Regret Lower Bounds for Decentralized Multi-Agent Stochastic Shortest Path Problems

Utkarsh U. Chavan[†] Prashant Trivedi[*] Nandyala Hemachandra[†]

[†]Indian Institute of Technology Bombay

[*]University of Petroleum & Energy Studies, Dehradun

December 2025

Background

- Stochastic Shortest Path problems (SSPs)
- Capture many RL problems [1] : navigation, swarm robotics, goal-oriented games, etc.

- Some notable work on learning of SSPs include
  - [Tarbouriech et al., 2020, Tarbouriech et al., 2021]
  - [Cohen et al., 2021, Min et al., 2021, Vial et al., 2021]

[1] Dimitri Bertsekas, Dynamic Programming and Optimal Control, Vols. I and II, Athena Scientific, 1995, (3rd Ed. Vol. I, 2005, 4th Ed. Vol. II, 2012)

Motivation & Challenges for Multi-Agent SSPs

- Focus had been on single-agent SSP learning
- Rarely a single agent traverses the network
- For two or more agents we need to model the *congestion*
- Often these agents act on their own but have a *common objective*
- The agents *affect* each others' *costs* and *transitions*
- **A natural approach**: use a central controller
  - This has scalability issues
- **Another extreme**: no communication
  - The common objective can't be attained
- [Trivedi and Hemachandra, 2023] introduce *fully decentralized* Multi-Agent version of the SSP learning problem

Setting: Learning MASSPs

- Defined by tuple $(\mathcal{V}, \mathcal{N}, \mathbb{P}, \mathcal{A})$
- Nodes $\mathcal{V} = \{v_1, \ldots, v_q\}$, agents $\mathcal{N} = \{1, \ldots, n\}$
- Global state $\boldsymbol{s} = (s_1, \ldots, s_n) \in \mathcal{S} = \mathcal{V}^n$
- Agents start at $\boldsymbol{s}_{\text{init}}$ to reach the goal state $\boldsymbol{g}$
- Global action $\boldsymbol{a}$, transition kernel $\mathbb{P}(\cdot \mid \boldsymbol{s}, \boldsymbol{a})$
- Cost for agent i: $c_i(\boldsymbol{s}, \boldsymbol{a}) \in [c_{\min}, 1]$
- Global cost $\bar{c}(\boldsymbol{s}, \boldsymbol{a}) = \frac{1}{n} \sum_{i=1}^{n} c_i(\boldsymbol{s}, \boldsymbol{a})$
- Agents communicate via a *communication* network

Objective

Learn a policy π^* that minimizes $V^\pi(\boldsymbol{s}_{\text{init}})$ over K episodes

$$V^\pi(\boldsymbol{s}_{\text{init}}) = \mathbb{E}\left[\sum_{t=1}^{\tau^\pi(\boldsymbol{s}^1)} c(\boldsymbol{s}^t, \pi(\boldsymbol{s}^t)) \;\middle|\; \boldsymbol{s}^1 = \boldsymbol{s}_{\text{init}}\right]$$

Linear Model and Performance Metric

Assumption (Linear Dynamics)

For every $(\boldsymbol{s}, \boldsymbol{a}, \boldsymbol{s}')$, there exist known features $\phi(\boldsymbol{s}'|\boldsymbol{s}, \boldsymbol{a}) \in \mathbb{R}^{nd}$ and unknown parameter $\theta \in \mathbb{R}^{nd}$ such that

$$\mathbb{P}(\boldsymbol{s}'|\boldsymbol{s}, \boldsymbol{a}) = \langle \phi(\boldsymbol{s}'|\boldsymbol{s}, \boldsymbol{a}), \theta \rangle$$

- This defines the family of *linear mixture MASSPs* (LM-MASSPs)

Performance Metric (Regret)

The *regret* over K episodes is

$$\mathbb{E}_{\theta,\pi}[R(K)] = \mathbb{E}\left[\sum_{k=1}^{K}\sum_{h=1}^{h_k} c(\boldsymbol{s}^{k,h}, \boldsymbol{a}^{k,h})\right] - K \cdot V^*(\boldsymbol{s}_{\text{init}}),$$

where h_k is the random length of episode k under the algorithm π

Central Question

Problem

What are the fundamental limits of learning in decentralized multi-agent SSPs with linear function approximation?

- **Solution:** Establish tight *regret lower bounds* via construction of *hard-to-learn* instances

Our Approach

- Construct families of provably hard-to-learn 2-node instances
- $\mathcal{V} = \{s, g\}; \qquad \mathcal{S} = \{s, g\}^n$
- Action set $\mathcal{A}_i = \{-1, 1\}^{d-1}; \qquad \mathcal{A} = \{-1, 1\}^{n(d-1)}$
- **Key ideas**
  - Design feature map $\phi : (\boldsymbol{s}, \boldsymbol{a}, \boldsymbol{s}') \to \mathbb{R}^{nd}$ to ensure valid linear transitions, parameterized by $\theta \in \mathbb{R}^{nd}$
  - Obtain an analytically tractable optimal policy
- An instance is given by $(n, \delta, \Delta, \theta)$

Optimal Policy Structure

- Partition states by *type*
- $\mathcal{S}_r :=$ states with exactly r agents at node $s \in \mathcal{V}$

Theorem (Optimal Policy Structure)

For any instance $(n, \delta, \Delta, \theta)$,

- *Optimal policy: choose* $\boldsymbol{a}_\theta$ *in every state*
- *Optimal value depends only on type* r*:* V_r^*
- $0 = V_0^* < \cdots < V_n^* = B^*$ *(B^* is SSP diameter)*

Lower Bound on Regret

Theorem (Lower bound)

For any decentralized learning algorithm π, $\delta \in (2/5, 1/2)$ *and* $\Delta < 2^{-n} \cdot \frac{1-2\delta}{1+n+n^2}$, *there exists an LM-MASSP instance such that for* $K > \frac{n(d-1)^2 \cdot \delta}{2^{10} B^* \left(\frac{1-2\delta}{1+n+n^2}\right)^2}$ *episodes,*

$$\mathbb{E}_{\theta,\pi}[R(K)] \geq \frac{d \cdot \sqrt{\delta} \cdot \sqrt{KB^*/n}}{2^{n+9}}$$

- This matches the $\sqrt{K}$ regret upper bound for LM-MASSPs [Trivedi and Hemachandra, 2023]
- When $n = 1$, it recovers the regret lower bound [Min et al., 2021]

References I

Cohen, A., Efroni, Y., Mansour, Y., and Rosenberg, A. (2021).
Minimax regret for stochastic shortest path.
Advances in neural information processing systems, 34:28350–28361.

Min, Y., He, J., Wang, T., and Gu, Q. (2021).
Learning stochastic shortest path with linear function approximation.
In *International Conference on Machine Learning*.

Tarbouriech, J., Garcelon, E., Valko, M., Pirotta, M., and Lazaric, A. (2020).
No-regret exploration in goal-oriented reinforcement learning.
In III, H. D. and Singh, A., editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 9428–9437. PMLR.

Tarbouriech, J., Zhou, R., Du, S. S., Pirotta, M., Valko, M., and Lazaric, A. (2021).
Stochastic shortest path: Minimax, parameter-free and towards horizon-free regret.
Advances in neural information processing systems, 34:6843–6855.

References II

Trivedi, P. and Hemachandra, N. (2023).
Multi-agent congestion cost minimization with linear function approximations.
In *International Conference on Artificial Intelligence and Statistics*, pages 7611–7643. PMLR.

Vial, D., Parulekar, A., Shakkottai, S., and Srikant, R. (2021).
Regret bounds for stochastic shortest path problems with linear function approximation.
In *International Conference on Machine Learning*.

Thank You!