Learning Nash Equilibria with Bandit Feedback

Maryam Kamgarpour

Electrical & Computer Engineering, UBC

Jan 29, 2021
Pacific Interdisciplinary Hub on Optimal Transport

Outline

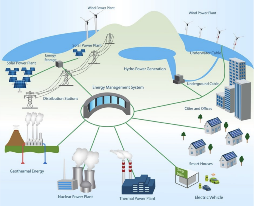
Background - dynamical systems and control

Decision-making in environments
that change and are uncertain

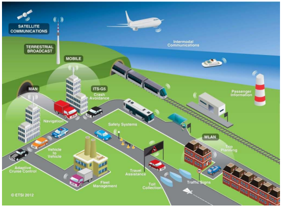
Control systems evolution

from single systems in predictable environments

to . . .



networks



dynamic interactions

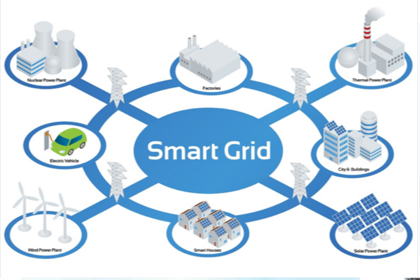
unknown terrains

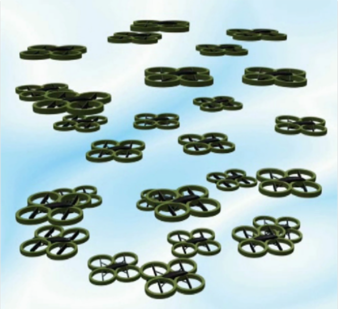
Research thread

- Develop fundamental understanding of decision-making under uncertainty
- Design algorithms with provable safety and performance guarantees

Multi-agent systems

Interacting agents with coupled objectives and constraints





How do players learn to optimize
given only local information?

The rest of the talk

with Tatiana Tatarenko, TU Darmstadt, Germany

- T. Tatarenko, M. Kamgarpour, Bandit Online Learning of Nash Equilibria in Monotone Games, 2020
- T. Tatarenko, M. Kamgarpour, Learning Generalized Nash Equilibria in a Class of Convex Games, IEEE Transactions on Automatic Control, 2019
- T. Tatarenko, M. Kamgarpour, Minimizing Regret of Bandit Online Optimization in Unconstrained Action Spaces, 2018

Outline

Players objectives and constraints

Game $\Gamma(N, \{A^i\}, \{J^i\})$ with N agents/players

- action $\boldsymbol{a}^i \in A^i \subset \mathbb{R}^d$
- joint action $\boldsymbol{a} \in \boldsymbol{A} = A^1 \times \cdots \times A^N \subseteq \mathbb{R}^{Nd}$
- cost $J^i : \mathbb{R}^{Nd} \to \mathbb{R}$, $J^i(\boldsymbol{a}^i, \boldsymbol{a}^{-i})$

Players objectives and constraints

Game $\Gamma(N, \{A^i\}, \{J^i\})$ with N agents/players

- action $\boldsymbol{a}^i \in A^i \subset \mathbb{R}^d$
- joint action $\boldsymbol{a} \in \boldsymbol{A} = A^1 \times \cdots \times A^N \subseteq \mathbb{R}^{Nd}$
- cost $J^i : \mathbb{R}^{Nd} \to \mathbb{R}$, $J^i(\boldsymbol{a}^i, \boldsymbol{a}^{-i})$

Convex game

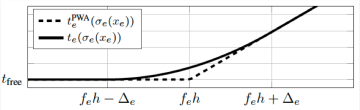
- A^i: convex and compact
- $J^i(\boldsymbol{a}^i, \boldsymbol{a}^{-i})$: continuously differentiable in $\boldsymbol{a}$, convex in $\boldsymbol{a}^i$

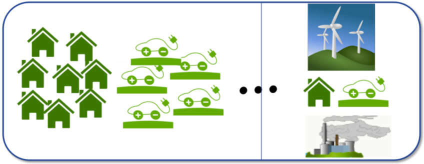
Examples of convex games

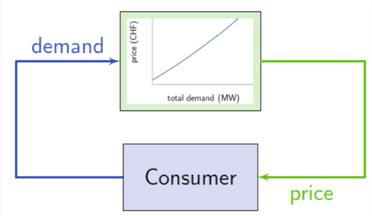
- Mixed strategy extensions of finite action games
  - A^i: probability simplex, $J^i(\boldsymbol{a}^i, \boldsymbol{a}^{-i})$ linear in $\boldsymbol{a}^i$

Examples of convex games

- Mixed strategy extensions of finite action games
  - A^i: probability simplex, $J^i(\boldsymbol{a}^i, \boldsymbol{a}^{-i})$ linear in $\boldsymbol{a}^i$
- Traffic networks, communication networks, power networks







Characterizing Nash equilibria

- $\boldsymbol{a}^* \in \boldsymbol{A}$ is a Nash equilibrium (NE): for each $i = 1, \ldots, N$

$$J^i(\boldsymbol{a}^{*i}, \boldsymbol{a}^{*-i}) \leq J^i(\boldsymbol{a}^i, \boldsymbol{a}^{*-i}), \quad \forall \boldsymbol{a}^i \in A^i$$

Characterizing Nash equilibria

- $\boldsymbol{a}^* \in \boldsymbol{A}$ is a Nash equilibrium (NE): for each $i = 1, \ldots, N$

$$J^i(\boldsymbol{a}^{*i}, \boldsymbol{a}^{*-i}) \leq J^i(\boldsymbol{a}^i, \boldsymbol{a}^{*-i}), \quad \forall \boldsymbol{a}^i \in A^i$$

- NE exists in convex games

Characterizing Nash equilibria

- $\boldsymbol{a}^* \in \boldsymbol{A}$ is a Nash equilibrium (NE): for each $i = 1, \ldots, N$

$$J^i(\boldsymbol{a}^{*i}, \boldsymbol{a}^{*-i}) \leq J^i(\boldsymbol{a}^i, \boldsymbol{a}^{*-i}), \quad \forall \boldsymbol{a}^i \in A^i$$

- NE exists in convex games

Variational inequality (VI) characterization of NE

- game mapping $\boldsymbol{M} : \mathbb{R}^{Nd} \to \mathbb{R}^{Nd}$

$$\boldsymbol{M}(\boldsymbol{a}) = [\nabla_{\boldsymbol{a}^i} J^i(\boldsymbol{a}^i, \boldsymbol{a}^{-i})]_{i=1}^N$$

- $\boldsymbol{a}^*$ is a NE $\iff \underbrace{\boldsymbol{M}(\boldsymbol{a}^*)^T(\boldsymbol{a} - \boldsymbol{a}^*) \geq 0, \forall \boldsymbol{a} \in \boldsymbol{A}}_{\text{VI problem given } \boldsymbol{M} \text{ and } \boldsymbol{A}}$

[Facchinei, Pang, 2007]

Games versus optimization problems

Variational Inequality problem VI($\boldsymbol{M}, \boldsymbol{A}$)

Given $\boldsymbol{M} : \mathbb{R}^{Nd} \to \mathbb{R}^{Nd}$, $\boldsymbol{A} \subset \mathbb{R}^{Nd}$, find $\boldsymbol{a}^* \in \boldsymbol{A}$

$$\boldsymbol{M}(\boldsymbol{a}^*)^T(\boldsymbol{a} - \boldsymbol{a}^*) \geq 0, \forall \boldsymbol{a} \in \boldsymbol{A}$$

- if $\boldsymbol{M} = \nabla f$ for some $f : \boldsymbol{A} \to \mathbb{R}$, then VI is the first-order optimality condition for $\min_{\boldsymbol{a} \in \boldsymbol{A}} f(\boldsymbol{a})$

Games versus optimization problems

Variational Inequality problem VI($\boldsymbol{M}, \boldsymbol{A}$)

Given $\boldsymbol{M} : \mathbb{R}^{Nd} \to \mathbb{R}^{Nd}$, $\boldsymbol{A} \subset \mathbb{R}^{Nd}$, find $\boldsymbol{a}^* \in \boldsymbol{A}$

$$\boldsymbol{M}(\boldsymbol{a}^*)^T(\boldsymbol{a} - \boldsymbol{a}^*) \geq 0, \forall \boldsymbol{a} \in \boldsymbol{A}$$

- if $\boldsymbol{M} = \nabla f$ for some $f : \boldsymbol{A} \to \mathbb{R}$, then VI is the first-order optimality condition for $\min_{\boldsymbol{a} \in \boldsymbol{A}} f(\boldsymbol{a})$

- in a game $\boldsymbol{M}(\boldsymbol{a}) = [\nabla_{\boldsymbol{a}^i} J^i(\boldsymbol{a}^i, \boldsymbol{a}^{-i})]_{i=1}^N$ is a pseudo-gradient
  - is gradient if the Jacobian $J\boldsymbol{M}(\boldsymbol{a})$ is symmetric

Example - matching pennies

- zero-sum game of matching pennies
  - row-player, column-player

$$\begin{array}{cc} & \begin{array}{cc} \text{head} & \text{tail} \end{array} \\ \begin{array}{c} \text{head} \\ \text{tail} \end{array} & \left[\begin{array}{cc} (1, -1) & (-1, 1) \\ (-1, 1) & (1, -1) \end{array} \right] \end{array}$$

Example - matching pennies

- zero-sum game of matching pennies
  - row-player, column-player

$$
\begin{array}{cc}
 & \begin{array}{cc} \text{head} & \text{tail} \end{array} \\
\begin{array}{c} \text{head} \\ \text{tail} \end{array} & \begin{bmatrix} (1,-1) & (-1,1) \\ (-1,1) & (1,-1) \end{bmatrix}
\end{array}
$$

- mixed strategies: $\boldsymbol{a}^i$ probability of player i choosing head

$$J^1(\boldsymbol{a}^1, \boldsymbol{a}^2) = \begin{bmatrix} \boldsymbol{a}^1 & 1-\boldsymbol{a}^1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{a}^2 \\ 1-\boldsymbol{a}^2 \end{bmatrix}$$

- game mapping is not a gradient

$$\boldsymbol{M}(\boldsymbol{a}^1, \boldsymbol{a}^2) = \begin{bmatrix} 0 & 4 \\ -4 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{a}^1 \\ \boldsymbol{a}^2 \end{bmatrix} + \begin{bmatrix} -2 \\ 2 \end{bmatrix}$$

Seeking equilibria with limited information

Each player observes only her cost for a played action

- zero-order information: $J_t^i = J^i(\boldsymbol{a}_t^i, \boldsymbol{a}_t^{-i})$
- black-box access to the function

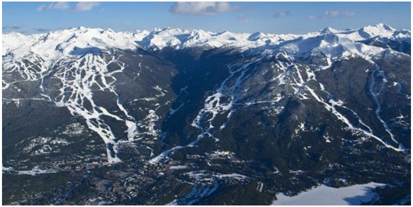
How should she play to ensure convergence to a Nash equilibrium?

Seeking equilibria with limited information

Each player observes only her cost for a played action

- zero-order information: $J_t^i = J^i(\boldsymbol{a}_t^i, \boldsymbol{a}_t^{-i})$
- black-box access to the function

How should she play to ensure convergence to a Nash equilibrium?



Zero-order information in games

Use function evaluations $J^i_t = J^i(\boldsymbol{a}^i_t, \boldsymbol{a}^{-i}_t)$ to estimate gradient?

- query J^i at $\boldsymbol{a}^i_{t+1} = \boldsymbol{a}^i_t + \delta$ and use finite difference
- feedback: $J^i_{t+1} = J^i(\boldsymbol{a}^i_{t+1}, \boldsymbol{a}^{-i}_{t+1})$, can't control $\boldsymbol{a}^{-i}_{t+1}$

How should she play to ensure convergence to a Nash equilibrium?

Zero-order information in games

Use function evaluations $J_t^i = J^i(\boldsymbol{a}_t^i, \boldsymbol{a}_t^{-i})$ to estimate gradient?

- query J^i at $\boldsymbol{a}_{t+1}^i = \boldsymbol{a}_t^i + \delta$ and use finite difference
- feedback: $J_{t+1}^i = J^i(\boldsymbol{a}_{t+1}^i, \boldsymbol{a}_{t+1}^{-i})$, can't control $\boldsymbol{a}_{t+1}^{-i}$

How should she play to ensure convergence to a Nash equilibrium?

Randomization helps in learning

- each player samples her action from a distribution
$$\boldsymbol{a}_t^i \sim p(\boldsymbol{\mu}_t^i, \sigma_t)$$
- mean $\boldsymbol{\mu}^i$: updated greedily based on player's observed cost
- variance σ^i: encourages exploring non-greedy strategies

Randomization helps in learning

- each player samples her action from a distribution
$$\boldsymbol{a}_t^i \sim p(\boldsymbol{\mu}_t^i, \sigma_t)$$
- mean $\boldsymbol{\mu}^i$: updated greedily based on player's observed cost
- variance σ^i: encourages exploring non-greedy strategies

decision making when faced with unknown cost functions:
exploitation and exploration

Learning-based algorithm iterates

- actions $\boldsymbol{a}^i$ and states $\boldsymbol{\mu}^i$ of each player are updated as

$$\text{play: } \boldsymbol{a}_t^i \sim \mathcal{N}(\boldsymbol{\mu}_t^i, \sigma_t^2 I), \text{ receive: } J_t^i = J^i(\boldsymbol{a}_t^i, \boldsymbol{a}_t^{-i})$$

$$\boldsymbol{\mu}_{t+1}^i = \mathsf{Proj}_{A^i}\big[\boldsymbol{\mu}_t^i - \beta_t J_t^i \frac{\boldsymbol{a}_t^i - \boldsymbol{\mu}_t^i}{\sigma_t^2}\big]$$

Learning-based algorithm iterates

- actions $\boldsymbol{a}^i$ and states $\boldsymbol{\mu}^i$ of each player are updated as

$$\text{play: } \boldsymbol{a}_t^i \sim \mathcal{N}(\boldsymbol{\mu}_t^i, \sigma_t^2 I), \text{ receive: } J_t^i = J^i(\boldsymbol{a}_t^i, \boldsymbol{a}_t^{-i})$$

$$\boldsymbol{\mu}_{t+1}^i = \mathsf{Proj}_{A^i}\left[\boldsymbol{\mu}_t^i - \beta_t J_t^i \frac{\boldsymbol{a}_t^i - \boldsymbol{\mu}_t^i}{\sigma_t^2}\right]$$

Learning-based algorithm iterates

- actions $\boldsymbol{a}^i$ and states $\boldsymbol{\mu}^i$ of each player are updated as

$$\boldsymbol{a}_t^i \sim \mathcal{N}(\boldsymbol{\mu}_t^i, \sigma_t^2 I)$$

$$\boldsymbol{\mu}_{t+1}^i = \mathsf{Proj}_{A^i}\left[\boldsymbol{\mu}_t^i - \beta_t \underbrace{J_t^i \frac{\boldsymbol{a}_t^i - \boldsymbol{\mu}_t^i}{\sigma_t^2}}_{\hat{\boldsymbol{M}}^i(\boldsymbol{a}_t, \boldsymbol{\mu}_t^i)}\right]$$

Learning-based algorithm iterates

- actions $\boldsymbol{a}^i$ and states $\boldsymbol{\mu}^i$ of each player are updated as

$$\boldsymbol{a}_t^i \sim \mathcal{N}(\boldsymbol{\mu}_t^i, \sigma_t^2 I)$$
$$\boldsymbol{\mu}_{t+1}^i = \mathsf{Proj}_{A^i}\big[\boldsymbol{\mu}_t^i - \beta_t \underbrace{J_t^i \frac{\boldsymbol{a}_t^i - \boldsymbol{\mu}_t^i}{\sigma_t^2}}_{\hat{\boldsymbol{M}}^i(\boldsymbol{a}_t, \boldsymbol{\mu}_t^i)}\big]$$

- samples of gradient with respect to cost in mixed strategies

$$\mathrm{E}_{\boldsymbol{a}_t}\{\hat{\boldsymbol{M}}^i(\boldsymbol{a}_t, \boldsymbol{\mu}_t^i)\} = \frac{\partial \tilde{J}^i(\boldsymbol{\mu}_t)}{\partial \boldsymbol{\mu}^i}$$
$$\tilde{J}^i(\boldsymbol{\mu}) = \int_{\mathbb{R}^{Nd}} J^i(\boldsymbol{y}) p_{\boldsymbol{\mu}^1}(\boldsymbol{y}^1) \dots p_{\boldsymbol{\mu}^N}(\boldsymbol{y}^N) d\boldsymbol{y}$$

Randomization for gradient estimation

Let $f : \mathbb{R}^n \to \mathbb{R}$, $p(\boldsymbol{y})$ a probability density function

$$f_\sigma(\boldsymbol{\mu}) = \int_{\mathbb{R}^n} f(\boldsymbol{\mu} + \sigma \boldsymbol{y}) p(\boldsymbol{y}) d\boldsymbol{y}$$

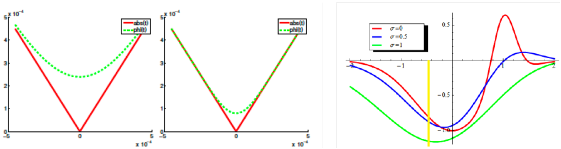
- bandit learning and regret minimization [Flaxman et al. 2006], [Bravo et al. 2019]
- stochastic and zero-order optimization [Nesterov 2010], [Ghadimi, Lan 2014]

Randomization for gradient estimation

Let $f : \mathbb{R}^n \to \mathbb{R}$, $p(\boldsymbol{y})$ a probability density function

$$f_\sigma(\boldsymbol{\mu}) = \int_{\mathbb{R}^n} f(\boldsymbol{\mu} + \sigma \boldsymbol{y}) p(\boldsymbol{y}) d\boldsymbol{y}$$

- bandit learning and regret minimization [Flaxman et al. 2006], [Bravo et al. 2019]
- stochastic and zero-order optimization [Nesterov 2010], [Ghadimi, Lan 2014]
- non-smooth optimization [Duchi et al. 2012]
- non-convex graduated optimization [Mobhai 2012], [Levy, Hazan 2015]



left: smoothing absolute value, right: graduated optimization

Interpretation as a stochastic optimization procedure

Player i: $\boldsymbol{\mu}_{t+1}^i = \mathsf{Proj}_{A^i}\left[\boldsymbol{\mu}_t^i - \beta_t J_t^i \frac{\boldsymbol{a}_t^i - \boldsymbol{\mu}_t^i}{\sigma_t^2}\right]$

Stacking players' iterates, the algorithm is

$$\boldsymbol{\mu}_t = \mathsf{Proj}_{\boldsymbol{A}}[\boldsymbol{\mu}_t - \beta_t\big(\boldsymbol{M}(\boldsymbol{\mu}_t) + \boldsymbol{Q}(\boldsymbol{\mu}_t, \sigma_t) + \boldsymbol{R}(\boldsymbol{\mu}_t, \boldsymbol{a}_t, \sigma_t)\big)]$$

- $\boldsymbol{M}$ game mapping, stacked gradients of players' cost functions
- $\boldsymbol{Q}$ difference in the gradient of the smoothed and original cost
- $\boldsymbol{R}$ stochastic noise term, $\mathrm{E}_{\boldsymbol{a}_t}\boldsymbol{R}(\boldsymbol{\mu}_t, \boldsymbol{a}_t, \sigma_t) = 0$

Convergence of the algorithm

Assumptions

- strictly monotone: $\big(M(\boldsymbol{a}) - M(\boldsymbol{a}')\big)^T(\boldsymbol{a} - \boldsymbol{a}') > 0 \; \forall \boldsymbol{a}, \boldsymbol{a}' \in \boldsymbol{A}$
- Lipschitz: $\|(M(\boldsymbol{a}) - M(\boldsymbol{a}')\| \leq L\|\boldsymbol{a} - \boldsymbol{a}'\| \; \forall \boldsymbol{a}, \boldsymbol{a}' \in \boldsymbol{A}$

Convergence of the algorithm

Assumptions

- strictly monotone: $\big(M(\boldsymbol{a}) - M(\boldsymbol{a}')\big)^T(\boldsymbol{a}-\boldsymbol{a}') > 0 \ \forall \boldsymbol{a}, \boldsymbol{a}' \in \boldsymbol{A}$
- Lipschitz: $\|(M(\boldsymbol{a}) - M(\boldsymbol{a}')\| \leq L\|\boldsymbol{a}-\boldsymbol{a}'\| \ \forall \boldsymbol{a}, \boldsymbol{a}' \in \boldsymbol{A}$

Theorem [TT, MK TAC 2019]

Choose $\beta_t, \sigma_t \to 0$ such that

$$\sum_{t=0}^{\infty} \beta_t = \infty, \ \sum_{t=0}^{\infty} \beta_t \sigma_t < \infty \ \sum_{t=0}^{\infty} \frac{\beta_t^2}{\sigma_t^2} < \infty$$

Then,

- state $\boldsymbol{\mu}_t$ converges almost surely to a Nash equilibrium $\boldsymbol{\mu}^*$
- action $\boldsymbol{a}_t$ converges in probability to $\boldsymbol{\mu}^*$

Proof sketch

Approach: show $\|\boldsymbol{\mu}_t - \boldsymbol{\mu}^*\|^2$ sufficiently decreases at each iteration

$$\boldsymbol{\mu}_{t+1} = \mathsf{Proj}_{\boldsymbol{A}}[\boldsymbol{\mu}_t - \beta_t\big(\boldsymbol{M}(\boldsymbol{\mu}_t) + \boldsymbol{Q}(\boldsymbol{\mu}_t, \sigma_t) + \boldsymbol{R}(\boldsymbol{\mu}_t, \boldsymbol{a}_t, \sigma_t)\big)]$$

$$\mathrm{E}\{\|\boldsymbol{\mu}_{t+1} - \boldsymbol{\mu}^*\|^2\} \quad \leq \|\boldsymbol{\mu}_t - \boldsymbol{\mu}^*\|^2 + \underbrace{\xi_t}_{O(\beta_t\sigma_t + \frac{\beta_t^2}{\sigma_t^2})} - \beta_t \underbrace{\boldsymbol{M}(\boldsymbol{\mu}_t)^T(\boldsymbol{\mu}_t - \boldsymbol{\mu}^*)}_{\geq 0}$$

[Robbins and Siegmund, 1985]

- $\|\boldsymbol{\mu}_t - \boldsymbol{\mu}^*\|^2$ converges as $t \to \infty$
- $\sum_{t=0}^{\infty} \beta_t \boldsymbol{M}(\boldsymbol{\mu}_t)^T(\boldsymbol{\mu}_t - \boldsymbol{\mu}^*) < \infty$

Proof sketch

Approach: show $\|\boldsymbol{\mu}_t - \boldsymbol{\mu}^*\|^2$ sufficiently decreases at each iteration

$$\boldsymbol{\mu}_{t+1} = \mathsf{Proj}_{\boldsymbol{A}}[\boldsymbol{\mu}_t - \beta_t\big(\boldsymbol{M}(\boldsymbol{\mu}_t) + \boldsymbol{Q}(\boldsymbol{\mu}_t, \sigma_t) + \boldsymbol{R}(\boldsymbol{\mu}_t, \boldsymbol{a}_t, \sigma_t)\big)]$$

$$\mathrm{E}\{\|\boldsymbol{\mu}_{t+1} - \boldsymbol{\mu}^*\|^2\} \quad \leq \|\boldsymbol{\mu}_t - \boldsymbol{\mu}^*\|^2 + \underbrace{\xi_t}_{O(\beta_t\sigma_t + \frac{\beta_t^2}{\sigma_t^2})} - \beta_t \underbrace{\boldsymbol{M}(\boldsymbol{\mu}_t)^T(\boldsymbol{\mu}_t - \boldsymbol{\mu}^*)}_{\geq 0}$$

[Robbins and Siegmund, 1985]

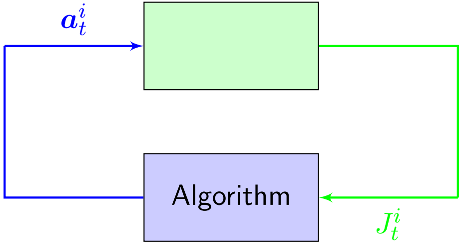
$$\left.\begin{array}{l} \blacktriangleright\ \|\boldsymbol{\mu}_t - \boldsymbol{\mu}^*\|^2 \text{ converges as } t \to \infty \\ \blacktriangleright\ \sum_{t=0}^{\infty} \beta_t \boldsymbol{M}(\boldsymbol{\mu}_t)^T(\boldsymbol{\mu}_t - \boldsymbol{\mu}^*) < \infty \end{array}\right\} \mu_t \to \mu^*$$

Summary

Convex game, zero-order information: $J_t^i = J^i(\boldsymbol{a}_t^i, \boldsymbol{a}_t^{-i})$

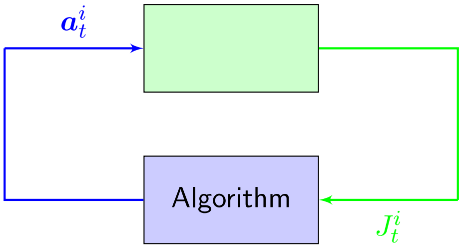
- player i: one-point estimation of her gradient
- $\boldsymbol{a}_t = (\boldsymbol{a}_t^1, \ldots, \boldsymbol{a}_t^N)$ convergence to $\boldsymbol{a}^* \in \boldsymbol{A}$

$$J^i(\boldsymbol{a}^{*i}, \boldsymbol{a}^{*-i}) \leq J^i(\boldsymbol{a}^i, \boldsymbol{a}^{*-i}), \quad \forall \boldsymbol{a}^i \in A^i$$

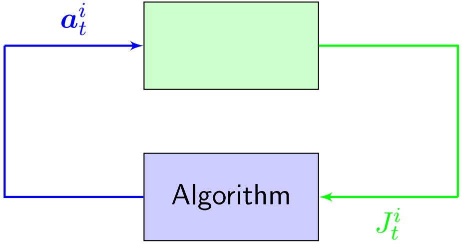


Outline

Learning in games as a bandit optimization problem



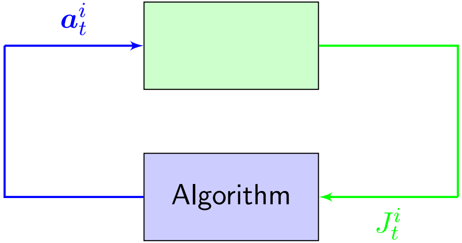
Learning in games as a bandit optimization problem



Regret: $R(T) = \sum_{t=0}^{T} J_t^i(\boldsymbol{a}_t^i) - \sum_{t=0}^{T} J_t^i(\boldsymbol{a}^i)$

- $\boldsymbol{a}_t^i$ played action, $J_t^i(\boldsymbol{a}_t^i) = J^i(\boldsymbol{a}_t^i, \boldsymbol{a}_t^{-i})$
- $\boldsymbol{a}^i$ best action in hindsight: $\min_{\tilde{a}^i \in A^i} \sum_{t=0}^{T} J_t^i(\tilde{a}^i)$

Learning in games as a bandit optimization problem



Regret: $R(T) = \sum_{t=0}^{T} J_t^i(\boldsymbol{a}_t^i) - \sum_{t=0}^{T} J_t^i(\boldsymbol{a}^i)$

- $\boldsymbol{a}_t^i$ played action, $J_t^i(\boldsymbol{a}_t^i) = J^i(\boldsymbol{a}_t^i, \boldsymbol{a}_t^{-i})$
- $\boldsymbol{a}^i$ best action in hindsight: $\min_{\tilde{a}^i \in A^i} \sum_{t=0}^{T} J_t^i(\tilde{a}^i)$

No-regret algorithm: $R(T) = o(T)$ as $T \to \infty$

[Flaxman et al. 2005], [Shamir 2013], [Bubeck 2016], ...

No-regret learning and convex games

Finite action games: each player adopts a no-regret algorithm

- $\frac{1}{T}\sum_{t=0}^{T} \boldsymbol{a}_t^i \rightarrow$ *coarse-correlated equilibrium*

No-regret learning and convex games

Finite action games: each player adopts a no-regret algorithm

- $\frac{1}{T}\sum_{t=0}^{T} \boldsymbol{a}_t^i \rightarrow$ *coarse-correlated equilibrium*

Convex games: our algorithm is no-regret [TT, MK 2018]

$$\boldsymbol{\mu}_{t+1}^i = \mathsf{Proj}_{A^i}\big[\boldsymbol{\mu}_t^i - \beta_t J_t^i \frac{\boldsymbol{a}_t^i - \boldsymbol{\mu}_t^i}{\sigma_t^2}\big]$$

- $\boldsymbol{a}_T^i \rightarrow \boldsymbol{a}^*$, $\boldsymbol{a}^*$: Nash equilibrium

No-regret learning and convex games

Finite action games: each player adopts a no-regret algorithm

- $\frac{1}{T}\sum_{t=0}^{T} \boldsymbol{a}_t^i \to$ *coarse-correlated equilibrium*

Convex games: our algorithm is no-regret [TT, MK 2018]

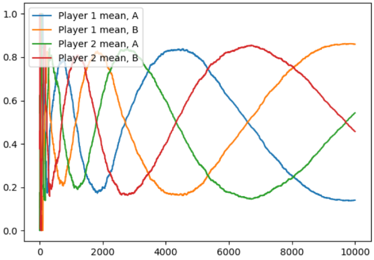
$$\boldsymbol{\mu}_{t+1}^i = \mathsf{Proj}_{A^i}\left[\boldsymbol{\mu}_t^i - \beta_t J_t^i \frac{\boldsymbol{a}_t^i - \boldsymbol{\mu}_t^i}{\sigma_t^2}\right]$$

- $\boldsymbol{a}_T^i \to \boldsymbol{a}^*$, $\boldsymbol{a}^*$: Nash equilibrium
- *under strict monotonicity of the game map*

Convex games versus optimization: zero-sum games

$$\begin{array}{cc} & \begin{array}{cc} \text{head} & \text{tail} \end{array} \\ \begin{array}{c} \text{head} \\ \text{tail} \end{array} & \begin{bmatrix} (1,-1) & (-1,1) \\ (-1,1) & (1,-1) \end{bmatrix} \end{array}$$

- Game mapping is not strictly monotone: $\big(M(\boldsymbol{a}) - M(\boldsymbol{a}')\big)^T(\boldsymbol{a}-\boldsymbol{a}') = 0$, $\forall \boldsymbol{a}, \boldsymbol{a}'$
- Our algorithm does not converge



matching pennies - [N. Kwan, USRA 2020]

Implications of non-strictly monotone game mapping

- All follow-the-regularized-leader algorithms (no-regret) diverge

  [Mertikopoulos et. al. 2018], [Bailey, 2020]

- Hamiltonian system interpretations [Balduzzi et al. 2018]

$$JM(\boldsymbol{a}) = \underbrace{P(\boldsymbol{a})}_{\text{symmetric}} + \underbrace{H(\boldsymbol{a})}_{\text{assymetric}}$$

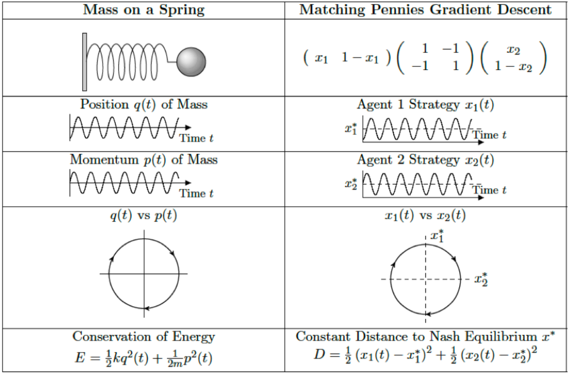
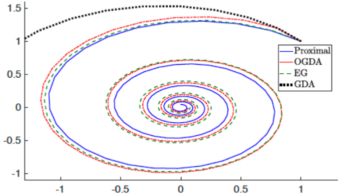


Figure – [Bailey & Piliouras 2019]

Zero-sum games beyond matching pennies

$\min_x \max_d f(x, d)$: robust optimization, robust control, training generative adversarial networks

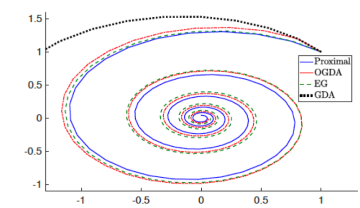
- algorithms for monotone VIs [Tseng 1995], [Facchinei, Pang 2007]
- extra gradient, optimistic mirror descent [Mokhtari et al. 2019]



Zero-sum games beyond matching pennies

$\min_x \max_d f(x, d)$: robust optimization, robust control, training generative adversarial networks

- algorithms for monotone VIs [Tseng 1995], [Facchinei, Pang 2007]
- extra gradient, optimistic mirror descent [Mokhtari et al. 2019]



- Limitation in our setup: $J_t^i = J^i(\boldsymbol{a}_t^i, \boldsymbol{a}_t^{-i})$
  - no (extra) gradients
  - no implicit algorithms

Bandit learning in non-strictly monotone games

Monotone game map: $\big(M(\boldsymbol{a}) - M(\boldsymbol{a}')\big)^T(\boldsymbol{a} - \boldsymbol{a}') \geq 0 \; \forall \boldsymbol{a}, \boldsymbol{a}' \in \boldsymbol{A}$

- single time-scale regularization

$$\boldsymbol{\mu}^i_{t+1} = \mathsf{Proj}_{A^i}\big(\boldsymbol{\mu}^i_t - \beta_t J^i_t \frac{\boldsymbol{a}^i_t - \boldsymbol{\mu}^i_t}{\sigma^2_t} + \epsilon_t \boldsymbol{\mu}^i_t\big)$$

- regularized cost: $J^i(\boldsymbol{a}) + \frac{\epsilon_t}{2}\|\boldsymbol{a}^i\|^2_2$

Convergence result

Assumptions

- $\boldsymbol{M} : \mathbb{R}^{Nd} \to \mathbb{R}^{Nd}$ is montone and Liptschitz

Convergence result

Assumptions

- $\boldsymbol{M} : \mathbb{R}^{Nd} \to \mathbb{R}^{Nd}$ is montone and Liptschitz

Theorem [TT, MK 2019]

Choose $\beta_t, \sigma_t, \epsilon_t \to 0$ such that

$$\sum_{t=0}^{\infty} \beta_t = \infty, \ \sum_{t=0}^{\infty} \beta_t \sigma_t < \infty, \ \sum_{t=0}^{\infty} \frac{\beta_t^2}{\sigma_t^2} < \infty,$$
$$\sum_{t=0}^{\infty} \frac{(\epsilon_{t-1} - \epsilon_t)^2}{\beta_t \epsilon_t^3} < \infty, \ \sum_{t=0}^{\infty} \beta_t \epsilon_t = \infty.$$

Then,

- state $\boldsymbol{\mu}_t$ converges almost surely to a Nash equilibrium $\boldsymbol{\mu}^*$
- action $\boldsymbol{a}_t$ converges in probability to $\boldsymbol{\mu}^*$

Proof sketch

Define $\boldsymbol{y}_t$ as solution of VI$(\boldsymbol{M}(\boldsymbol{a}) + \epsilon_t \boldsymbol{a}, \mathbf{A})$

- converges to a solution of VI$(\boldsymbol{M}(\boldsymbol{a}), \mathbf{A})$ [Facchinei, Pang 2007]

Show $\|\boldsymbol{\mu}_t - \boldsymbol{y}_t\|^2$ sufficiently decreases at each iteration

$$\mathrm{E}\{\|\boldsymbol{\mu}_{t+1} - \boldsymbol{y}_{t+1}\|^2\} \leq (1 - \epsilon_t \beta_t)\|\boldsymbol{\mu}_t - \boldsymbol{y}_t\|^2 + \xi_t$$

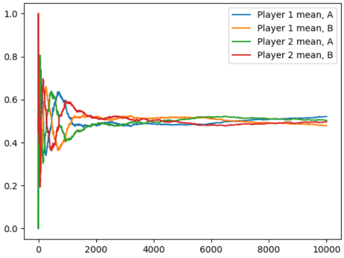
- $\|\boldsymbol{y}_t - \boldsymbol{y}_{t-1}\|_2^2 = O(\frac{|\epsilon_t - \epsilon_{t-1}|^2}{\epsilon_t^2}) \Rightarrow \xi_t = O(\beta_t \sigma_t + \frac{\beta_t^2}{\sigma_t^2} + \frac{|\epsilon_{t-1} - \epsilon_t|^2}{\beta_t \epsilon_t^3})$
- $\|\boldsymbol{\mu}_t - \boldsymbol{y}_t\|$ goes to zero almost surely

Learning in matching pennies

Choose $\beta_t = \frac{1}{t^p}, \sigma_t = \frac{1}{t^q}, \epsilon_t = \frac{1}{t^l}$, $p, q, l > 0$ such that

$$\sum_{t=0}^{\infty} \beta_t = \infty, \ \sum_{t=0}^{\infty} \beta_t \sigma_t < \infty, \ \sum_{t=0}^{\infty} \frac{\beta_t^2}{\sigma_t^2} < \infty,$$

$$\sum_{t=0}^{\infty} \frac{(\epsilon_{t-1} - \epsilon_t)^2}{\beta_t \epsilon_t^3} < \infty, \ \sum_{t=0}^{\infty} \beta_t \epsilon_t = \infty$$



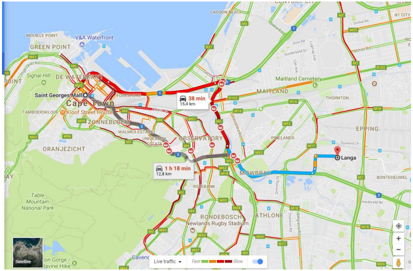
matching pennies - [N. Kwan, USRA 2020]

Extension - games with coupling constraints

- sharing limited capacity resources
  - transmission lines, roads, bandwidth
- convex coupling constraint $\mathbf{g} : \mathbb{R}^{Nd} \to \mathbb{R}^m$

$$C := \{\boldsymbol{a} \in \mathbb{R}^{Nd} \mid \mathbf{g}(\boldsymbol{a}) \leq \mathbf{0}\}$$

- jointly convex game $\Gamma(N, \boldsymbol{A} \cap C, \{J^i\})$



Challenges due to coupled action spaces

Generalized Nash equilibria (GNE) for $\Gamma(N, \boldsymbol{A} \cap C, \{J^i\})$

For each player i

$$J^i(\boldsymbol{a}^{*i}, \boldsymbol{a}^{*-i}) \leq J^i(\boldsymbol{a}^i, \boldsymbol{a}^{*-i}), \ \forall \boldsymbol{a}^i \in \{\boldsymbol{a}^i \in A^i \,|\, \mathbf{g}(\boldsymbol{a}^i, \boldsymbol{a}^{*-i}) \leq 0\}$$

- uniqueness and computation [Rosen 1965], [Facchinei, Pang, Kanzow 2009-2010]

Variational equilibria $\subset$ GNE

If $\boldsymbol{a}^* \in \boldsymbol{A} \cap C$ satisfies $\boldsymbol{M}(\boldsymbol{a}^*)^T(\boldsymbol{a} - \boldsymbol{a}^*) \geq 0, \ \forall \boldsymbol{a} \in \boldsymbol{A} \cap C$
Then $\boldsymbol{a}^*$ is a GNE [Facchinei and Pang, 2009]

Decoupling the constraints for distributed computation

Associate a player to the coupling constraint $\mathbf{g} : \mathbb{R}^{Nd} \to \mathbb{R}^m$

- a new game with an additional fictitious player, $\boldsymbol{\lambda} \in \mathbb{R}^m_{\geq 0}$

$$\bar{\Gamma}(N+1, \{\{A^i\}_{i=1,\dots,N}, \mathbb{R}^n_{\geq 0}\}, \{\bar{J}^i\})$$

- cost functions in extended game $\bar{\Gamma}$

$$\begin{aligned} \bar{J}^i(\boldsymbol{a}^i, \boldsymbol{a}^{-i}, \boldsymbol{\lambda}) &= J^i(\boldsymbol{a}^i, \boldsymbol{a}^{-i}) + \boldsymbol{\lambda}^T \mathbf{g}(\boldsymbol{a}^i, \boldsymbol{a}^{-i}), \quad i = 1, \dots, N \\ \bar{J}_{N+1}(\boldsymbol{a}, \boldsymbol{\lambda}) &= -\boldsymbol{\lambda}^T \mathbf{g}(\boldsymbol{a}) \end{aligned}$$

$[\boldsymbol{a}^*, \boldsymbol{\lambda}^*]$ Nash equilibrium in $\bar{\Gamma} \Rightarrow \boldsymbol{a}^*$ variational equilibrium in Γ

Non-monotonicity of the game mapping

Example: quadratic cost and affine coupling constraint

- $J^i(\boldsymbol{a}) = \frac{1}{2}\boldsymbol{a}^T H^i \boldsymbol{a}$, $i = 1, \ldots, N$
- $\mathbf{g}(\boldsymbol{a}) = F\boldsymbol{a} + f$, $F : \mathbb{R}^{Nd} \to \mathbb{R}^m$

Non-monotonicity of the game mapping

Example: quadratic cost and affine coupling constraint

- $J^i(\boldsymbol{a}) = \frac{1}{2}\boldsymbol{a}^T H^i \boldsymbol{a}$, $i = 1, \ldots, N$
- $\mathbf{g}(\boldsymbol{a}) = F\boldsymbol{a} + f$, $F : \mathbb{R}^{Nd} \to \mathbb{R}^m$

$$\bar{\boldsymbol{M}}(\boldsymbol{a}, \boldsymbol{\lambda}) = \begin{bmatrix} H & F^T \\ -F & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{\lambda} \end{bmatrix}$$

Zero-order learning in games with coupling constraints

Zero-order information: $\bar{J}_t^i = J^i(\boldsymbol{a}_t) + \boldsymbol{\lambda}_t \mathbf{g}(\boldsymbol{a}_t)$

$$\boldsymbol{a}_t^i \sim \mathcal{N}(\boldsymbol{\mu}_t^i, \sigma_t^2 I)$$
$$\boldsymbol{\mu}_{t+1}^i = \mathsf{Proj}_{A^i}\left[\boldsymbol{\mu}_t^i - \beta_t \bar{J}_t^i \frac{\boldsymbol{a}_t^i - \boldsymbol{\mu}_t^i}{\sigma_t^2}\right]$$
$$\boldsymbol{\lambda}_{t+1} = \mathsf{Proj}_{\mathbb{R}^n_{\geq 0}}[\boldsymbol{\lambda}_t + \beta_t \mathbf{g}(\boldsymbol{a}_t)]$$

Theorem

- Assume $\boldsymbol{M}(\boldsymbol{a})$ is symmetric and strictly monotone
- Choose β_t, σ_t as in the strictly monotone case

$\boldsymbol{\mu}_t$ converges almost surely to the variational equilibrium.

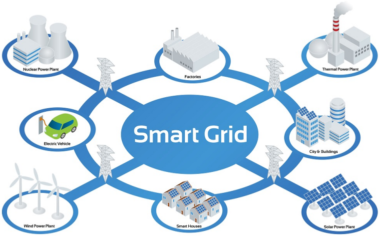
Example - Cournot game in electricity markets

Consumers minimizing their electricity bills

- consumption profile over d periods $\boldsymbol{a}^i = [a_1^i, \ldots, a_d^i]^\top \in \mathbb{R}^d$
- local consumption bounds

$$0 \le a_k^i \le \bar{a}_k^i,\ k = 1, \ldots, d, \qquad \sum_{k=1}^{d} a_k^i = \bar{a}^i$$

- network capacity constraint $\sum_{i=1}^{N} a_k^i \le \bar{a}_k,\ k = 1, \ldots, d$

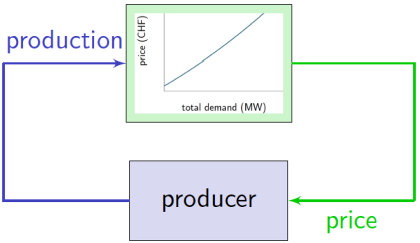


Convex game formulation

- electricity price $\mathbf{p}(\boldsymbol{a})$
- player i's cost function

$$J^i(\boldsymbol{a}^i, \boldsymbol{a}^{-i}) = P^i(\boldsymbol{a}^i) + \mathbf{p}(\boldsymbol{a})\boldsymbol{a}^i$$

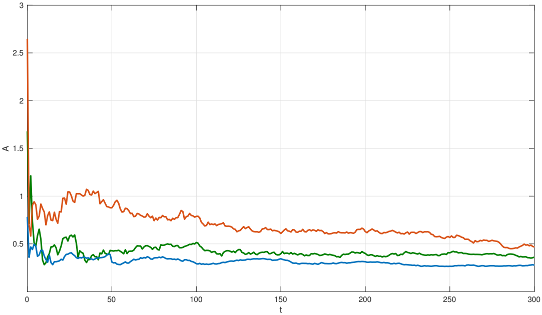
- P^i convex quadratic, $\mathbf{p}$ linear
  - convex game with strictly convex potential function
  - learning optimal consumption profile using payoff information



Simulation result

Relative error $\frac{\|\boldsymbol{\mu}_t - \boldsymbol{a}^*\|}{\|\boldsymbol{a}^*\|}$

- fast initial decrease, very slow convergence
- lower bounds on convergence rates?



Colors blue, green, red corresponding to $N = 3, 10, 30$

Outline

Introduction

Learning in convex games - setup & algorithm

Learning in games - connections & extensions

Conclusions

Summary

Learning in convex games

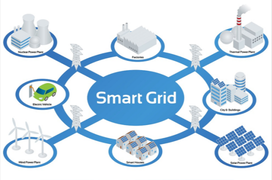
- Nash equilibria solve a variational inequality problem
- learn Nash equilibria using zero-order information

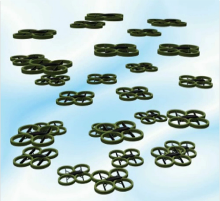
Proposed algorithm

- bandit feedback: no knowledge of the cost functions
- convergence to Nash equilibrium under monotonicity

Outlook

- Connections of no-regret learning and convex games
- Exploring lower bounds for convergence rate
- Learning in non-convex games
- Learning in dynamic and feedback games






Thank you for your time and attention!

Learning players' cost functions via Gaussian processes

1. P. G. Sessa, I. Bogunovic, A. Krause and M. Kamgarpour, Contextual Games: Multi-Agent Learning with Side Information, NeurIPS 2020
2. P. G. Sessa, I. Bogunovic, M. Kamgarpour and A. Krause, No-Regret Learning in Unknown Games with Correlated Payoffs, NeurIPS 2019
3. P. G. Sessa, I. Bogunovic, M. Kamgarpour and A. Krause, Mixed Strategies for Robust Optimization of Unknown Objectives, AISTATS 2020

Efficiency of Nash equilibria, mechanism design, applications

1. O. Karaca and M. Kamgarpour, Designing Coalition-Proof Reverse Auctions over Continuous Goods, IEEE Transactions on Automatic Control, 2019
2. P. G. Sessa, M. Kamgarpour, A. Krause, Bounding Inefficiency of Equilibria in Continuous actions games using submodularity and curvature, AISTATS 2019
3. O. Karaca*, P. G. Sessa*, A. Leidi and M. Kamgarpour, No-regret Learning from Partially Observed Data in Repeated Auctions, IFAC 2019, *: equal contribution
4. B. Shea, M. Schmidt and M. Kamgarpour, A Multiagent Model of Efficient and Sustainable Financial Markets, NeurIPS 2020 Workshop

Convergence of random variables

Robbins and Siegmund on non-negative random variables

Theorem

(Ω, F, P): probability space, $F_1 \subset F_2 \subset \dots$ sub-σ-algebras of F, $z_t, b_t, \xi_t,$ and ζ_t be non-negative F_t-measurable random variables with

$$\mathrm{E}(z_{t+1}|F_t) \leq z_t(1 + b_t) + \xi_t - \zeta_t.$$

- *almost surely* $\lim_{t\to\infty} z_t$ *exists and is finite*
- $\sum_{t=1}^{\infty} \zeta_t < \infty$ *almost surely on* $\{\sum_{t=1}^{\infty} b_t < \infty,\ \sum_{t=1}^{\infty} \xi_t < \infty\}$