Research Statement

Siddharth Chandak

Ph.D. Candidate, Electrical Engineering, Stanford University

How can we design distributed algorithms that ensure convergence to equilibria with global guarantees?

Answering this question can have sustained impact on engineering infrastructures, including large-scale communication, transportation, and computing networks. My work in **multi-agent systems** and **games** seeks to formulate and tackle this problem across multiple fronts. I study what constitutes *desirable* equilibria in different multi-agent settings. My research further includes game control, where the aim is to steer players toward an equilibrium that aligns with both local and global objectives. A central challenge in such settings is establishing convergence guarantees under limited and noisy feedback. Addressing this challenge requires analytical tools that motivate the second thread of my research:

*How can we analyze iterative algorithms across **optimization, reinforcement learning and distributed control** through a unified viewpoint?*

My research in **stochastic approximation** provides such a viewpoint, enabling analysis of a broad class of iterations under various noise models and multiple time-scales. Within this thread, I also work on reinforcement learning, designing new algorithms and establishing sharper finite-time guarantees.

Multi-Agent Learning and Control

Large-scale networks such as in communication, transportation, and computing consist of many interacting agents making autonomous decisions based on local information. These local decisions influence not only each agent's own performance but also the outcomes experienced by others, and can be naturally modeled using game theory, where each agent's decision affects the incentives and payoffs of other agents. Because agents lack a global view, such decentralized decision-making can lead to globally inefficient outcomes. In principle, a centralized authority can instruct the agents how to act to optimize a global objective. But such centralized architectures are typically infeasible due to substantial communication requirement, high computational complexity, and poor scalability. This motivates the need for designing distributed algorithms.

Game Control: Distributed algorithms that converge to a Nash Equilibrium (NE) have been extensively studied for various games. However an NE does not optimize global objectives and can lead to poor system performance. This gives rise to the problem of game control, where a manager adjusts system parameters so that the resulting NE aligns with a desired global objective. In [1], we studied strongly monotone games in which the manager controls a linear term in the players' utilities, and designed an algorithm that steers the system toward an NE satisfying prescribed linear constraints. For example, in resource allocation games, the global objective is load balancing, and the controlled linear term represents pricing or subsidies. In [2], we studied scenarios where the manager chooses from a discrete set of policies, each affecting the game's parameters, thereby the resulting NE. Our bandit-based algorithm enables the manager to identify the policy with the highest global objective at equilibrium. In resource allocation games, this corresponds to the manager controlling the set of resources available to each user. We are currently exploring how these techniques can be used for fine-tuning large machine learning models and selecting optimal hyperparameters.

Beyond Nash Equilibrium: In the vast literature focusing on NE, the inherent assumption is that players are focused solely on maximizing their local objective. However, in many real-world systems, agents are not selfish, and only require that their reward exceeds a minimum threshold, termed the Quality-of-Service (QoS) requirement. Examples include power control in wireless communication, where transmitters only want the signal-to-interference ration (SIR) to be sufficiently high. In general, achieving QoS guarantees can require significant coordination among players. In [3], we identified a class of continuous action games called *Tug-of-War games*, in which increasing one player's action decreases the reward for all others. We showed that a simple distributed algorithm converges to an action profile satisfying QoS requirements in these games. In [4], we extended this to *meta-games*, where multiple games run simultaneously, and each player chooses both the game to participate in and the action to take in it. A multi-channel wireless system is a natural example: each channel is a game, and each transmitter chooses a channel and the transmission power (action). Beyond QoS objectives, we studied a sensor network scenario in [5], where a set of sensors need to communicate a joint observation to the server through a shared medium. Unlike most existing medium access schemes, the goal is to ensure that the shared message gets through, regardless of the sender.

Reinforcement Learning via Stochastic Approximation

Stochastic Approximation (SA) is a class of iterative algorithms to find the fixed point of an operator given its noisy realizations. It provides a unified viewpoint for analyzing and developing algorithms across optimization, reinforcement learning (RL), and stochastic control.

Finite-Time Bounds in RL: Due to their applicability in RL, there has been growing interest in bounding the finite-time performance of SA algorithms. Because RL algorithms operate on Markov Decision Processes (MDPs), the noise in their SA formulation is *Markovian* rather than independent. We addressed this challenge in [6], obtaining concentration bounds for Q-Learning by using the Poisson equation to decompose the Markovian noise into a martingale difference component. We extended this analysis to the TD(0) algorithm in [7], using relaxed concentration inequalities to handle the possibility of unbounded iterates.

Two-Time-Scale Stochastic Approximation: Two-time-scale (TTS) SA algorithms involve two coupled iterations, each updating at a different rate, to find the fixed points of two coupled operators. My interest in analyzing these algorithms is closely connected to my research on game control, where players typically update their actions on a faster time-scale, and the manager updates the system parameters on a slower one. The algorithm we developed in [1] is in fact a two-time-scale iteration. In [8], we obtained the first mean square bound for TTS SA with arbitrary norm contractions, enabling the analysis of Stochastic Shortest Path (SSP) Q-Learning algorithm for average reward MDPs. In [9], I obtained an $O(1/k)$ mean square bound on TTS fixed-point iterations with contractive mappings, the first such bound without additional assumptions. While these works and most existing finite-time analyses focus on settings where both time-scales have contractive mappings, in [10] I analyzed iterations in which the slower time-scale is merely non-expansive, and showed how algorithms such as Lagrangian optimization can be incorporated into this framework.

RL beyond standard assumptions: I am also interested in studying how the behavior of classical algorithms changes and developing new algorithms when one or more standard assumptions of the RL framework are not satisfied. Traditionally, RL assumes that decision-makers behave rationally and maximize their expected cumulative reward. In [11], we studied how prospect-theoretic distortions of future returns modify the equilibrium that Q-learning converges to. While RL algorithms are designed for controlled Markov chains, many real-life systems are not Markovian and exhibit complex temporal dependencies. We studied RL for non-Markovian environments in [12], identifying the additional error when Q-Learning is applied in these environments, and designing an observation-to-state mapping aimed at minimizing this error.

Future Directions

My research in multi-agent learning and stochastic approximation spans several areas including game control, optimization, and reinforcement learning. In the near future, I aim both to deepen these lines of work and to unify them wherever possible, while also incorporating more realistic modeling considerations. In existing works on game control, the costs incurred by a manager for observing the underlying system or for communicating the updated system parameters are often not taken into account. I am currently exploring the resulting tradeoff between convergence speed and such costs. I am also interested in extending the meta-game framework from [4] further as many real-world systems allow agents to switch between different environments (or 'games') or to exit the system entirely. On the theoretical front, I am working on general-purpose results that can act as black-box tools for analyzing a broad class of stochastic approximation algorithms.

I am also motivated by how my research can support informed decision making in the real world. In [2], we studied how our model can be used for epidemic control, and how policy-makers can use our algorithm to identify the best course of action. Driven by this motivation, I have recently been working on **Remote Patient Monitoring (RPM)**. Advances in wearable medical devices, such as continuous glucose monitors and smartwatches, have enabled continuous monitoring of patients in their daily environments. A central challenge in RPM is determining the optimal intensity of patient monitoring, a decision that affects both patient burden and clinician workload. We modeled this problem using controlled Markov chains, and developed an approach for adaptively selecting monitoring intensity that achieves optimal performance while remaining practical and interpretable [13, 14, 15]. I aim to pursue this direction further, focusing on the practical challenges faced by clinicians in implementing RPM systems.

Going forward, I aim to broaden the scope of my research while deepening my understanding of the foundational tools of probability and optimization. My goal remains to design algorithms that meaningfully impact domains such as distributed training of machine learning models, communication networks, and transportation systems. I am keen to work more closely with practitioners so that my theoretical insights can translate into deployable and relevant solutions. This means not only designing algorithms using my core expertise but also paying attention to practical constraints, implementation details, and the system-level trade-offs that appear in real-world applications.

References

[1] S. Chandak, I. Bistritz, and N. Bambos, "Learning to control unknown strongly monotone games," *arXiv preprint arXiv:2407.00575*, 2024.

[2] S. Chandak, I. Bistritz, and N. Bambos, "Equilibrium bandits: Learning optimal equilibria of unknown dynamics," in *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*, pp. 1336–1344, 2023.

[3] S. Chandak, I. Bistritz, and N. Bambos, "Tug of peace: Distributed learning for quality of service guarantees," in *2023 62nd IEEE Conference on Decision and Control (CDC)*, pp. 2346–2351, 2023.

[4] S. Chandak, I. Bistritz, and N. Bambos, "Choose your battles: Distributed learning over multiple tug of war games," *arXiv preprint arXiv:2509.20147*, 2025.

[5] S. U. Haque, S. Chandak, F. Chiariotti, D. Günduz, and P. Popovski, "Learning to speak on behalf of

a group: Medium access control for sending a shared message," *IEEE Communications Letters*, vol. 26, no. 8, pp. 1843–1847, 2022.

[6] S. Chandak, V. S. Borkar, and P. Dodhia, "Concentration of contractive stochastic approximation and reinforcement learning," *Stochastic Systems*, vol. 12, no. 4, pp. 411–430, 2022.

[7] S. Chandak and V. S. Borkar, "A concentration bound for TD (0) with function approximation," *arXiv preprint arXiv:2312.10424*, 2023. *Accepted at Stochastic Systems.*

[8] S. Chandak, S. U. Haque, and N. Bambos, "Finite-time bounds for two-time-scale stochastic approximation with arbitrary norm contractions and Markovian noise," *arXiv preprint arXiv:2503.18391*, 2025. *Accepted at IEEE CDC 2025.*

[9] S. Chandak, "$O(1/k)$ finite-time bound for non-linear two-time-scale stochastic approximation," *arXiv preprint arXiv:2504.19375*, 2025.

[10] S. Chandak, "Non-expansive mappings in two-time-scale stochastic approximation: Finite-time analysis," *arXiv preprint arXiv:2501.10806*, 2025.

[11] V. S. Borkar and S. Chandak, "Prospect-theoretic Q-learning," *Systems & Control Letters*, vol. 156, p. 105009, 2021.

[12] S. Chandak, P. Shah, V. S. Borkar, and P. Dodhia, "Reinforcement learning in non-Markovian environments," *Systems & Control Letters*, vol. 185, p. 105751, 2024.

[13] S. Chandak, I. Thapa, N. Bambos, and D. Scheinker, "Tiered service architecture for remote patient monitoring," in *2024 IEEE International Conference on E-health Networking, Application & Services (HealthCom)*, pp. 1–7, IEEE, 2024.

[14] S. Chandak, I. Thapa, N. Bambos, and D. Scheinker, "Optimal control for remote patient monitoring with multidimensional health states," in *ICC 2025 - IEEE International Conference on Communications*, pp. 3186–3192, 2025.

[15] R. Tamizholi, S. Chandak, I. Thapa, N. Bambos, and D. Scheinker, "Learning adaptive control policies for remote patient monitoring," *Submitted to ICC 2026 - IEEE International Conference on Communications*, 2026.