal{P}_i)$ and suffers cost

$$C_i(p_i^t, p_{-i}^t) := \sum_{e \in p_i^t} c_e(\ell_e(p_i^t, p_{-i}^t)).$$

  Each agent $i \in [n]$ *only observes* the costs $c_e(\ell_e(p_i^t, p_{-i}^t))$ of its selected resources, $e \in p_i^t$ (*semi-bandit*).

end for

- Each agent selects $\pi_i^t \in \Delta(\mathcal{P}_i)$ to *minimize its overall cost*.

No-regret

Regret

Given a sequence of mixed strategy profiles $\pi_1, \ldots, \pi_T$, *the regret of agent* $i \in [n]$

$$\mathcal{R}_i(T) := \underbrace{\sum_{t=1}^{T} \mathbb{E}_{\pi_i^t, \pi_{-i}^t} \left[C_i(p_i^t, p_{-i}^t) \right]}_{\text{expected cost}} - \underbrace{\min_{p_i \in \mathcal{P}} \sum_{t=1}^{T} \mathbb{E}_{\pi_{-i}^t} \left[C_i(p_i, p_{-i}^t) \right]}_{\text{expected cost of best fixed strategy}}$$

No-regret

Regret

Given a sequence of mixed strategy profiles $\pi_1, \ldots, \pi_T$, *the regret of agent* $i \in [n]$

$$\mathcal{R}_i(T) := \underbrace{\sum_{t=1}^{T} \mathbb{E}_{\pi_i^t, \pi_{-i}^t} \left[C_i(p_i^t, p_{-i}^t) \right]}_{\text{expected cost}} - \underbrace{\min_{p_i \in \mathcal{P}} \sum_{t=1}^{T} \mathbb{E}_{\pi_{-i}^t} \left[C_i(p_i, p_{-i}^t) \right]}_{\text{expected cost of best fixed strategy}}$$

An algorithm $\mathcal{A}$ is *no-regret* if and only if $\mathcal{R}_i^{\mathcal{A}}(T) = o(T)$ for any $\pi_{-i}^1, \ldots, \pi_{-i}^T$.

No-regret

Regret

Given a sequence of mixed strategy profiles $\pi_1, \ldots, \pi_T$, *the regret of agent* $i \in [n]$

$$\mathcal{R}_i(T) := \underbrace{\sum_{t=1}^{T} \mathbb{E}_{\pi_i^t, \pi_{-i}^t} \left[C_i(p_i^t, p_{-i}^t) \right]}_{\text{expected cost}} - \underbrace{\min_{p_i \in \mathcal{P}} \sum_{t=1}^{T} \mathbb{E}_{\pi_{-i}^t} \left[C_i(p_i, p_{-i}^t) \right]}_{\text{expected cost of best fixed strategy}}$$

An algorithm $\mathcal{A}$ is *no-regret* if and only if $\mathcal{R}_i^{\mathcal{A}}(T) = o(T)$ for any $\pi_{-i}^1, \ldots, \pi_{-i}^T$.

◦ Time-averaged experienced cost $\rightarrow$ time-averaged cost of the *best fixed strategy*!

Nash Equilibrium

- What if all agents $i \in [n]$ use a no-regret algorithm $\mathcal{A}$?
- Does the overall system converge to a steady state?

ϵ-Mixed Nash Equilibrium

A mixed strategy profile $\pi^\star := (\pi_1^\star, \dots, \pi_n^\star)$ is an ϵ-*Mixed Nash Equilibrium* iff

$$\underbrace{\mathbb{E}_{\pi_i^\star, \pi_{-i}^\star}[C_i(p_i, p_{-i})]}_{\text{expected cost of agent } i} \leq \underbrace{\min_{\pi_i} \mathbb{E}_{\pi_i, \pi_{-i}^\star}[C_i(p_i, p_{-i})]}_{\text{best response of agent } i} \quad + \quad \epsilon \qquad \text{for each agent } i \in [n]$$

- No agents can decrease be more than ϵ!

Convergence to NE of Semi-Bandit Dynamics

Question ([Cui et al., 2022])

Is there an online learning algorithm $\mathcal{A}$ (under semi-bandit feedback) such that

1. *is no-regret,* $\mathcal{R}_i^{\mathcal{A}}(T) = o(T)$
2. *once adopted by all agents* $\rightarrow$ *convergence to Nash Equilibrium?*

Convergence to NE of Semi-Bandit Dynamics

Question ([Cui et al., 2022])

Is there an online learning algorithm $\mathcal{A}$ (under semi-bandit feedback) such that

1. *is no-regret,* $\mathcal{R}_i^{\mathcal{A}}(T) = o(T)$
2. *once adopted by all agents* $\rightarrow$ *convergence to Nash Equilibrium?*

- $\underbrace{\text{[Awerbuch et al. '04, Dani et al. '08, Audibert et al. '14]}}_{\text{previous no-regret algorithms}}$ $\rightarrow$ **do not guarantee convergence to NE.**
- $\underbrace{\text{[Cui et al., 2022]}}_{\text{convergence to NE}}$ $\rightarrow$ **does not guarantee sublinear regret.**

Our Results

◦ We present an online learning algorithm $\mathcal{A}$ (*Online Gradient Descent with Caratheodory Exploration*)

No-regret guarantees

If agent $i \in [n]$ adopts OGDCE then with probability $1 - \delta$,

$$\underbrace{\sum_{t=1}^{T} \sum_{e \in p_i^t} c_e^t - \min_{p_i^\star \in \mathcal{P}_i} \sum_{e \in p_i^\star} c_e^t}_{\text{regret of agent } i} \leq \mathcal{O}\left(mT^{4/5} \log(1/\delta)\right)$$

Convergence to NE

If all agents adopt OGDCE for $T \geq \Theta(n^{6.5} m^7 / \epsilon^5)$ then with prob. $\geq 1 - \delta$,

Our Results

◦ We present an online learning algorithm $\mathcal{A}$ (*Online Gradient Descent with Caratheodory Exploration*)

No-regret guarantees

If agent $i \in [n]$ adopts OGDCE then with probability $1-\delta$,

$$\underbrace{\sum_{t=1}^{T}\sum_{e\in p_i^t} c_e^t - \min_{p_i^\star \in \mathcal{P}_i}\sum_{e\in p_i^\star} c_e^t}_{\text{regret of agent } i} \leq \mathcal{O}\left(mT^{4/5}\log(1/\delta)\right)$$

Convergence to NE

If all agents adopt OGDCE for $T \geq \Theta(n^{6.5}m^7/\epsilon^5)$ then with prob. $\geq 1-\delta$, $(1-\delta)T$ strategy profiles are ϵ/δ^2-approximate Mixed NE.

Exploration with Bounded-Away Polytopes

Implicit Description $\mathcal{P}_i$

$$\mathcal{P}_i = \{\text{extreme points of polytope } \underbrace{\mathcal{X}_i := \{x \in [0,1]^m : \; A_i \cdot x \leq b_i\}}_{\text{description polytope}}\}$$

Exploration with Bounded-Away Polytopes

Implicit Description $\mathcal{P}_i$

$$\mathcal{P}_i = \{\text{extreme points of polytope } \underbrace{\mathcal{X}_i := \{x \in [0,1]^m : \; A_i \cdot x \leq b_i\}}_{\text{description polytope}}\}$$

$\mathcal{P}_i = \{\text{all possible paths form } s_i \text{ to } t_i\}$ exponential description!!

Exploration with Bounded-Away Polytopes

Implicit Description $\mathcal{P}_i$

$$\mathcal{P}_i = \{\text{extreme points of polytope } \underbrace{\mathcal{X}_i := \{x \in [0,1]^m : \ A_i \cdot x \leq b_i\}}_{\text{description polytope}}\}$$

Exploration with Bounded-Away Polytopes

Implicit Description $\mathcal{P}_i$

$$\mathcal{P}_i = \{\text{extreme points of polytope } \underbrace{\mathcal{X}_i := \{x \in [0,1]^m : \; A_i \cdot x \leq b_i\}}_{\text{description polytope}}\}$$

$$\mathcal{P}_i = \{\text{extreme points of } \mathcal{X}_i\}$$

$$\mathcal{X}^i = \left\{ \begin{aligned} &\sum_{e \in \mathrm{Out}(s_i)} x_e = 1 \\ &\sum_{e \in \mathrm{Out}(s_i)} x_e = 1 \\ &\sum_{e \in \mathrm{Out}(v)} x_e = \sum_{e \in \mathrm{In}(v)} x_e \text{ for all } v \in V \setminus \{s_i, t_i\} \\ &0 \leq x_e \leq 1 \end{aligned} \right\}$$

Exploration with Bounded-Away Polytopes

Implicit Description $\mathcal{P}_i$

$$\mathcal{P}_i = \{\text{extreme points of polytope } \underbrace{\mathcal{X}_i := \{x \in [0,1]^m : \ A_i \cdot x \leq b_i\}}_{\text{description polytope}}\}$$

μ-Bounded Description Polytope

For any parameter $\mu > 0$,

$$\mathcal{X}_i^\mu := \{x \in \mathcal{X}_i : \ x_e \geq \mu \text{ for all } e \in E\}$$

Exploration with Bounded-Away Polytopes

Implicit Description $\mathcal{P}_i$

$$\mathcal{P}_i = \{\text{extreme points of polytope } \underbrace{\mathcal{X}_i := \{x \in [0,1]^m : \ A_i \cdot x \leq b_i\}}_{\text{description polytope}}\}$$

μ-Bounded Description Polytope

For any parameter $\mu > 0$,

$$\mathcal{X}_i^\mu := \{x \in \mathcal{X}_i : \ x_e \geq \mu \text{ for all } e \in E\}$$

Online Gradient Descent over X_i^μ

$$x_i^{t+1} \leftarrow \Pi_{\mathcal{X}_i^\mu}\left[x_i^t - \gamma \cdot \hat{c}^t\right] \quad \text{where } \hat{c}_e^t \leftarrow \frac{c_e^t}{x_e^t} \cdot \mathbb{1}\left[e \in p_i^t\right]$$

Exploration with Bounded-Away Polytopes

Implicit Description $\mathcal{P}_i$

$$\mathcal{P}_i = \{\text{extreme points of polytope } \underbrace{\mathcal{X}_i := \{x \in [0,1]^m : \ A_i \cdot x \leq b_i\}}_{\text{description polytope}}\}$$

μ-Bounded Description Polytope

For any parameter $\mu > 0$,

$$\mathcal{X}_i^{\mu} := \{x \in \mathcal{X}_i : \ x_e \geq \mu \text{ for all } e \in E\}$$

Online Gradient Descent over X_i^{μ}

$$x_i^{t+1} \leftarrow \Pi_{\mathcal{X}_i^{\mu}} \left[x_i^t - \gamma \cdot \hat{c}^t\right] \quad \text{where } \hat{c}_e^t \leftarrow \frac{c_e^t}{x_e^t} \cdot \mathbb{1}\left[e \in p_i^t\right]$$

- $\mathrm{E}\left[\hat{c}_t\right] = c_t$ (*Unbias*)
- $||\hat{c}_t|| \leq \mathcal{O}(1/\mu)$ (*Bounded Variance*)

Summary

Take-Away

- OGDCE admits $o(T)$ regret no matter the choices of the other agents.
- Once adopted by all agents $\rightarrow$ convergence to NE.

Summary

Take-Away

- OGDCE admits $o(T)$ regret no matter the choices of the other agents.
- Once adopted by all agents $\rightarrow$ convergence to NE.

Future Directions

- Tighter regret guarantees ($\mathcal{O}(\sqrt{T})$)?
- Faster convergence rates to NE?
- Stronger notion of convergence (last-iterate)?

Summary

Take-Away

- OGDCE admits $o(T)$ regret no matter the choices of the other agents.
- Once adopted by all agents $\rightarrow$ convergence to NE.

Future Directions

- Tighter regret guarantees ($\mathcal{O}(\sqrt{T})$)?
- Faster convergence rates to NE?
- Stronger notion of convergence (last-iterate)?

Thank you!

References I

[0] Cui, Q., Xiong, Z., Fazel, M., and Du, S. S. (2022).
Learning in congestion games with bandit feedback.
16, 17