Independent Learning in Constrained Markov Potential Games

Philip Jordan Anas Barakat Niao He

Motivation

Multi-agent RL in Markov Potential Games with:

- **independent learning** for:
  (a) scaling (breaking curse of multi-agents),
  (b) privacy (no information sharing),
  (c) avoid communication cost.
- common coupled **constraints**; e.g., collision avoidance in autonomous driving, or power constraints in signal transmission

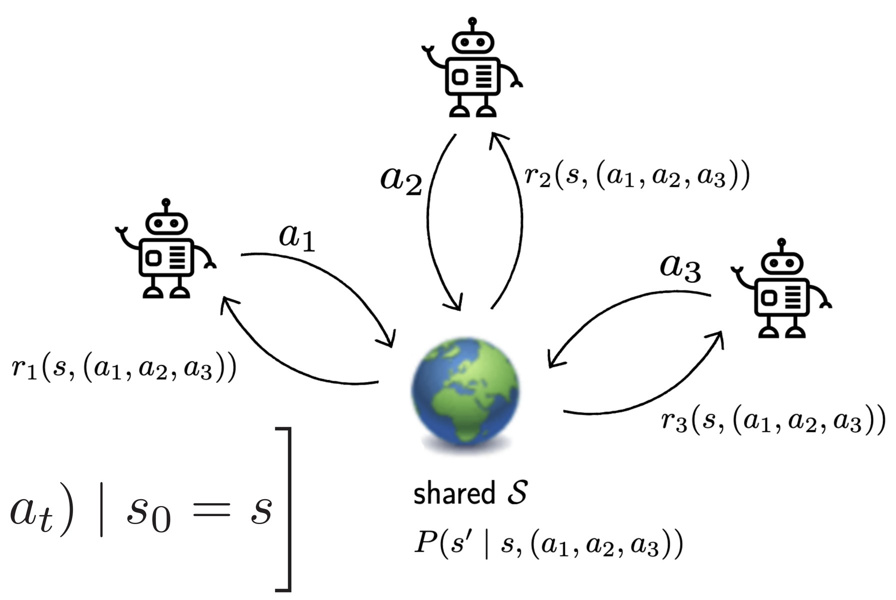
Related Work

	centralized	independent
MPG	Nash-CA; [1]; $\mathcal{O}(\epsilon^{-3})$	Ind. PGA; [2]; $\mathcal{O}(\epsilon^{-5})$
CMPG	CA-CMPG; [3]; $\tilde{\mathcal{O}}(\epsilon^{-5})$	*Ours (iProxCMPG)*; $\tilde{\mathcal{O}}(\epsilon^{-7})$

Problem Setting

Markov Game $\mathcal{G} = (\mathcal{S}, \mathcal{N}, \{\mathcal{A}_i, r_i\}_{i\in\mathcal{N}}, c, \alpha, \mu, P, \kappa)$

- shared state space $\mathcal{S}$
- players $\mathcal{N} = \{1, \ldots, m\}$
- joint policy space
$$\Pi = \prod_{i\in\mathcal{N}} \Delta(\mathcal{A}_i)^{\mathcal{S}}$$
- indiv. value functions
$$V_{r_i}(\pi) := \mathbb{E}_{s\sim\mu}\left[\sum_{t=0}^{T} r_i(s_t, a_t) \mid s_0 = s\right]$$



- potential structure:
$\exists \Phi : \Pi \to \mathbb{R}$ s.t. $\forall i \in \mathcal{N}$, $(\pi_i, \pi_{-i}) \in \Pi$, and $\pi_i' \in \Pi_i'$,
$$V_{r_i}(\pi_i, \pi_{-i}) - V_{r_i}(\pi_i', \pi_{-i}) = \Phi(\pi_i, \pi_{-i}) - \Phi(\pi_i', \pi_{-i})$$
- constr. threshold $\alpha \in \mathbb{R}$; feasible set $\Pi_c := \{\pi \in \Pi \mid V_c(\pi) \leq \alpha\}$,
$$V_c(\pi) := \mathbb{E}_{s\sim\mu}\left[\sum_{t=0}^{T} c(s_t, a_t) \mid s_0 = s\right]$$
- **solution concept:** ϵ-approx. NE: $\pi^* \in \Pi$ s.t. $\forall i \in \mathcal{N}, \pi_i' \in \Pi_c^i(\pi_{-i}^*)$,
$$V_{r_i}(\pi^*) - V_{r_i}(\pi_i', \pi_{-i}^*) \leq \epsilon.$$

Challenges

- nonconvex objective **and** constraint; constr. opt. challenge
- constraint **couples** π_i's; how to learn independently?
- **no** strong duality [3]; prohibits CMDP primal-dual methods

Main Contributions

- design of an algorithm for **independent** learning of **constrained** ϵ**-approximate Nash equilibria** in CMPGs
- establish **sample complexity** with poly. dependency on ϵ and problem parameters
- two CMPG **applications**: pollution tax & energy marketplace

Method

- proximal-point-like update
$$\pi^{(t+1)} = \arg\min_{\pi\in\Pi} \left\{\Phi(\pi) + \tfrac{1}{2\eta}\|\pi - \pi^{(t)}\|^2 \;\middle|\; V_c(\pi) + \tfrac{1}{2\eta}\|\pi - \pi^{(t)}\|^2 \leq \alpha\right\}$$
converges to ϵ-KKT policy $\Rightarrow$ ϵ-approx. constr. NE
- Φ and V_c weakly cvx $\Rightarrow$ subproblem obj. and constr. strongly cvx
$\rightarrow$ solve via gradient switching
- **observation:**
$\nabla_{\pi_i}\Phi_{\eta,\pi'}(\pi) = \nabla_{\pi_i}\Phi(\pi) + \frac{1}{\eta}(\pi_i - \pi_i') = \nabla_{\pi_i}V_{r_i}(\pi) + \frac{1}{\eta}(\pi_i - \pi_i')$
$\Rightarrow$ implementable as independent PG steps

Algorithm (iProxCMPG)

for $t = 0, \ldots, T-1$ **do**
  $\pi_i^{(t,0)} = \pi_i^{(t)}$ for $i \in \mathcal{N}$
  for $k = 0, \ldots, K-1$ and $i \in \mathcal{N}$ simultaneously **do**
$$\pi_i^{(t,k+1)} = \begin{cases} \mathcal{P}_{\Pi^{i,\xi}}\left[\pi_i^{(t,k)} - \nu_k \hat{\nabla}_{\pi_i} V_{r_i}(\pi^{(t,k)}) - \frac{\nu_k}{\eta}(\pi_i^{(t,k)} - \pi_i^{(t)})\right] & \text{if } \hat{V}_c(\pi^{(t,k)}) + \beta - \alpha \leq \delta_k \\ \mathcal{P}_{\Pi^{i,\xi}}\left[\pi_i^{(t,k)} - \nu_k \hat{\nabla}_{\pi_i} V_c(\pi^{(t,k)}) - \frac{\nu_k}{\eta}(\pi_i^{(t,k)} - \pi_i^{(t)})\right] & \end{cases}$$
  $\pi_i^{(t+1)} = \pi_i^{(t,\hat{k})}$ s.t. $\mathbb{P}(\hat{k} = k) = \left(\sum_{k\in\mathcal{B}^{(t)}} \rho_k\right)^{-1} \rho_k$
output: $\pi_i^{(T)}$ for $i \in \mathcal{N}$

Convergence Results

Assumptions.

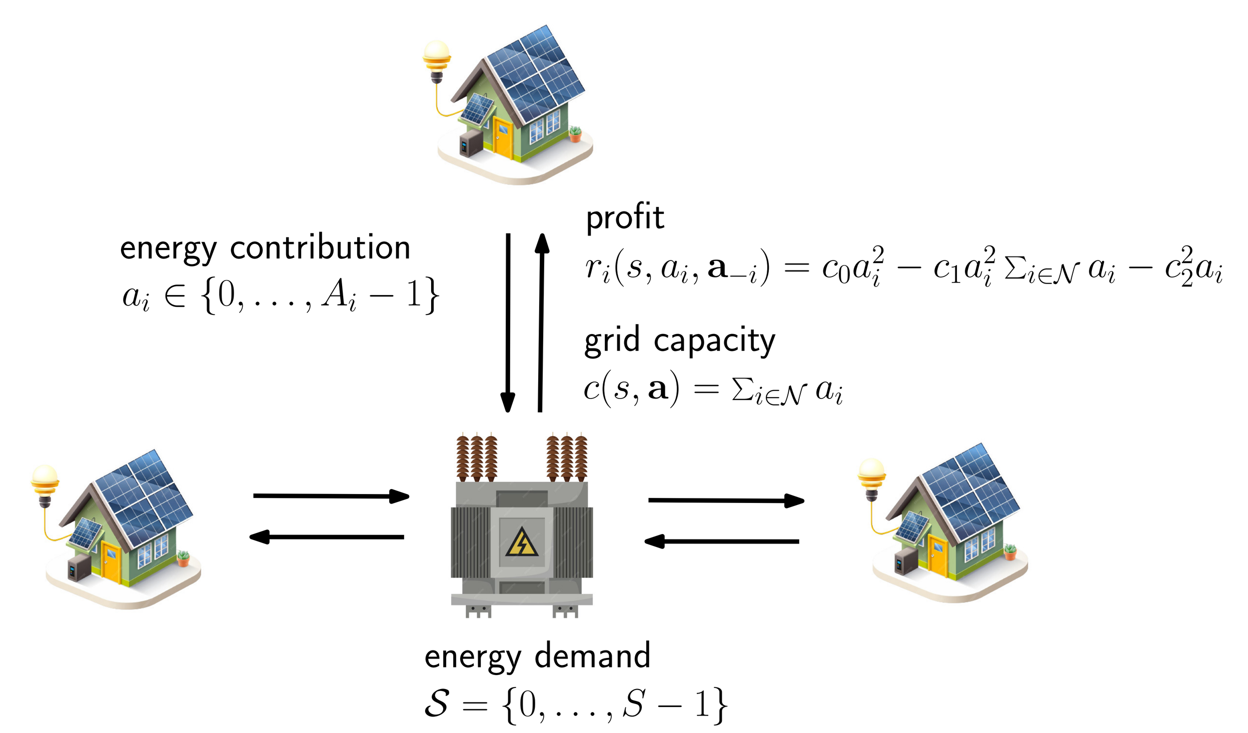
1. **initial feasibility:** $\pi^{(0)}$ satisfies $V_c(\pi^{(0)}) < \alpha$
2. **uniform Slater's condition:**
$$\exists\zeta > 0 \text{ s.t. } \forall \pi' \in \Pi \text{ with } V_c(\pi') < \alpha, \exists \pi \in \Pi \text{ s.t. } V^c_{\eta,\pi'}(\pi) \leq \alpha - \zeta$$

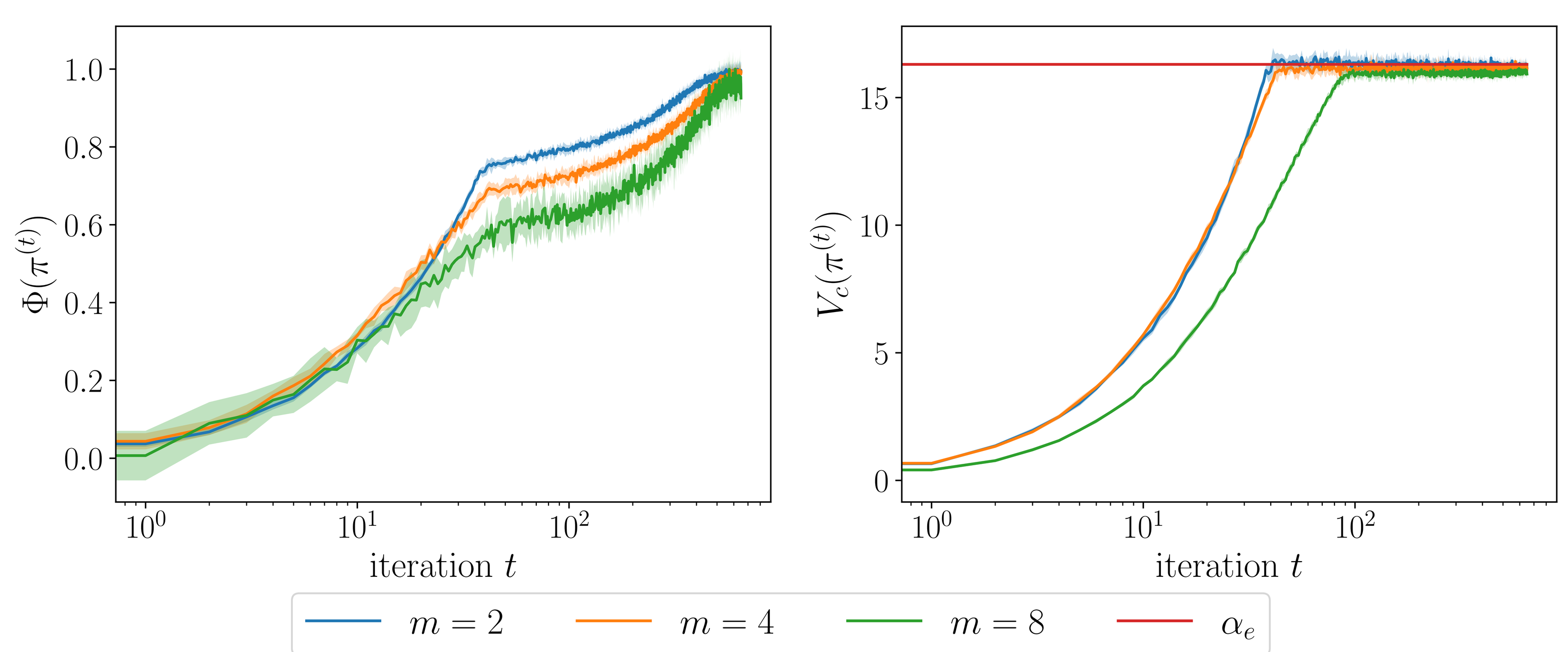
Theorem. For $\epsilon > 0$, using ϵ-greedy exploration, after running iProxCMPG for suitably chosen η, T, K, and $\{(\nu_k, \delta_k)\}_{0\leq k\leq K}$, there exists $t \in [T]$ s.t. in expectation $\pi^{(t)}$ is a constrained ϵ-NE.

- **exact gradients:** total iteration complexity[a] $\tilde{\mathcal{O}}(\epsilon^{-4})$
- **finite sample:** total sample complexity[1] $\tilde{\mathcal{O}}(\epsilon^{-7})$

[a] $\tilde{\mathcal{O}}(\cdot)$ hides logarithmic dependencies in $1/\epsilon$, and polynomial dependencies in $m, S, A_{\max}, 1-\gamma, \zeta$, and D.

Simulations: Energy Marketplace





Future Work

- learning constrained NEs beyond CMPGs
- "fully" independent learning (different stepsizes/algorithms)
- coupled playerwise (instead of common) constraints

References

[1] Ziang Song, Song Mei, and Yu Bai. When can we learn general-sum markov games with a large number of players sample-efficiently? In *ICLR*, 2022.
[2] Stefanos Leonardos, Will Overman, Ioannis Panageas, and Georgios Piliouras. Global convergence of multi-agent policy gradient in markov potential games. In *ICLR*, 2022.
[3] Pragnya Alatur, Giorgia Ramponi, Niao He, and Andreas Krause. Provably learning nash policies in constrained markov potential games. In *AAMAS*, 2024.