Continual Safe Adaptation in Multi-agent Domains

Yarden As[1] and Daphne Cornelisse[2]

[1]Learning and Adaptive Systems Group, ETH Zürich
[2]New York University

1 Background

Cooperative intelligence—the capacity to foster cooperation between humans, machines, or organizations—hinges on the ability to take *calculated risks* to better understand the behavior of others. This skill is essential in settings where outcomes are interdependent but incentives may not be fully aligned, such as in climate negotiations (Bengio et al., 2023), monitoring agentic AI systems (Reuel et al., 2024), or even navigating traffic in a city with different norms than what you are accustomed to. Humans are particularly good at continuously exploring and discerning patterns in the responses of others. We use this information to construct a *model* that describes how others think and act (Grosz and Kraus, 1996; Xiang et al., 2023). We then adapt accordingly when needed, achieving mutually beneficial outcomes without inflicting harm along the way.

For artificial agents to exhibit similar behaviors, they too must be able to navigate such risks while adapting their model of how their co-players "think". This challenge can be framed within the paradigm of sequential decision-making under constraints, where the objective is to refine models of other agents' behavior while pursuing individual goals, all while ensuring that safety and other constraints are upheld *continuously during learning*. In the single-agent setting, where agents typically learn models of the environment dynamics, this problem is known as *safe exploration* and has been extensively studied. Various methods have been developed to ensure that optimal policies maintain safety throughout the learning and adaptation process, particularly in smaller-scale environments (Sui et al., 2015; Berkenkamp et al., 2021), with recent advancements addressing more complex, high-dimensional, and non-stationary environments (As et al., 2022, 2024).

In the multi-agent setting, despite extensive research on multi-agent constrained Markov decision processes (CMPD, Altman, 1999; Garg et al., 2024), current approaches often struggle with scalability and with ensuring continuous constraint satisfaction while learning. This is partly because extending existing multi-agent frameworks to problems with non-stationary dynamics under constraints is non-trivial. A fundamental distinction between the single- and multi-agent settings is the necessity for agents to model *each other's* decision-making and engage in higher-order reasoning (Schroeder de Witt et al., 2019; Dafoe et al., 2020). If these methods can be effectively scaled to advanced AI systems, however, they could serve as effective tools for ensuring safe adaptation in high-stakes environments. To take steps in this direction, we ask the following questions:

Q.1 What is an appropriate framework for extending single-agent safe exploration algorithms to multi-agent problems, where agents must learn and adapt to behaviors of *other unfamiliar agents* while preventing harmful outcomes?

Q.2 How can we use constraints in a practical and scalable manner to induce cooperation in multi-agent settings with mixed incentives?

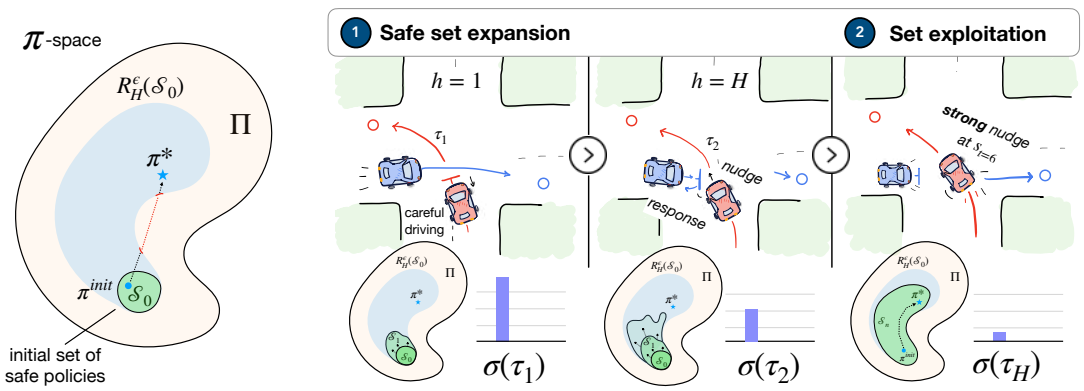

Figure 1: Illustration of approach. To safely adapt to new domains, we use ActSafe from As et al. (2024) to expand a safe set of policies by reducing our epistemic uncertainty σ. We are initially given a set of safe policies ($\mathcal{S}_0$) that was obtained with the baseline dynamics. The goal is to collect safely trajectories so as to learn about how other agents behave in new domains.

Our work aims to address these questions by developing a theoretically grounded foundation for artificial agents that can safely adapt to the behavior of others in uncertain, multi-agent environments.

2 Proposal

Consider an agent that is initially trained with a baseline policy derived from some baseline dynamics, and is ought to safely adapt when deployed in a new environment with different dynamics. For instance, a self-driving car policy trained with data from Zurich must adjust when deployed in New York, where driving behavior and culture differ significantly, as depicted in Figure 1. The goal is for the agent to gather new trajectory data safely to update its dynamics model without causing accidents or critical errors.

A starting point for solving the above problem, would be to extend existing safe reinforcement learning algorithms, such as LAMBDA or ActSafe (As et al., 2022, 2024), to multi-agent problems. In particular, we the above problem can be simulated in a controlled environment like *GPUDdrive* (Kazemkhani et al., 2024), which which provides a consistent benchmark for measuring an agent's adaptability to various driving norms. Through diverse datasets from different locations (Caesar et al., 2020; Ettinger et al., 2021), the adaptability of the agent can be assessed in complex, real-world-like situations.

3 Supervision

If you are a Master's student with

- basic knowledge in reinforcement learning, for instance, by taking *Probabilistic Artificial Intelligence* or *Foundations of Reinforcement Learning* courses;
- strong programming background in Python,

please reach out to Yarden As or Daphne Cornelisse.

References

E. Altman. *Constrained Markov Decision Processes*. Chapman and Hall, 1999.

Yarden As, Ilnura Usmanova, Sebastian Curi, and Andreas Krause. Constrained policy optimization via bayesian world models. *ICLR*, 2022.

Yarden As, Bhavya Sukhija, Lenart Treven, Carmelo Sferrazza, Stelian Coros, and Andreas Krause. Actsafe: Active exploration with safety constraints for reinforcement learning. *arXiv preprint arXiv:5921425*, 2024.

Yoshua Bengio, Prateek Gupta, Lu Li, Soham Phade, Sunil Srinivasa, Andrew Williams, Tianyu Zhang, Yang Zhang, and Stephan Zheng. Ai for global climate cooperation 2023 competition proceedings. *arXiv preprint arXiv:2307.06951*, 2023.

Felix Berkenkamp, Andreas Krause, and Angela P Schoellig. Bayesian optimization with safety constraints: safe and automatic parameter tuning in robotics. *Machine Learning*, 2021.

Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *ICCV*, 2020.

Allan Dafoe, Edward Hughes, Yoram Bachrach, Tantum Collins, Kevin R McKee, Joel Z Leibo, Kate Larson, and Thore Graepel. Open problems in cooperative ai. *arXiv preprint arXiv:2012.08630*, 2020.

Scott Ettinger, Shuyang Cheng, Benjamin Caine, Chenxi Liu, Hang Zhao, Sabeek Pradhan, Yuning Chai, Ben Sapp, Charles R Qi, Yin Zhou, et al. Large scale interactive motion forecasting for autonomous driving: The waymo open motion dataset. In *ICCV*, 2021.

Kunal Garg, Songyuan Zhang, Oswin So, Charles Dawson, and Chuchu Fan. Learning safe control for multi-robot systems: Methods, verification, and open challenges. *Annual Reviews in Control*, 2024.

Barbara J Grosz and Sarit Kraus. Collaborative plans for complex group action. *Artificial Intelligence*, 1996.

Saman Kazemkhani, Aarav Pandya, Daphne Cornelisse, Brennan Shacklett, and Eugene Vinitsky. Gpudrive: Data-driven, multi-agent driving simulation at 1 million fps. *arXiv preprint arXiv:2408.01584*, 2024.

Anka Reuel, Ben Bucknall, Stephen Casper, Tim Fist, Lisa Soder, Onni Aarne, Lewis Hammond, Lujain Ibrahim, Alan Chan, Peter Wills, et al. Open problems in technical ai governance. *arXiv preprint arXiv:2407.14981*, 2024.

Christian Schroeder de Witt, Jakob Foerster, Gregory Farquhar, Philip Torr, Wendelin Boehmer, and Shimon Whiteson. Multi-agent common knowledge reinforcement learning. *NeurIPS*, 2019.

Yanan Sui, Alkis Gotovos, Joel Burdick, and Andreas Krause. Safe exploration for optimization with gaussian processes. In *ICML*, 2015.

Yang Xiang, Natalia Vélez, and Samuel J Gershman. Collaborative decision making is grounded in representations of other people's competence and effort. *Journal of Experimental Psychology: General*, 2023.