Self-Sufficient Cooperation and Exploration: From Optimality Principles to Efficient Algorithms

Max Muchen Sun
Center for Robotics and Biosystems, Northwestern University, Evanston, IL 60208, USA.
Email: msun@u.northwestern.edu

I. Research Overview

For robots to have a greater impact on society, they must be self-sufficient. This means being able to perceive, reason, and adapt to uncertain and unstructured environments and tasks using only onboard resources and no external supervision. Unlike virtual learning systems, robots have the agency to actively interact with the environment, other robots, and humans. Such agency necessitates two crucial capabilities to support self-sufficiency: cooperation and exploration. Cooperation enables robots to leverage the intelligence and agency of other robots and humans, and exploration enables robots to gather high-quality data to adapt in unfamiliar scenarios.

The goal of my research is to **develop *optimal* and *computationally efficient* strategies for cooperation and exploration, enabling self-sufficient robot autonomy across environments and tasks**. Formal optimality principles lead to interpretable and provable properties, while computational efficiency is necessary for physical intelligence using only onboard resources. My work spans theory to algorithm design, with a focus on real-world validation to bridge the gap between theoretical development and practical deployment.

II. Previous Research Contributions

A. Game-theoretic optimality for self-sufficient cooperation

Motivation. Self-sufficient cooperation requires robots to coordinate actions with humans or other robots without relying on explicit communication (e.g., verbal commands, predefined protocols). In such settings, each agent can only estimate others' intent, introducing inherent uncertainty in decision-making. This calls for both theoretical frameworks and practical algorithms to model emergent cooperative behaviors from individual decisions under uncertainty. My previous work explored this topic through the social navigation problem—safe and efficient navigation alongside humans in unstructured environments. Without accounting for cooperation from humans for collision avoidance, robots could exhibit overly aggressive or conservative actions, known as the "freezing robot problem" [22].

Game theory provides a principled framework for cooperation, where agents optimize individual objectives that depend on others' actions. The Nash equilibrium is defined as an optimality principle [11] where no agent can improve their objective by unilaterally changing their action; when decisions are represented as distributions, this extends to the mixed strategy Nash equilibrium (see Fig. 1). Compared to other human-robot interaction methods, game-theoretic optimality enables robots to account for how their actions influence others [15], leading to better cooperation performance [16, 8] and providing a structured framework for learning cooperation strategies [12, 7]. However, applying game theory for scalable real-time decision-making remains an open challenge due to high computational cost and the need to hand-craft agent objectives that align with real-world human behavior.

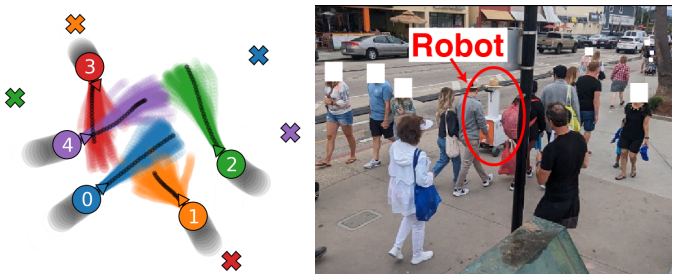

Fig. 1: *(Left)* Nash-optimal trajectory distributions for cooperative collision avoidance. *(Right)* Social navigation in Santa Cruz, CA, using only onboard perception and computation.

Contribution 1. In [21, 10], I developed scalable, real-time inference of mixed strategy Nash equilibrium for untethered social navigation in the real world. In [21], I introduced a structured decision-making formulation that splits each agent's objective into collective (e.g., collision avoidance) and individual (e.g., goal-directed navigation) components. This formulation enables an efficient recursive Bayesian inference scheme with guaranteed convergence to a mixed strategy Nash equilibrium [10], which outperforms both learning-based and non-learning-based social navigation methods in simulated and real-world dataset benchmarks. I further developed a full-stack social navigation system that was deployed on an untethered wheeled robot at Honda Research Institute and a quadruped robot, with the former used for a large-scale field study in Santa Cruz, CA (see Fig. 1).

Contribution 2. The entanglement between individual behavior and group behavior often obscures the learning of cooperative policies. In [20], I applied game-theoretic optimality to address this issue by developing a differentiable optimization layer using the mixed strategy Nash equilibrium algorithm from [10]. The proposed method effectively guides the learning process to distinguish individual policy and cooperation among individual policies without compromising the fidelity

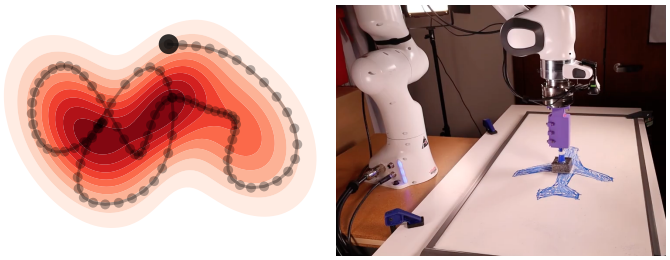

Fig. 2: *(Left)* A coverage trajectory under the ergodic optimality. *(Right)* Ergodic coverage applied to a vision-based erasing task under uncertainty from the onboard camera.

of the learned policy. Structuring the learning process using Nash optimality significantly improved the data efficiency, the robustness of the policy when facing novel behaviors, and enabled learning from varying numbers of agents.

B. Ergodic optimality for self-sufficient exploration

Motivation. Self-sufficient exploration requires robots to continuously search across uncertain, unstructured environments over extended periods of time. Uncertainty from the environment, sensor measurements, and robot dynamics leads to uneven information distributions, making repeated visitation of certain regions inevitable. These requirements call for non-myopic and even non-Markovian decision-making. In particular, the conventional state tracking-based notion of optimality (e.g., next best view), which focuses on how each specific point-in-time state individually contributes to exploration performance, is insufficient compared to *coverage*-based methods that account for spatial and temporal correlation across the trajectory horizon [4, 23].

Ergodic coverage [9] provides a formal notion of optimality for coverage-based exploration, where the ergodic metric measures the difference between the spatial distribution of the robot trajectory and a target distribution (see Fig. 2). Ergodic coverage generates trajectories that allocate more time covering regions with higher information density while maintaining guaranteed asymptotic coverage, outperforming greedy information maximization and uniform coverage [14].

Contribution 1. In [18], I significantly improved the scalability and computational efficiency of ergodic coverage through the use of kernel functions, enabling real-time long-horizon ergodic coverage in 6D space and on Lie groups. The proposed method was applied to a peg-in-hole insertion task, where the problem is formulated as an exploration problem over the state visitation distribution of human demonstrations. Ergodic coverage reliably solves the problem without any learning at all, even under suboptimal human demonstrations, and the asymptotic coverage property leads to a 100% success rate given sufficient time.

Contribution 2. In [19], I introduced the flow matching method [5] from generative model learning to ergodic coverage, enabling statistical inference methods previously infeasible for ergodic coverage, including Stein variational gradient descent [6] and optimal transport [3]. Integrating these state-of-the-art inference techniques significantly improves exploration performance in scenarios challenging to existing methods, such as with unnormalized target distributions and non-smooth distributions with irregular supports. These advantages are demonstrated on hardware in a series of vision-based erasing tasks, where the proposed method robustly erases hand-drawn patterns through exploration, relying only on the noisy on-board camera (see Fig. 2).

III. Future Research Agenda

With the promise of machine learning for robotics, my future research answers how cooperation and exploration can further advance robot learning to support self-sufficiency.

Imitation learning of cooperation. *Given demonstrations of two humans carrying a table together, can a robot learn to carry the table with another human?* While imitation learning has gained significant attention in recent years [24, 2], the tasks addressed are predominantly single-agent or centralized. However, learning decentralized cooperative policies from multi-agent demonstrations imposes significant challenges: the policy depends on not only the environment, but also the actions of other agents. Robots lack full controllability over others' actions and must plan actions that align with estimation of their intent—and vice versa, align estimation of others' intent with the robot's planned actions. My future research will address these challenges by guiding the supervised learning process with structured cooperation models, such as differentiable mixed strategy Nash equilibrium layers [20]. This allows the learned policy to better distinguish between individual intent and inter-agent influences, infer others' intents, account for how the robot's actions could influence others, and improve data efficiency.

Optimal data collection for robot learning. Data quality is crucial for learning, and data collection for robots poses unique challenges: robots are constrained by their dynamics and geometry, make decisions in continuous time and space, and often face non-Markovian tasks. Recently, several coverage-based exploration methods have been formally shown to be optimal data collection strategies for robot learning [1, 13]. These methods achieve optimality by optimizing the independent and identically distributed (i.i.d.) property of collected data. However, existing works are limited to a small set of learning frameworks, such as model-based reinforcement learning [1] and PAC learning [13]. Moreover, the integrations with learning frameworks are post hoc, often compromising model fidelity. My future research will expand exploration-based data collection strategies focusing on the i.i.d. property to broader robot learning frameworks, such as imitation learning, where coverage has already been shown to be an effective demonstration strategy [17]. I will develop deeper integration of exploration-based data collection without compromising model performance. My previous work [19] directly integrated generative model training methods into coverage-based exploration. I will continue this direction, embedding exploration as an inherent component of the learning processes, such that the model produces not only inference results but also actions for continual exploration and adaptation.

References

[1] Thomas A. Berrueta, Allison Pinosky, and Todd D. Murphey. Maximum diffusion reinforcement learning. *Nature Machine Intelligence*, 6(5):504–514, 2024.

[2] Cheng Chi, Zhenjia Xu, Siyuan Feng, Eric Cousineau, Yilun Du, Benjamin Burchfiel, Russ Tedrake, and Shuran Song. Diffusion policy: Visuomotor policy learning via action diffusion. *The International Journal of Robotics Research*, 2024.

[3] Jean Feydy, Thibault Séjourné, François-Xavier Vialard, Shun-ichi Amari, Alain Trouve, and Gabriel Peyré. Interpolating between Optimal Transport and MMD using Sinkhorn Divergences. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, pages 2681–2690. 2019.

[4] Ayoung Kim and Ryan M. Eustice. Active visual SLAM for robotic area coverage: Theory and experiment. *The International Journal of Robotics Research*, 34(4-5):457–475, 2015.

[5] Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow Matching for Generative Modeling. In *The Eleventh International Conference on Learning Representations*, 2022.

[6] Qiang Liu and Dilin Wang. Stein Variational Gradient Descent: A General Purpose Bayesian Inference Algorithm. In *Advances in Neural Information Processing Systems*, volume 29. 2016.

[7] Xinjie Liu, Lasse Peters, Javier Alonso-Mora, Ufuk Topcu, and David Fridovich-Keil. Auto-Encoding Bayesian Inverse Games. In *Algorithmic Foundations of Robotics XVI*, Chicago IL USA, 2024. Springer International Publishing.

[8] Negar Mehr, Mingyu Wang, Maulik Bhatt, and Mac Schwager. Maximum-Entropy Multi-Agent Dynamic Games: Forward and Inverse Solutions. *IEEE Transactions on Robotics*, 39(3):1801–1815, 2023.

[9] Lauren M. Miller, Yonatan Silverman, Malcolm A. MacIver, and Todd D. Murphey. Ergodic Exploration of Distributed Information. *IEEE Transactions on Robotics*, 32(1):36–52, 2016.

[10] Max Muchen Sun, Francesca Baldini, Katie Hughes, Peter Trautman, and Todd Murphey. Mixed strategy Nash equilibrium for crowd navigation. *The International Journal of Robotics Research*, page 02783649241302342, 2024.

[11] John F. Nash. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36 (1):48–49, 1950.

[12] Lasse Peters, Vicenç Rubies-Royo, Claire J Tomlin, Laura Ferranti, Javier Alonso-Mora, Cyrill Stachniss, and David Fridovich-Keil. Online and offline learning of player objectives from partial observations in dynamic games. *The International Journal of Robotics Research*, 42(10):917–937, 2023.

[13] Allison Pinosky and Todd D. Murphey. Embodied Active Learning of Generative Sensor-Object Models, 2024. arXiv:2410.11130 [cs].

[14] Ahalya Prabhakar and Todd Murphey. Mechanical intelligence for learning embodied sensor-object relationships. *Nature Communications*, 13(1):4108, 2022.

[15] Dorsa Sadigh, Shankar Sastry, Sanjit A. Seshia, and Anca D. Dragan. Planning for Autonomous Cars that Leverage Effects on Human Actions. In *Proceedings of Robotics: Science and Systems*, AnnArbor, Michigan, 2016.

[16] Wilko Schwarting, Alyssa Pierson, Javier Alonso-Mora, Sertac Karaman, and Daniela Rus. Social behavior for autonomous vehicles. *Proceedings of the National Academy of Sciences*, 116(50):24972–24978, 2019.

[17] Max Simchowitz, Daniel Pfrommer, and Ali Jadbabaie. The Pitfalls of Imitation Learning when Actions are Continuous, 2025. arXiv:2503.09722 [cs].

[18] Max Muchen Sun, Ayush Gaggar, Pete Trautman, and Todd Murphey. Fast Ergodic Search With Kernel Functions. *IEEE Transactions on Robotics*, 41:1841–1860, 2025.

[19] Max Muchen Sun, Allison Pinosky, and Todd Murphey. Flow Matching Ergodic Coverage. In *Proceedings of Robotics: Science and Systems*. 2025.

[20] Max Muchen Sun, Pete Trautman, and Todd Murphey. Inverse Mixed Strategy Games with Generative Trajectory Models. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, 2025.

[21] Muchen Sun, Francesca Baldini, Peter Trautman, and Todd Murphey. Move Beyond Trajectories: Distribution Space Coupling for Crowd Navigation. In *Proceedings of Robotics: Science and Systems*. 2021.

[22] Peter Trautman, Jeremy Ma, Richard M. Murray, and Andreas Krause. Robot navigation in dense human crowds: the case for cooperation. In *2013 IEEE International Conference on Robotics and Automation*, pages 2153–2160, 2013.

[23] David Vutetakis and Jing Xiao. Active perception network for non-myopic online exploration and visual surface coverage. *The International Journal of Robotics Research*, 44(2):247–272, 2025.

[24] Tony Z. Zhao, Vikash Kumar, Sergey Levine, and Chelsea Finn. Learning Fine-Grained Bimanual Manipulation with Low-Cost Hardware. In *Proceedings of Robotics: Science and Systems*, 2023.