Game-Theoretic Decision Making in Multi-Agent Systems under Constraints and Welfare Objectives

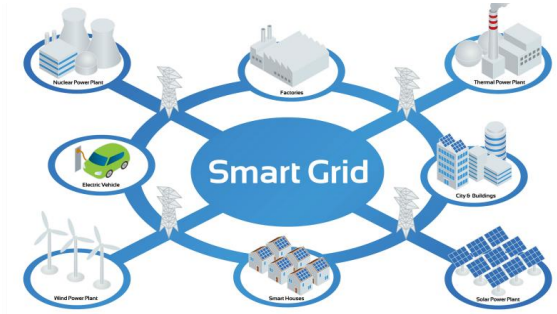
Anna Maria Maddux

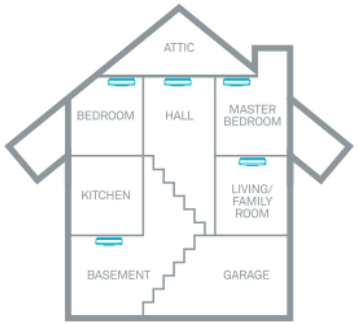
Ph.D. defense, 19. March 2026, EPFL

EPFL sycamore National Centre of Competence in Research Automation Swiss National Science Foundation

Multi-agent systems

Multiple self-interested agents that interact





Essential: Agents are coupled in their objectives

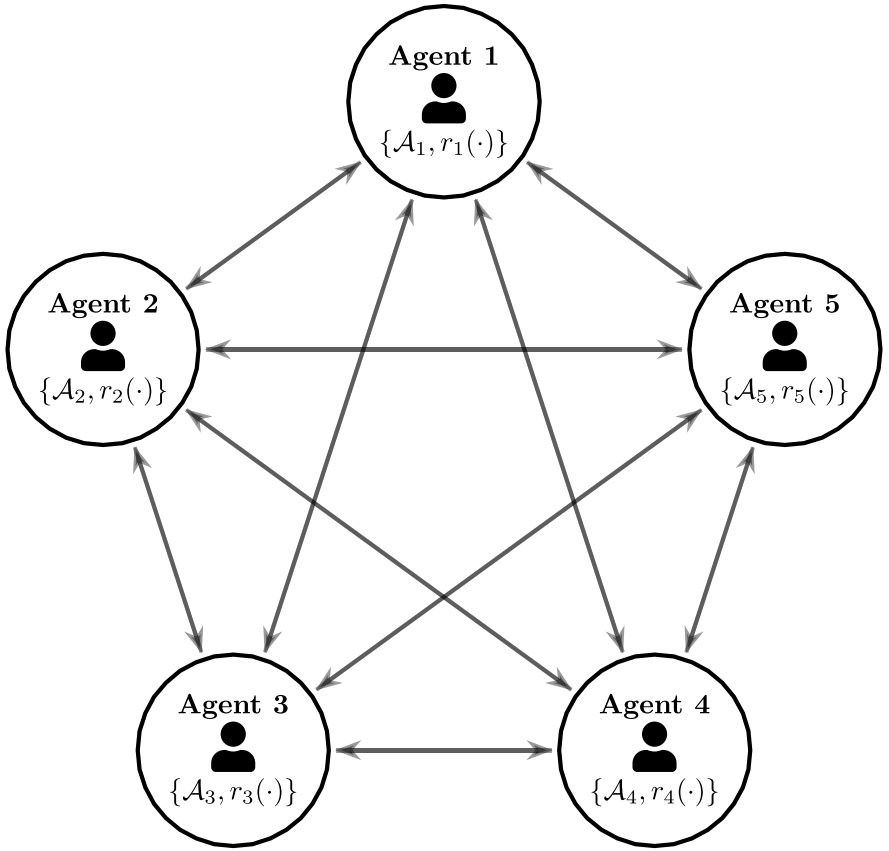
Multi-agent systems as games

Game theory: natural framework for analyzing interactions

Definiton of a **game**:

$$\Gamma = \{\mathcal{N}, \{\mathcal{A}_i\}_{i\in\mathcal{N}}, \{r_i\}_{i\in\mathcal{N}}\}$$

→ Characterize **stable outcomes** of multi-agent systems



Multi-agent systems with constraints

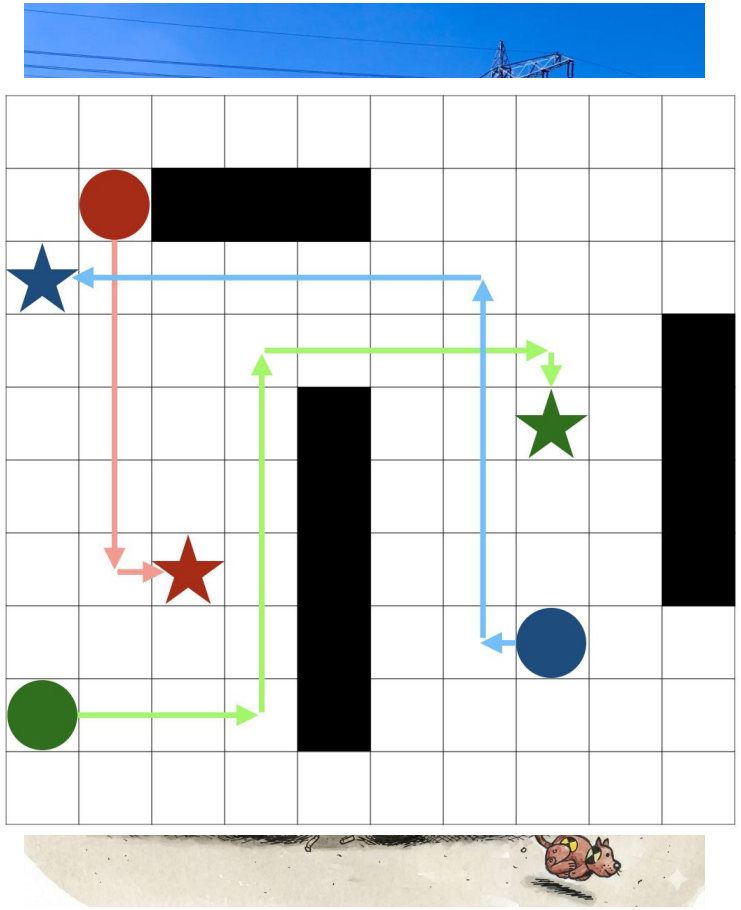
Central to multi-agent systems: **Constraints**

→ Encode physical limitations, safety or regulartory requirements, or inter-agent dependencies

→ Ignoring constraints in multi-agent decision making results in **physical failures and safety risks**

→ Action space $\mathcal{A}_i$ is restricted to feasible actions

How to ensure **outcome** is not only stable but **feasible**?



Multi-agent systems with welfare objectives

Central to multi-agent systems: **Inefficiency**

→ Self-interested decision-making can lead to **inefficient outcomes** (congestion, overusage, inequality, etc.)

Individual perspective:

→ Agent wants to achieve own goal:

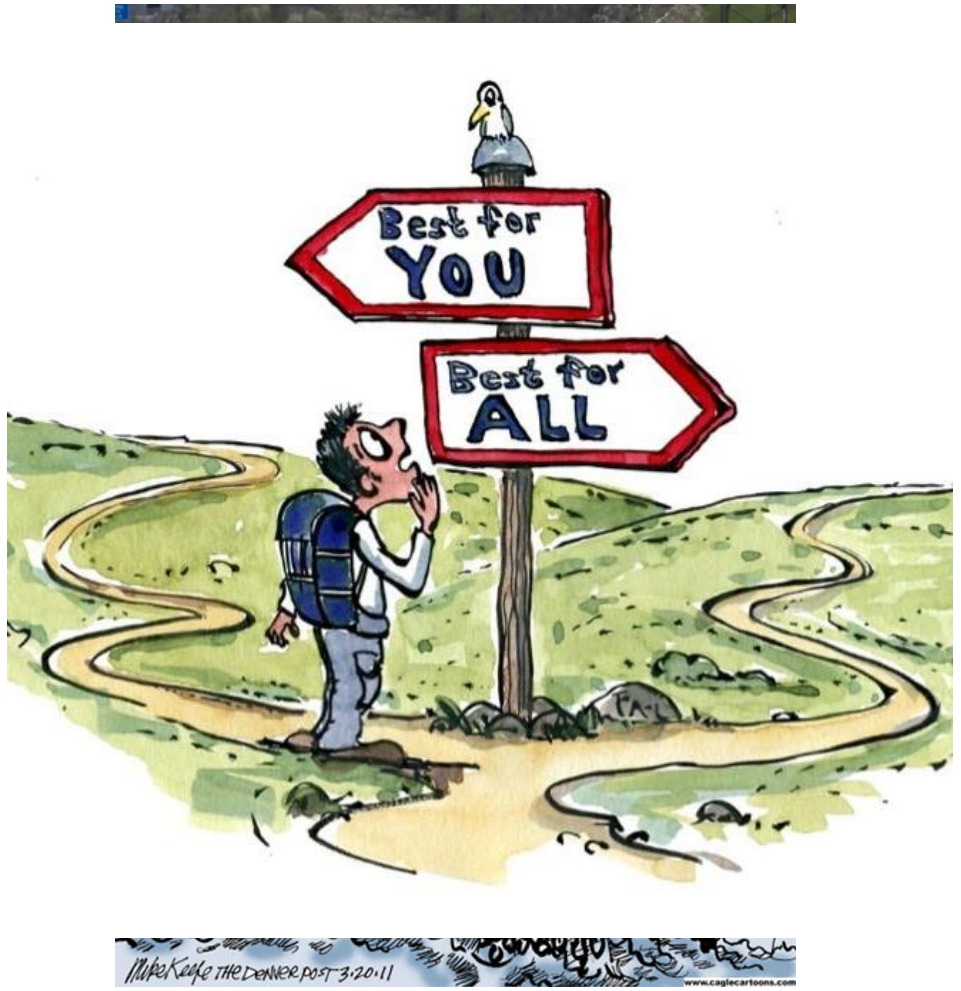
$$r_i : \mathcal{A} \to \mathbb{R}$$

Global perspective:

→ Central authorithy wants to maximize welfare objetive:

$$W : \mathcal{A} \to \mathbb{R}$$

How to ensure **outcome** is not only stable but **efficient**?



Ph.D. research overview

Part I:
Multi-agent systems with constraints

- **Chapter 3:** Existence of constrained equilibria in games with coupling constraints
  [Under review at Automatica]
- **Chapter 4:** Learning constrained equilibria in games with private constraints
  [AISTATS 2024]

Part II:
Multi-agent systems with welfare objectives

- **Chapter 5:** Steering to efficient outcomes via learning in Stackelberg games
  [CDC 2025]
- **Chapter 6:** Learning efficient equilibria in potential games via log-linear learning
  [TCNS 2026]

Other:
- Eliciting truthful feedback for preference-based learning *[AISTATS 2026]*
- Reinforcement learning for scaffold-free construction *[SCF 2023]*
- Preference-based distributed welfare maximization *[Under review at ICML]*

Multi-agent systems with constraints

$$\Gamma = \{\mathcal{N}, \{\mathcal{A}_i\}_{i\epsilon\mathcal{N}}, \{r_i\}_{i\epsilon\mathcal{N}}, \{g_i\}_{i\epsilon\mathcal{N}}\}$$

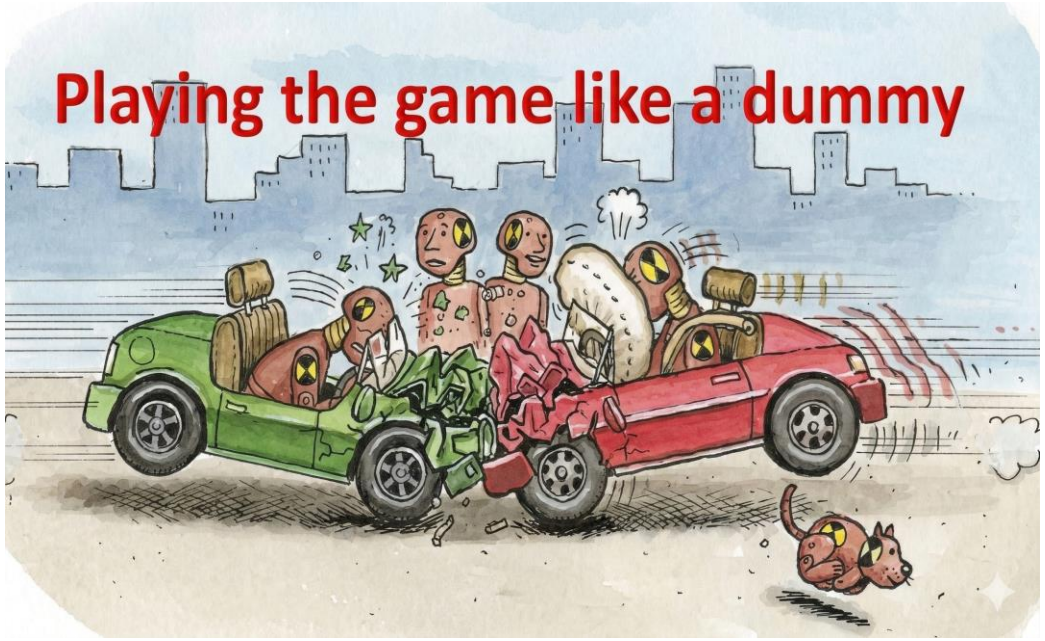
→ Games are subject to constraints $g_i(\cdot)$

→ Action $a\epsilon\mathcal{A}$ is feasible for agent i if $g_i(a) \leq 0$

How to ensure **outcome** is not only stable but **feasible**?

Existence conditions for constrained equilibria

Learning constrained equilibria



Chapter 3

Existence of constrained equilibria in games with coupling constraints

Games with constraints

$$\Gamma = \{\mathcal{N}, \{\mathcal{A}_i\}_{i\epsilon\mathcal{N}}, \{r_i\}_{i\epsilon\mathcal{N}}, \{g_i\}_{i\epsilon\mathcal{N}}\}$$

- **Playerwise coupling** constraints:

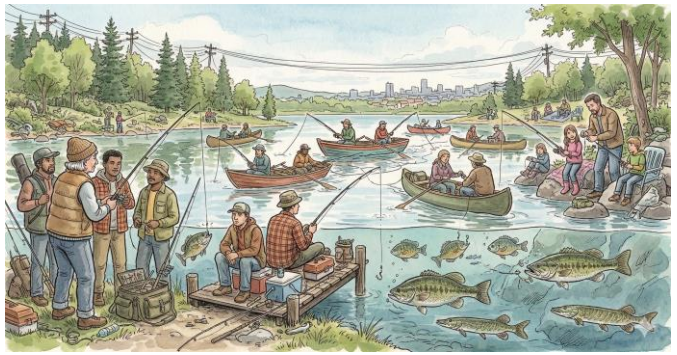
$$g_i : \mathcal{A} \to \mathbb{R} \quad \forall i\epsilon\mathcal{N}$$

- **Common coupling** constraints:

$$g : \mathcal{A} \to \mathbb{R} \quad \forall i\epsilon\mathcal{N}$$

- **Private** constraints:

$$g_i : \mathcal{A}_i \to \mathbb{R} \quad \forall i\epsilon\mathcal{N}$$



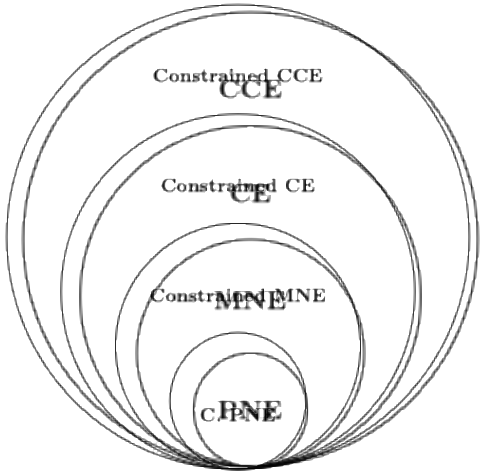
Existence of constrained correlated equilibrium

Definition

A distribution $\sigma \epsilon \Delta(\mathcal{A})$ is an equilibrium if:

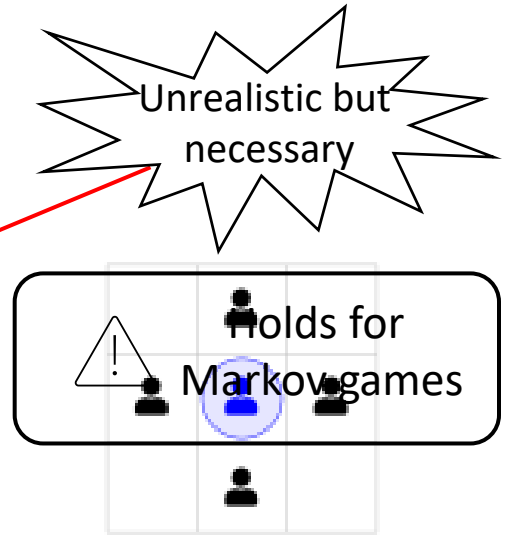
- For each agent i: $\mathbf{E}_{a \sim \sigma}[r_i(a)] \geq \max_{\phi_i \in \Phi_i} \mathbf{E}_{a \sim \phi_i \circ \sigma}[r_i(a)]$

→ Deviating is either **unprofitable or infeasible**



Existence of constrained correlated equilibrium?

- Games with playerwise coupling constraints: Strong Slater's condition [1]
- Games with common coupling constraints: Weak Slater's condition [2]



[1] E. Altman and A. Shwartz. "Constrained Markov games: Nash equilibria". In: Advances in Dynamic Games and Applications. Springer, 2000.
[2] T. Ni*, A. Maddux*, and M. Kamgarpour, "On characterization and existence of constrained correlated equilibria in Markov games". Under review: Automatica. 2025.

Chapter 4

Learning constrained equilibria in games with private constraints

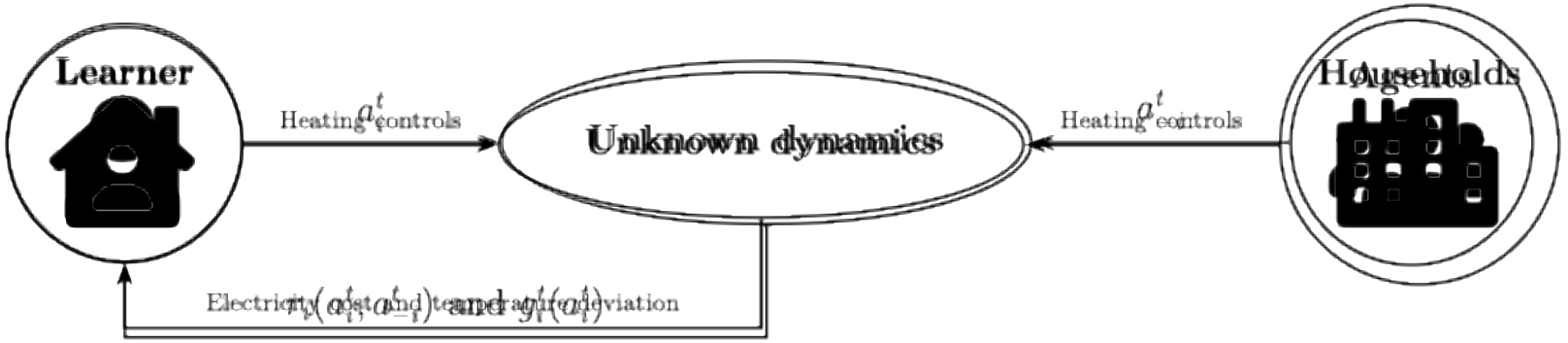
Repeated game with private constraints

Consider **game** $\Gamma = \{\mathcal{N}, \{\mathcal{A}_i\}_{i\epsilon\mathcal{N}}, \{r_i\}_{i\epsilon\mathcal{N}}, \{g_i\}_{i\epsilon\mathcal{N}}\}$ with **private constraints** $g_i: \mathcal{A}_i \to \mathbb{R}$

a priori unknown → feasible action set $\mathcal{A}_i^f = \{a_i \epsilon \mathcal{A}_i \mid g_i(a_i) \leq 0\}$ unknown

At round t:

1. Agent i selects parameters $\boldsymbol{a_i^t} = (\boldsymbol{K_i}, \boldsymbol{SP_i}, \boldsymbol{ST_i})$ of building's proportional controller
2. Agent i observes **electricity cost** and **temperature deviation** from the setpoint temperature



Constrained regret

Constrained regret of agent i:

$$R_i(T) = \max_{a_i \in \mathcal{A}_i'} \sum_{t=1}^{T} r_i(a_i, a_{-i}^t) - \sum_{t=1}^{T} r_i(a_i^t, a_{-i}^t)$$

Cumulative constraint violations of agent i:

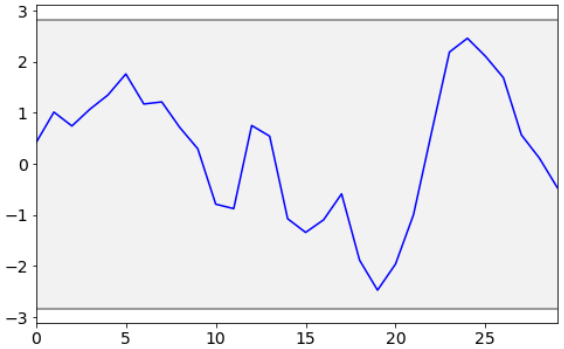
$$\mathcal{V}_i(T) = \sum_{t=1}^{T} [g_i(a_i^t)]_+$$

Goal of learning agent i:

- No regret $\frac{R_i(T)}{T} \to 0$ as $T \to \infty$
- No violation $\frac{\mathcal{V}_i(T)}{T} \to 0$ as $T \to \infty$

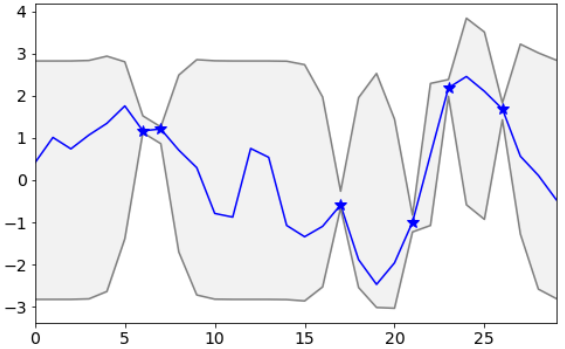
Estimates of the reward and constraint functions

1. Smoothness Assumption: $a \approx a' \Rightarrow r_i(a) \approx r_i(a')$
   $a_i \approx a_i' \Rightarrow g_i(a_i) \approx g_i(a_i')$
2. Feedback: agent i observes: other agents' action a_{-i}^t
   noisy reward $r_i(a^t) + \varepsilon_{r_i}^t$ and noisy constraint $g_i(a_i^t) + \varepsilon_{g_i}^t$



Smoothness assumption
+ feedback

Gaussian process regression



Confidence lemma [1]: With probability at least $1 - \delta$:

→ Reward and constraint functions estimates

[1] S.R. Chowdhury and A. Gopalan. "On kernelized multi-armed bandits". In: International Conference on Machine Learning. PMLR. 2017.

No-regret, no-violation algorithm

for $t = 1, \dots, T$ **do**

/ Check feasibility */*

if $\min_{a_i \in \mathcal{A}_i} \mathrm{LCB}^t_{g_i}(a_i) > 0$ **then**

Declare infeasibility

end if

/ Sample action from mixed strategy */*

Compute distribution p^t using $\mathcal{H}^{t-1}_{r_i} = \{(a^\tau_i, a^\tau_{-i}, r^\tau_i)\}_{\tau=1}^{t-1}$

Sample $a^t_i \sim \overline{p}^t$, where

$$\overline{p}^t(a_i) = \begin{cases} \frac{p^t(a_i)}{\sum_{\overline{a}_i \in \mathcal{A}_i} \overline{p}^t(\overline{a}_i)}, & \text{if } \mathrm{LCB}^t_{g_i}(a_i) \leq 0, \\ 0, & \text{otherwise.} \end{cases}$$

/ Update posterior mean and standard deviation */*

Observe $\tilde{r}^t_i = r_i(a^t) + \epsilon^t_{r_i}$, $\tilde{g}^t_i = g_i(a^t_i) + \epsilon^t_{g_i}$, and a^t_{-i}.

$\mathcal{H}^t_{r_i} = \mathcal{H}^{t-1}_{r_i} \cup \{(a^t, \tilde{r}^t_i)\}$ and $\mathcal{H}^t_{g_i} = \mathcal{H}^{t-1}_{g_i} \cup \{(a^t_i, \tilde{g}^t_i)\}$

Update $\mu^t_{r_i}(\cdot), \sigma^t_{r_i}(\cdot), \mu^t_{g_i}(\cdot)$, and $\sigma^t_{g_i}(\cdot)$

end for

At round t, action a_i is feasible if:

$$a_i \epsilon \mathcal{A}_i^{f,t} = \{a_i \epsilon \mathcal{A}_i \mid LCB^t_{g_i}(a_i) \leq 0\}$$

Repeated game with private constraints
≡
Sleeping expert problem [1]

→ Action a_i is feasible ≡ Expert a_i is awake

Set p^t as the **update rule of the sleeping expert algorithm** AdaNormalHedge [2]

[1] Y. Freund et al. “Using and Combining Predictors That Specialize”. In: Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing. Association for Computing Machinery, 1997.

[2] H. Luo and R.E. Schapire. “Achieving All with No Parameters: AdaNormalHedge”. In: Proceedings of The 28th Conference on Learning Theory. PMLR, 2015.

No-regret, no-violation algorithm - Guarantees

⚠ Holds for contextual Settings [1]

Theorem:

Let Assumptions hold and set $\beta_{r_i}^t$ and $\beta_{g_i}^t$ adequately. Then for $\delta\epsilon(0,1)$ with probability at least $1-\delta$ it holds:

- $R(T) = \mathcal{O}(\underbrace{\sqrt{T(\log(K)+\log(B))}}_{\text{Sleeping expert regret}} + \underbrace{\sqrt{T\log(2/\delta)}}_{\text{Probabilistic guarantees}} + \underbrace{\beta_{r_i}^T\sqrt{T\gamma_{r_i}^T}}_{\text{Estimate reward via GP framework}})$

- $V_{g_i}(T) = \mathcal{O}(\underbrace{\beta_{g_i}^T\sqrt{T\gamma_{g_i}^T}}_{\text{Estimate constraint via GP framework}})$

Empirical joint distribution: $\sigma^T(a) = \frac{1}{T}|\{t \in [T] \mid a^t = a\}|$

Proposition:

After T rounds, σ^T is an ε-constrained CCE with $\varepsilon \leq \max\{max_{i\epsilon\mathcal{N}}\frac{R_i(T)}{T}, max_{i\epsilon\mathcal{N}}\frac{V_i(T)}{T}\}$ [1]

} No-regret,no-violation ⟺ learning constrained CCE

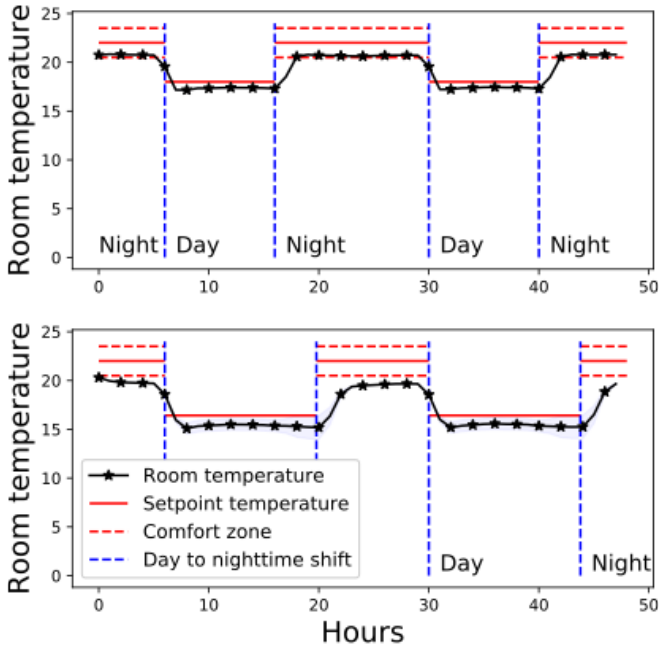
[1] A. Maddux and M. Kamgarpour, "Multi-player Learning in Contextual Games under Unknown Constraints". In: Proceedings of the 27th International Conference on Artificial Intelligence and Statistics (AISTATS). 2024.

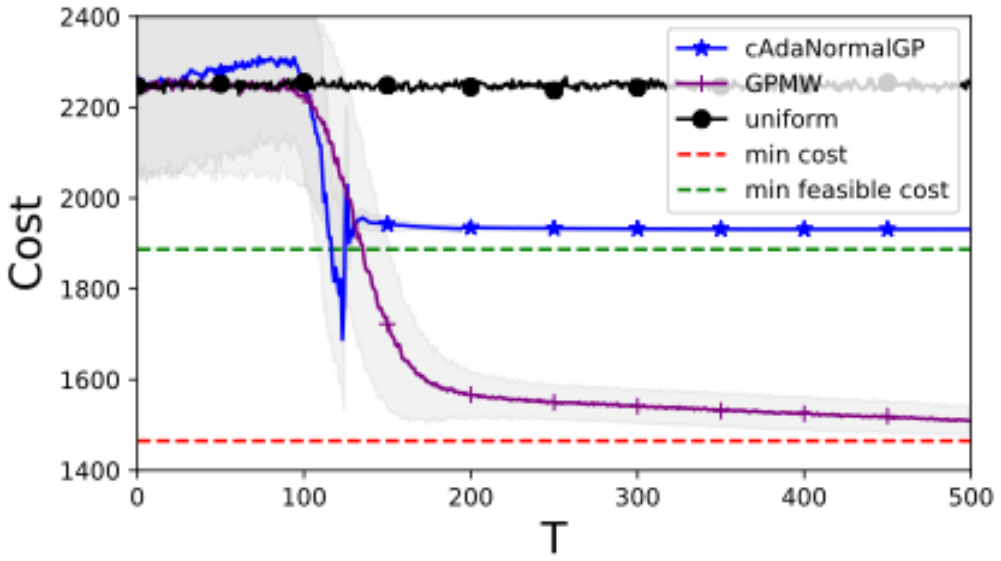
Multi-building temperature controller design

- Each building's heatpump is controlled by a P controller
- **Problem:**

$$min_{a_i} \quad \text{energy cost } J_i(a_i, a_{-i}) \text{ s.t. temperature deviation } g_i(a_i) \leq threshold,$$

where $a_i = (K_i, SP_i, ST_i) = (controller\ gain, setpoint\ temperature, switching\ times)$





Ph.D. research overview

Part I:
Multi-agent systems with constraints

- **Chapter 3:** Existence of constrained equilibria in games with coupling constraints
  [Under review at Automatica]
- **Chapter 4:** Learning constrained equilibria in games with private constraints
  [AISTATS 2024]

Part II:
Multi-agent systems with welfare objectives

- **Chapter 5:** Steering to efficient outcomes via learning in Stackelberg games
  [CDC 2025]
- **Chapter 6:** Learning efficient equilibria in potential games via log-linear learning
  [TCNS 2026]

Other:
- Eliciting truthful feedback for preference-based learning *[AISTATS 2026]*
- Reinforcement learning for scaffold-free construction *[SCF 2023]*
- Preference-based distributed welfare maximization *[Under review at ICML]*

Multi-agent systems with welfare objectives

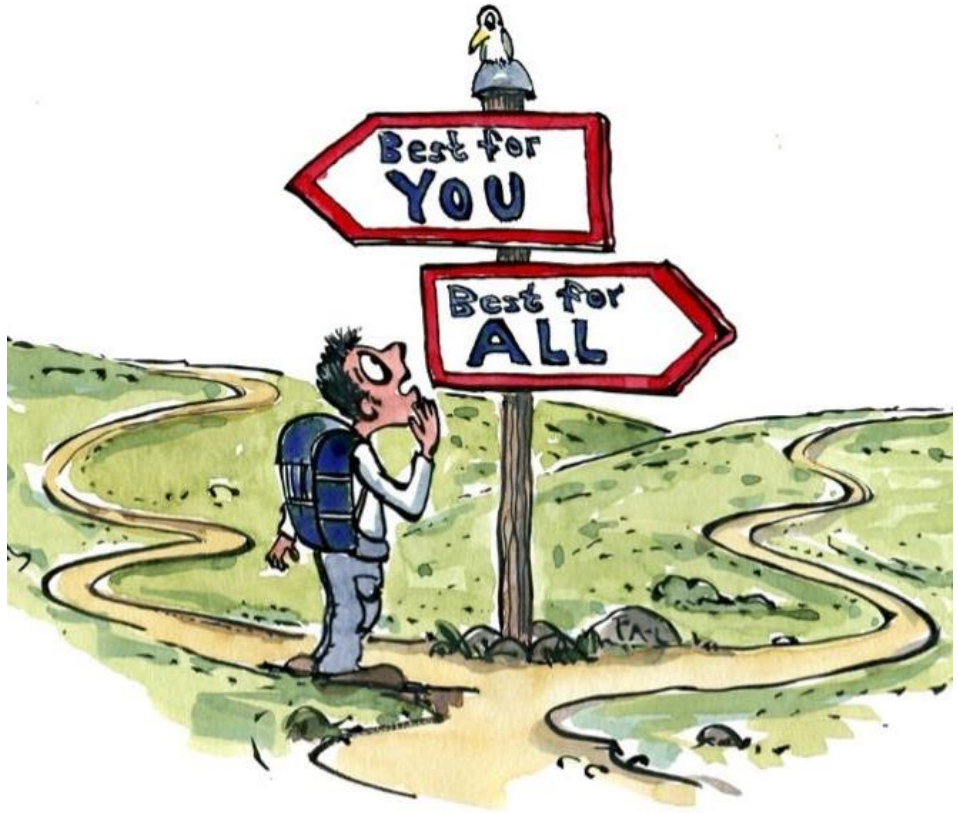
$$\Gamma = \{\mathcal{N}, \{\mathcal{A}_i\}_{i\epsilon\mathcal{N}}, \{r_i\}_{i\epsilon\mathcal{N}}, W\}$$

→ Games admit **multiple equilibria** with different values of welfare $W(\cdot)$

How to ensure **outcome** is not only stable but **efficient**?

Incentive design: Steering to efficient outcomes

Equilibrium selection: Learning efficient outcomes



Chapter 5

Steering to efficient outcomes via learning in Stackelberg games

Incentive design

Multi-agent system (e.g. ride-hailing markets) with:

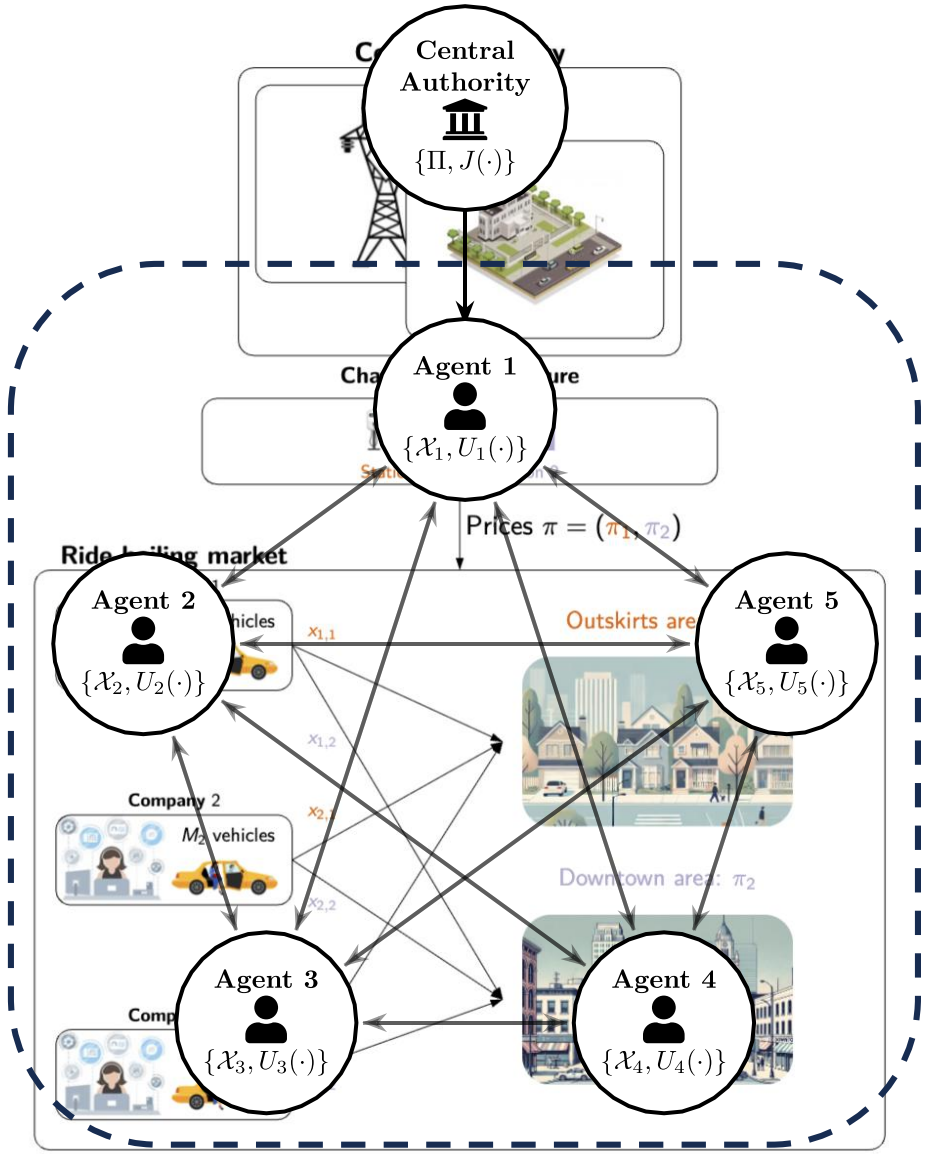
Agent (e.g. electric vehicle): maximize utility $U_i : \mathcal{X} \times \Pi \to \mathbb{R}$

→ Inefficient decision-making

Central authority (e.g. governement): introduces tolls, pricing schemes, subsidies etc.

→ Minimize social cost $J : \Pi \times \mathcal{X} \to \mathbb{R}$

Goal: Steer agents towards socially desirable behavior



Stackelberg game

1. Leader (Central authority): Set $\pi \in \Pi \subset \mathbb{R}^d$
2. Followers (Agents): Observe π
   Set $x_i \in \mathcal{X}_i \subseteq \mathbb{R}^p$

} Lower-level is a game $\Gamma(\pi) = \{N, \{\mathcal{X}_i\}_{i \in N}, \{U_i\}_{i \in N}; \pi\}$

$$\min_{\pi \in \Pi} \overbrace{J(\pi, x^*(\pi))}^{\text{Leader}}, \quad \text{s.t.} \quad \underbrace{x_i^*(\pi) \in \arg\max_{x_i \in \mathcal{X}_i} \overbrace{U_i(x_i, x_{-i}^*(\pi); \pi)}^{\text{Followers}}, \quad \forall i \in N}_{x^*(\pi) \text{ is a unique Nash equilibrium (NE) of } \Gamma(\pi)}$$ [1]

→ A stable outcome is the Stackelberg equilibrium $(\pi^*, x^*(\pi^*))$

[1] J B. Rosen. “Existence and uniqueness of equilibrium points for concave n-person games”. In: Econometrica: Journal of the Econometric Society (1965).

Challenges in Stackelberg games

Challenge: Leader does not know the agents' utilities $\{U_i\}_{i \in N}$

Remedy: Leader can observe its realized cost $J(\pi, \bar{x}(\pi))$

$$\min_{\pi \in \Pi} \underbrace{J(\pi, \bar{x}(\pi))}_{\text{Leader}} \quad \text{s.t.} \quad \underbrace{\bar{x}(\pi) \in \arg\min_{x \in X} J(\pi, x)}_{\text{Leader}}$$

Assumption: Followers respond $\bar{x}(\pi)$ s.t.

$$\mathbb{E}[\|\bar{x}(\pi) - x^*(\pi)\|^2] \leq \epsilon$$

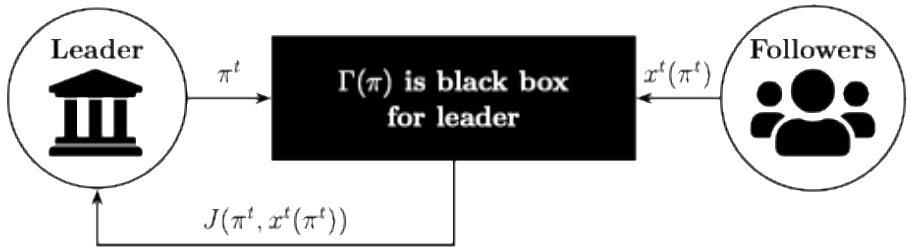
→ Direct computation of a Stackelberg equilibrium is impossible!

Can the leader learn an ϵ-Stackelberg equilibrium from observations of its realized cost?

Repeated Stackelberg games

At each round t:

1. Leader sets π^t
2. Followers respond with an approximate NE $x^t(\pi^t)$
3. Leader observes $J(\pi^t, x^t(\pi^t))$



Leader's regret:

$$R(T) = \sum_{t=1}^{T} J(\pi^t, x^t(\pi^t)) - \min_{\pi \in \Pi} J(\pi, x^*(\pi))$$

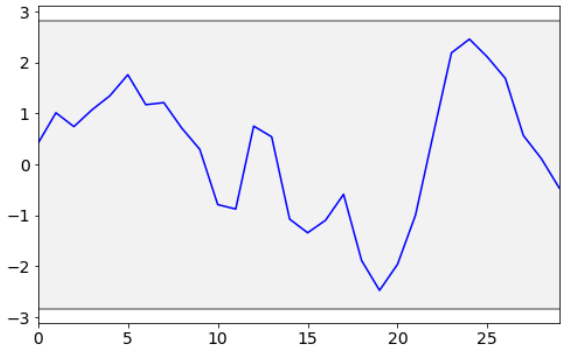
$$= \underbrace{\sum_{t=1}^{T} J(\pi^t, x^t(\pi^t)) - J(\pi^t, x^*(\pi^t))}_{1.} + \underbrace{\sum_{t=1}^{T} J(\pi^t, x^*(\pi^t)) - \min_{\pi \in \Pi} J(\pi, x^*(\pi))}_{2.}$$

→ **Leader's additional cost** beyond its optimal value due to:

1. Followers learning an approximation $x^t(\pi^t)$ of the unique NE $x^*(\pi^t)$
2. A priori not knowing $\pi^* \in \arg\min_{\pi \in \Pi} J(\pi, x^*(\pi))$

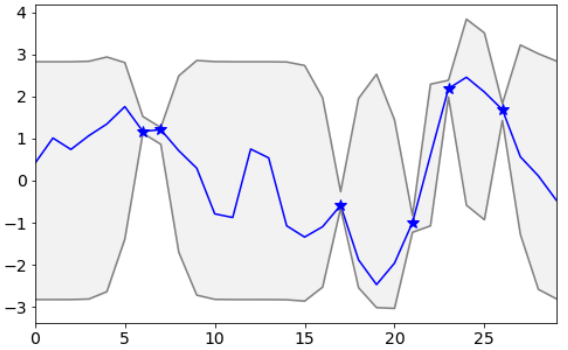
Confidence bounds for misspecified feedback

1. Smoothness assumption: $\pi \approx \bar{\pi} \quad \Rightarrow \quad J(\pi, x^*(\pi)) \approx J(\bar{\pi}, x^*(\bar{\pi}))$
2. Misspecified feedback: leader observes $J\left(\pi^t, x^t(\pi^t)\right)$ rather than $J(\pi^t, x^*(\pi^t))$



Smoothness assumption
+ misspecified feedback

Gaussian process regression



Confidence lemma [1]: With probability at least $1 - \delta$:

misspecification adjustment

$$\underbrace{\mu^{t-1}(\pi) - (\beta^t + L_J \epsilon \sqrt{t})\sigma^{t-1}(\pi)}_{\underline{J}^t(\pi)} \leq J(\pi) \leq \underbrace{\mu^{t-1}(\pi) + (\beta^t + L_J \epsilon \sqrt{t})\sigma^{t-1}(\pi)}_{\overline{J}^t(\pi)}$$

[1] I. Bogunovic and A. Krause. “Misspecified gaussian process bandit optimization”. In: Advances in Neural Information Processing Systems (NeurIPS) (2021).

Learning Stackelberg equilibria - Algorithm

1: **for** $t = 1, \ldots, T$ **do**

/ Followers learn approximate Nash equilibrium */*

2: **Inner loop:** Within K rounds, followers compute $\tilde{x}(\pi^t) = \text{ApproxNE}(\pi^t, K)$ s.t.

$$\mathbb{E}[\|x^*(\pi^t) - \tilde{x}(\pi^t)\|^2] \leq C(\pi^t)K^{-c}$$

3: Followers set $x^t(\pi^t) = \tilde{x}(\pi^t)$.

/ Leader updates its cost function estimate */*

4: Leader observes $J(\pi^t, x^t(\pi^t))$.

5: Leader updates μ^{t+1} and σ^{t+1} via GP framework.

6: Leader chooses and announces:

$$\pi^{t+1} \in \arg\min_{\pi \in \Pi} \underline{J}^{t+1}(\pi) = \mu^t(\pi) - \left(\beta^{t+1} + \frac{\epsilon\sqrt{t+1}}{\sigma}\right)\sigma^t(\pi)$$

7: **end for**

No-regret algorithm for the leader

→ No need for gradient-based techniques.

Learning Stackelberg equilibria - Guarantees

Theorem: [1]

Let Assumptions hold. For $\varepsilon, \delta \epsilon (0,1)$ and adequate β^t, the following holds:

1. With probability at least $1-\delta$:

$$R(T) \leq \mathcal{O}\Big(\underbrace{\big(\tfrac{1}{\delta}+\sqrt{\gamma^T}\big)TK^{-\frac{1}{2}}}_{\text{NE approximation}} + \underbrace{\beta^T\sqrt{\gamma^T T}}_{J \text{ unknown}}\Big)$$

2. Assume $U_i\colon \mathcal{X}\times\Pi \to \mathbb{R}$ is L_{U_i} −Lipschitz continuous in $x\epsilon\mathcal{X}$. If $K(\varepsilon) = \mathcal{O}\left(T^{\frac{1}{c}}\right)$, then with probaility at least $1-\delta$ there exists a $t^*\epsilon[T]$ such that:

$$\left(\pi^{t^*}, \mathrm{x}^{t^*}\left(\pi^{t^*}\right)\right) \text{ is an } \varepsilon\text{-Stackelberg equilibrium.}$$

[1] A. Maddux*, M. Maljkovic*, N. Geroliminis, and M. Kamgarpour, “No-Regret Learning in Stackelberg Games with an Application to Electric Ride-Hailing”. In: 64th IEEE Conference on Decision and Control (CDC). 2025.

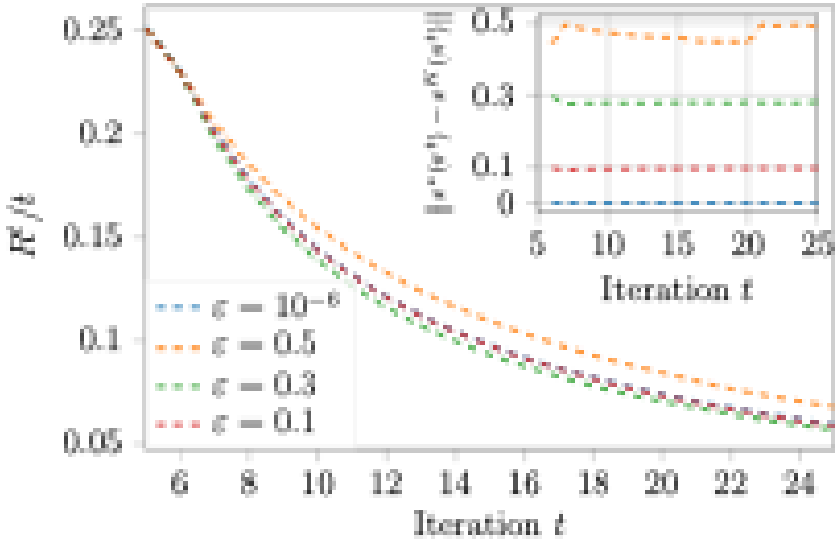
Learning in ride-hailing markets

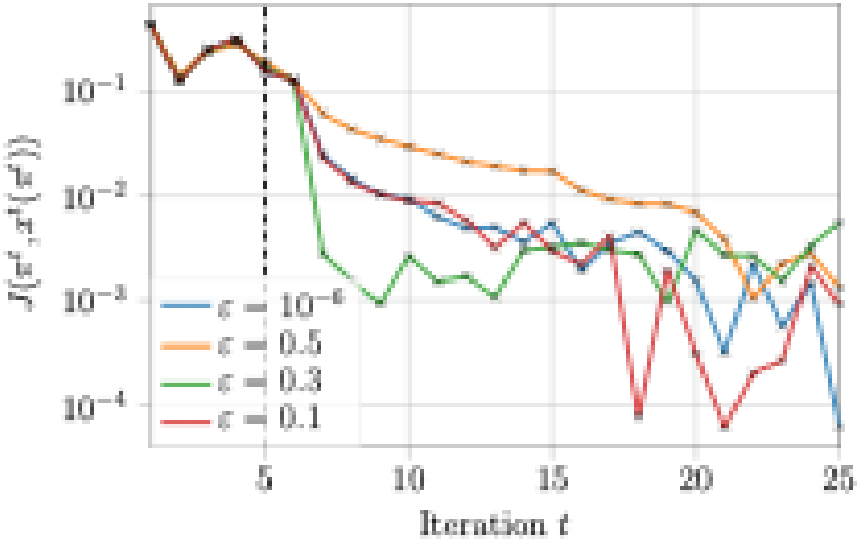
- **Ride-haling companies (Followers):** Allocate electric vehicles to districts with

$$x_i \in [0, x_{i,\max}]^d \cap \{x_i \in \mathbb{R}^d \mid 1^\top x_i \leq M_i\}$$

- **Government/ Power provider (Leader):** Set electricity prices $\pi \epsilon [0, \pi_{max}]^d$ and minimize social cost

$$J(\pi) = \left\| \frac{x(\pi)}{1^\top x(\pi)} - \xi^* \right\|^2$$





Chapter 6

Learning efficient equilibria in potential games via log-linear learning

Potential game

Game Γ is **potential** with $\Phi: \mathcal{A}^N: \rightarrow [0,1]$, if:

$$U_i(a_i, a_{-i}) - U_i(a_i', a_{-i}) = \Phi(a_i, a_{-i}) - \Phi(a_i', a_{-i})$$

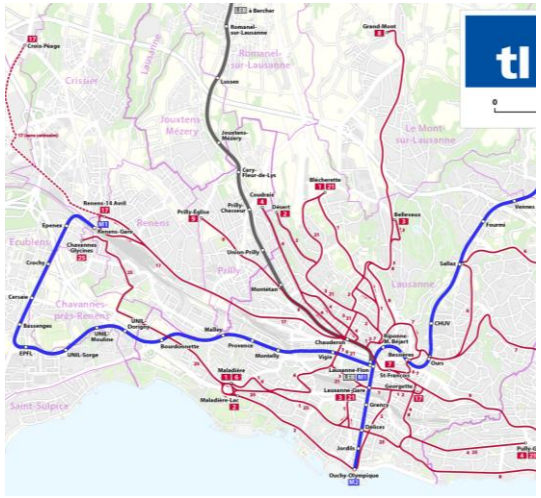
→ Examples: Congestion games, Identical interest games

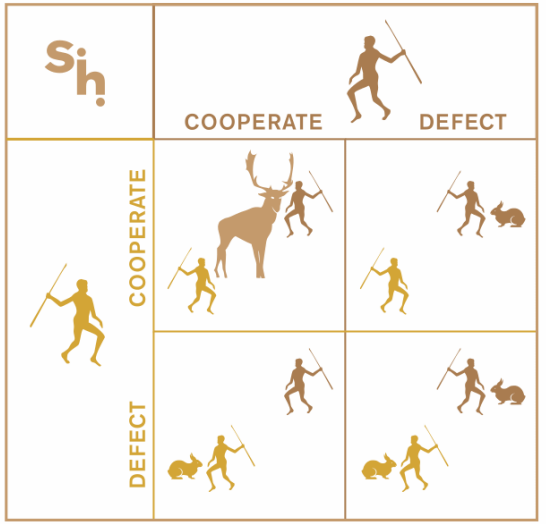
→ In potential games a **Nash equilibrium exists** [1]:

$$a^* \in \arg\max_{a \in \mathcal{A}^N} \Phi(a)$$

→ Nash equilibrium a^* is termed **efficient**

Can an efficient Nash equilibrium be learned in potential games?





By Christopher X Jon Jensen (CXJJensen) & Greg Riestenberg (Own work)
[CC BY-SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0)] via Wikimedia Commons

[1] D. Monderer and L.S. Shapley. "Potential games". In: Games and Economic Behavior (1996)

Log-linear learning

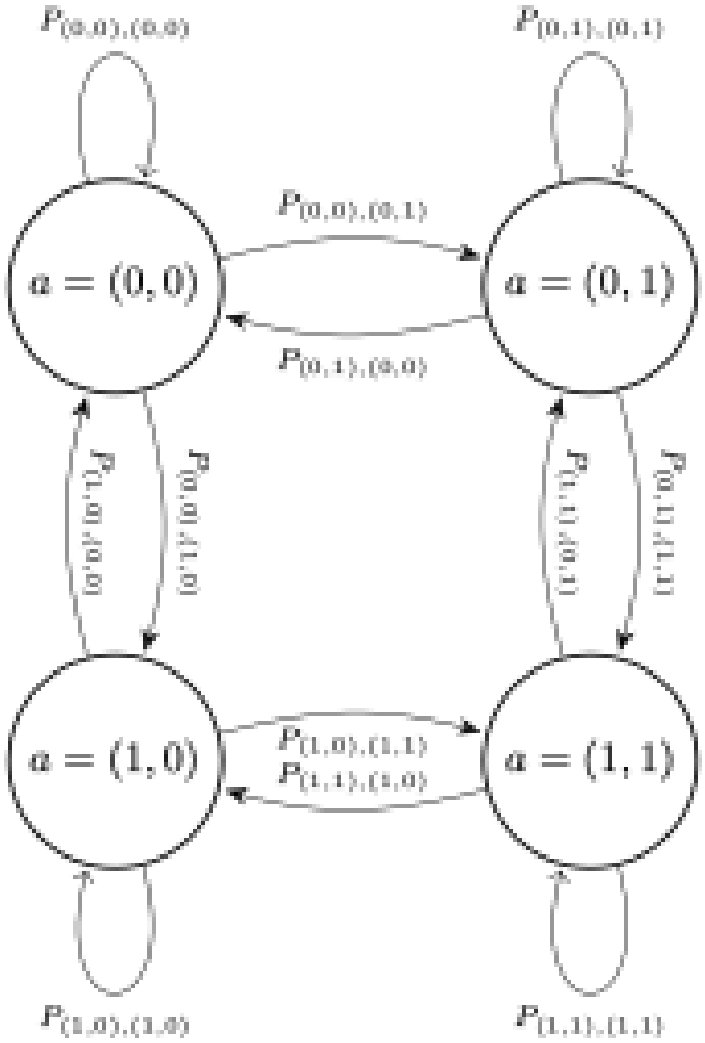
For $t = 1,2, \ldots$

1. Agent i is randomly chosen
2. All other agents repeat their action, i.e. $a_{-i}^{t} = a_{-i}^{t-1}$
3. Agent i samples $a_i \sim p_i^t \epsilon \Delta(\mathcal{A})$ with

$$p_i^t(a_i) = \frac{e^{\beta U_i(a_i, a_{-i}^{t-1})}}{\sum_{a_i' \in \mathcal{A}} e^{\beta U_i(a_i', a_{-i}^{t-1})}}$$

→ Log-linear learning induces an irreducible and aperiodic **Markov chain** $\{X_t\}_{t\epsilon\mathbb{N}}$ with:[1,2]

$$P_{\mathbf{a},\tilde{\mathbf{a}}} = \underbrace{\frac{1}{N}}_{1.} \underbrace{\frac{e^{\beta U_i(\tilde{a}_i, \tilde{a}_{-i})}}{\sum_{a_i' \in \mathcal{A}} e^{\beta U_i(a_i', \tilde{a}_{-i})}}}_{2.} \underbrace{1_{\mathcal{N}(a)}(\tilde{a})}_{3.}$$



[1] J.R. Marden and J.S. Shamma. " Revisiting log-linear learning: Asynchrony, completeness and payoff-based implementation ". In: Games and Economic Behavior (2012).
[2] L.E. Blume. "The statistical mechanics of strategic interaction". In: Games and Economic Behavior (1993).

Convergence time of log-linear learning

Theorem: [1]

For $\varepsilon\epsilon(0,1)$ and initial distribution μ^0, assume agents **follow log-linear learning** with $\beta = \tilde{\Omega}\left(\frac{1}{\Delta}\log\frac{A^N}{\epsilon}\right)$. Then

$$\mathbb{E}_{a\sim\mu^t}[\Phi(a)] \geq \max_{a\in\mathcal{A}^N}\Phi(a) - \epsilon$$

for

$$t = \tilde{\mathcal{O}}\left(N^2A^5\left(\frac{A^N}{\epsilon}\right)^{\frac{1}{\Delta}}\right)$$

where $\mu^t = \mu^0 P^t$ and $\Delta = \min_{a\in\mathcal{A}^N:\Phi(a)<\Phi(a^*)}(\Phi(a^*) - \Phi(a))$ is the **suboptimality gap**.

- **First finite-time convergence rate** to ε-efficient Nash equilbrium in potential games
- Convergence time grows **polynomially in A and** $1/\varepsilon$ and **exponentially in N**

By exploiting suboptimality gap

unavoidable

[1] A. Maddux*, R. Ouhamma*, H. Catic, and M. Kamgarpour, "Finite-time convergence to an ϵ-efficient Nash equilibrium in potential games". In: IEEE Transactions on Control of Network Systems (TCNS). 2026.

Key takeaways & future work

Existence of constrained correlated equilibria

- Strong Slater's condition necessary for playerwise coupling constraints
- Existence under weakened Slater's conditon for common coupling constraints

Learning constrained coarse correlated equilibria (CCE)

- Developed a no-regret, no-violation algorithm
- Convergence to constrained (CCE) in games with private constraints

The type of constraint sets dictate the existence condition

Learning in games with coupling constraints

- Conditions?
- Settings?
- Type of constraints?

Global equilibrium notions can be recovered from local strategy improvements

Key takeaways & future work

Incentive design in Stackelberg games

- Developed no-regret algorithm for the leader
- Convergence to ε-Stackelberg equilibria without needing followers' private utility gradients

Learning efficient Nash equilibria

- Finite-time convergence guarantees of log-linear learning to ε-efficient Nash equilibria in potential games

Central authority learns optimal incentive that minimizes social cost

- Extend to contextual Stackelberg games
- Consider preference feedback for the leader
- Design incentive scheme with sublinearly decreasing central intervention

Agents can independently learn efficient outcomes in specific game settings

Thank you!

Anna Maria Maddux

Ph.D. defense, 19. March 2026, EPFL

EPFL sycamore National Centre of Competence in Research Automation Swiss National Science Foundation

Appendix

Constrained correlated equilibrium

- Stochastic modification of agent i: $\phi_i : \mathcal{A}_i \to \Delta(\mathcal{A}_i)$
- Modified distribution $\phi_i \circ \sigma$ with $\sigma \epsilon \Delta(\mathcal{A})$: $(\phi_i \circ \sigma)(a) = \sum_{a'_i \in \mathcal{A}_i} \phi_i(a_i \mid a'_i)\sigma(a'_i, a_{-i})$
- ϕ_i is i-feasible: $\mathbf{E}_{a \sim \phi_i \circ \sigma}[g_i(a)] \leq 0$

Definition

A distribution $\sigma \epsilon \Delta(\mathcal{A})$ is a **constrained correlated equilibrium** if:

- σ is feasible: $\mathbf{E}_{a \sim \sigma}[g_i(a)] \leq 0$
- For each agent i: $\mathbf{E}_{a \sim \sigma}[r_i(a)] \geq \max_{\phi_i \text{ is i-feasible}} \mathbf{E}_{a \sim \phi_i \circ \sigma}[r_i(a)]$

→ Deviating via a stochastic modification is either **unprofitable or infeasible**

Strong Slater's condition

unrealistic

Strong Slater's assumption

For any agent i and distribution σ, there exists modification ϕ_i such that: $\mathbf{E}_{a\sim\phi_i\circ\sigma}[g_i(a)] \leq 0$

→ Under strong Slater's condition a **constrained Nash equilibrium exists** [1]

→ Implies existence of a constrained correlated equilibrium

⚠ Holds for Markov games

Can strong Slater's condition be weakened?

→ Not in games with **playerwise coupling constraints**

[1] E. Altman and A. Shwartz. "Constrained Markov games: Nash equilibria". In: Advances in dynamic games and applications. Springer, 2000.

Existence condition

→ Restrict to **common coupling constraints** ($g_i(\cdot) = g(\cdot)$ for each agent i)

→ Deterministic modification of agent i: $\phi_i(k) : \mathcal{A}_i \to \mathcal{A}_i$ with $k \epsilon [K_i]$ and $K_i = |\{\phi_i(k) : \mathcal{A}_i \to \mathcal{A}_i\}|$

Weaker Slater's assumption

For any agent i and distribution σ such that $\mathbb{E}_{a\sim\sigma}|g_i(a)| = 0$, there exists an $\alpha \epsilon \Delta(K_i)$ with $\alpha_k > 0$ for all k such that:

$$\sum_{k=1}^{K_i} \alpha_k \mathbb{E}_{a\sim\phi_i(k)\circ\sigma}[g(a)] \leq 0$$

Theorem

Let the weaker Slater assumption hold. Then, in games with **common coupling constraints** there exists a **constrained correlated equilibrium**.

⚠ Holds for Markov games [1]

[1] T. Ni*, A. Maddux*, and M. Kamgarpour, "On characterization and existence of constrained correlated equilibria in Markov games". Under review: Automatica. 2025.

Constrained coarse correlated equilibrium

Definition:

A distribution $\sigma \epsilon \Delta(\mathcal{A})$ is a **constrained coarse correlated equilibrium (CCE)** if:

- σ is feasible: $\mathbf{E}_{a\sim\sigma}[g_i(a_i)] \leq 0$
- For each agent i: $\mathbf{E}_{a\sim\sigma}[r_i(a)] \geq \max_{a_i' \in \mathcal{A}_i'} \mathbf{E}_{a\sim\sigma}[r_i(a_i', a_{-i})]$

Empirical joint distribution: $\sigma^T(a) = \frac{1}{T}|\{t \in [T] \mid a^t = a\}|$

Proposition:

After T rounds, σ^T is an ε-constrained CCE with $\varepsilon \leq \max\{max_{i\epsilon\mathcal{N}} \frac{R_i(T)}{T}, max_{i\epsilon\mathcal{N}} \frac{\mathcal{V}_i(T)}{T}\}$ [1]

!"#$%&'()*+,-./0123456789:;ÿ<=>?@ABCDEFGHIJKLMNOPQ#&):'0<689,24.5-$%(/13=>?@ABCDEFGHIJKLMNOPQRSÖÜáàâäãåçéêëíîïñóòôöõúùûüý RSTUVWXYZ[\]^_`abcdefghi

Proof outline

By Hölder's inequality:

$$\mathbb{E}_{a\sim\mu^t}[\Phi(a)] \geq \underbrace{\mathbb{E}_{a\sim\mu}[\Phi(a)]}_{\text{First term}} - 2\underbrace{\|\mu^t - \mu\|_{TV}}_{\text{Second term}} \underbrace{\max_{a\in A^N} \Phi(a)}_{\leq 1}$$

$$\geq \max_{a\in A^N} \Phi(a) - \frac{\epsilon}{2} - 2\|\mu^t - \mu\|_{TV} \max_{a\in A^N} \Phi(a)$$

Lemma:

For $\beta = \tilde{\Omega}\left(\frac{1}{\Delta}\log\frac{A^N}{\epsilon}\right)$ it holds that $\mathbb{E}_{a\sim\mu}[\Phi(a)] \geq \max_{a\in A^N}\Phi(a) - \epsilon/2$

Lemma:

For $t = \tilde{\mathcal{O}}\left(N^2A^5\left(\frac{A^N}{\epsilon}\right)^{\frac{1}{\Delta}}\right)$ it holds that $\|\mu^t - \mu\|_{TV} \leq \epsilon/2$. We used that:

1. $\|\mu^t - \mu\|_{TV} \leq \epsilon/2$ if $t \geq \frac{1}{\rho(PP^*)}\left(\log\log\frac{1}{\mu_{\min}} + 2\log\frac{2}{\epsilon}\right)$
2. For log-linear-learning it holds that $\rho(PP^*) \geq \frac{16\pi^2 A^{N-2}\mu_{\min}p_{\min}^2}{25N^2}$ with $\mu_{\min} \geq \frac{\epsilon^{-\beta}}{A^N}$ and $p_{\min} \geq \frac{\epsilon^{-\beta}}{A^N}$

Extensions of log-linear learning

1. Reduced dependence on number of agents [1]

Consider **symmetric potential games**:

$$U_i(a_1, \dots, a_N) = U_{\sigma(i)}(a_{\sigma(1)}, \dots, a_{\sigma(N)})$$

Assume agents follow modified log-linear learning:

- Potential games:

$$t = \tilde{\mathcal{O}}\left(\frac{A^N}{\epsilon}^{\frac{1}{\Delta}}\right)$$

- Symmetric potential games

$$t = \tilde{\mathcal{O}}\left(\frac{N^A}{\epsilon}^{\frac{1}{\Delta}}\right)$$

2. Reduced feedback [2]

Consider **binary log-linear learning:**
Select between $a_i' \sim Unif(\mathcal{A})$ and a_i^t

- Feedback LLL: $\{U_i(a_i, a_{-i}^t)\}_{a_{-i} \in \mathcal{A}}$
- Feedback binary LLL: $U_i(a_i', a_{-i}^t)$ and $U_i(a_i^t, a_{-i}^t)$

→ Same convergence time up to constants

3. Corrupted utilities

Consider **noisy utilities** with bounded noise ξ_i: $\bar{U}_i(a) = U_i(a) + \xi_i(a)$

Assume agents follow log-linear learning:

- Noiseless utilites:

$$\mathbb{E}_{a \sim u^t}[\Phi(a)] \geq \max_{a \in A^N} \Phi(a) - \epsilon$$

- Noisy utilities:

$$\mathbb{E}_{a \sim u^t}[\Phi(a)] \geq \max_{a \in A^N} \Phi(a) - \epsilon - \mathcal{O}(L)$$

[1] D. Shah and J. Shin. "Dynamics in congestion games". In: ACM SIGMETRICS Performance Evaluation Review (2010).
[2] G. Arslan, J. R. Marden, and J.S. Shamma. "Autonomous VehicleTarget Assignment: A Game-Theoretical Formulation". In: Journal of Dynamic Systems, Measurement, and Control (2007).