Shaocong Ma | Research Statement

385-439-4778 • scma0908@umd.edu • mshaocong.github.io

My current research focuses on emerging topics at the intersection of machine learning, optimization, and the natural sciences. Specifically, I study a range of problems in reinforcement learning and stochastic optimization, with applications in control, large language models, and computational fluid dynamics.

Summary of Recent Research Projects

- **Stochastic Optimization Theory and Algorithms.**
  - *Adaptive Memory-Efficient Zeroth-Order Optimization.* In many modern machine learning applications, particularly those involving external physical simulators or large language models, explicit gradients of the objective function are either unavailable or prohibitively expensive to compute. This motivates the development of zeroth-order optimization algorithms that rely solely on function evaluations. However, such methods often suffer from high sample complexity, especially in high-dimensional and nonconvex settings. My research addresses these challenges by designing adaptive zeroth-order algorithms tailored to the intrinsic structure of specific tasks. We propose a novel stochastic perturbation technique, *Directionally Aligned Perturbations* (DAPs), which adaptively enhance accuracy along critical directions while maintaining the minimum variance characteristic of classical two-point estimators. The DAP method significantly improves optimization performance in tasks such as mesh optimization and language model fine-tuning, where gradients exhibit strong sparsity. Additionally, we observe that parameter-efficient fine-tuning (PEFT) methods often yield limited improvements in downstream tasks, while full-parameter fine-tuning is computationally expensive. To bridge this gap, we develop a hybrid fine-tuning strategy: applying first-order optimization to the PEFT module and zeroth-order optimization to the base model. However, this novel training paradigm leads to complex hybrid nonconvex landscapes. To address this new challenge, we establish a comprehensive convergence framework, achieving provably faster rates and reduced computational overhead compared to classical approaches. Extensive experiments on large language model fine-tuning demonstrate substantial improvements in both efficiency and scalability.

    Research Outcome: one publication on ICLR 2025 (Spotlight) [2]; one conference submission.
  - *Sample-Efficient Stochastic Optimization under Non-IID Data.* Stochastic gradient-type algorithms are fundamental tools in modern machine learning, particularly for large-scale data problems. A major challenge arises when data are not independently and identically distributed (non-IID), where standard convergence analyses often fail. My research focuses on understanding the convergence behavior of stochastic optimization algorithms in highly nonconvex settings under complex, non-IID sampling schemes, and on developing new algorithmic frameworks that improve sample and computational efficiency. In particular, I have established sharp convergence guarantees for stochastic gradient descent (SGD) under random reshuffling, a non-IID sampling scheme, demonstrating reduced optimization errors over standard SGD. I have also studied the convergence performance of online SGD algorithm under scenarios where data is highly dependent; we observed that the impact of data dependency will be reduced if we increase the batch size, which motivate us to leverage large batch sizes to accelerate convergence while maintaining statistical efficiency.

    Research Outcome: two publications at ICML 2020 [11] and UAI 2022 [8].
- **Reinforcement Learning Theory and Applications.**
  - *Variance-Reduced TD Learning and Q-Learning Algorithms.* Temporal difference (TD) learning and Q-learning are the most fundamental reinforcement learning (RL) algorithms for policy evaluation and policy optimization. But their convergence is known to suffer from a large optimization variance due to the stochastic and dependent samples queried from the dynamic environment. Conventional approaches address this issue by adopting a diminishing stepsize that significantly slows down the convergence in practice. In my research, we developed a two-timescale variance reduction scheme for the classic TD with correction algorithm for off-policy evaluation, and further generalized the two-timescale variance reduction scheme to the classic greedy-GQ algorithm for off-policy control.

For both algorithms, we developed new analysis tools that are based on a recursive refinement proof strategy, and utilized them to establish improved finite-time convergence rates and sample complexities of these algorithms under linear function approximation and Markovian sampling over the state-of-the-art results.

Research Outcome: two publications at NeurIPS 2020 [12] and ICLR 2021 [10]; a monograph [3].

◦ *Equilibrium of Markov Games.* Two-player zero-sum Markov games are fundamental problems in reinforcement learning and game theory, where players compete to achieve a certain equilibrium. Although many algorithms have been proposed to these problems, most of them either require full knowledge of the environment or lack sample efficiency. We develop a fully decentralized stochastic policy extragradient algorithm for solving tabular zero-sum Markov games with improved sample complexity, leveraging multiple stochastic estimators for accurate value estimation and entropy regularization to accelerate convergence. Constrained Markov games further generalize this setting by introducing behavioral constraints among multiple players. While the existence of Nash equilibrium has been established, computing it is PPAD-complete and thus infeasible in polynomial time. To address this, we propose a surrogate notion of correlated equilibrium (CE) for constrained Markov games, showing that its modification structure fundamentally differs from the unconstrained case, and prove that the associated Lagrangian function has zero duality gap. Based on these results, we develop the first primal-dual algorithm that provably converges to a CE in constrained Markov games. Moreover, classical Markov game formulations often ignore environment model uncertainty, which is ubiquitous in practice. To tackle this challenge, we propose a new notion of robust correlated equilibrium (robust CE) for Markov games under model uncertainty and prove that its characterization critically depends on the uncertainty set. We further develop a fully decentralized robust V-learning algorithm that computes robust CE with polynomial episode complexity, providing the first non-asymptotic convergence guarantee for robust multi-play general-sum Markov games under environment uncertainty.

Research Outcome: multiple publications at ICLR 2022 [9], NeurIPS 2022 [7], JMLR 2023 [4].

❖ **Machine Learning for Physical Science.**

◦ *Computational Fluid Dynamics.* Deep learning has been widely applied to solving partial differential equations (PDEs) in computational fluid dynamics. Recent research has introduced a PDE correction framework that leverages deep learning to improve solutions obtained from PDE solvers on coarse meshes. However, end-to-end training of such correction models over both solver-dependent parameters, such as mesh configurations, and neural network parameters requires the PDE solver to support automatic differentiation through its iterative numerical process, a feature not readily available in many existing solvers. In our study, we explore the feasibility of end-to-end training for a hybrid model that couples a black-box PDE solver with a deep learning model for fluid flow prediction. Specifically, we investigate a hybrid framework that integrates a black-box PDE solver into a differentiable deep graph neural network. To enable training, we employ a zeroth-order gradient estimator to approximate gradients via forward evaluations of the PDE solver. Experimental results demonstrate that the proposed zeroth-order approach produces correction models that outperform baseline models trained with first-order methods under a frozen mesh configuration. Extending this idea, we propose a multi-fidelity mixture-of-experts (MF-MoE) framework that combines a neural operator with solver-based hybrid models of varying fidelity. A physics-aware gating network dynamically selects the most suitable expert based on input characteristics, balancing computational cost and predictive accuracy. By formulating training as a constrained optimization problem, the model achieves fast inference for in-distribution samples while maintaining strong generalization to out-of-distribution cases. Experiments on fluid flow prediction demonstrate that MF-MoE consistently improves both efficiency and accuracy over baseline approaches.

Research Outcome: one publication at NeurIPS 2023 workshop [5]; a journal extension at Machine Learning in 2025 [1]; one conference submission.

Plans for Future Research

❖ **Machine Learning for Science and Engineering.**

In recent years, machine learning has proved its great success in solving challenging tasks in broad engineering and science areas, including robotics, communication, language processing, material science and neural science, etc. This has clearly demonstrate the great potential and wide applicability of machine learning, and will be a long-lasting trend in the next decade. Currently, I am expanding my machine learning research into other science and engineering fields, and developing specialized ML systems/algorithms for domain-specific applications. For example, I has been collaborating with scientists from the Livermore National Lab to develop hybrid machine learning systems that integrates black-box PDE solvers for accelerating simulations in computational fluid dynamics. These cross-discipline projects not only apply the existing ML techniques, but also inspire developing new ML algorithms. We envision that machine learning will continue to generate a big impact on future technology, education and society. In the future, we will continuously seek for interdisciplinary collaborations from other fields. We aim to develop more advanced machine learning techniques and optimization algorithms that exploit the knowledge in different domains and apply them to solve problems that can lead to a positive impact on the real world.

❖ **Memory Efficiency in Large-Scale Optimization and Reinforcement Learning.**

Large-scale optimization and reinforcement learning (RL) have become indispensable tools in modern machine learning, particularly for fine-tuning large generative models such as large language models (LLMs) and diffusion models. In practice, tasks like instruction tuning, dialogue safety alignment, and aesthetic preference optimization for diffusion models heavily rely on reinforcement learning from human feedback (RLHF) or similar preference alignment techniques. However, scaling these methods to multi-billion parameter models introduces severe memory challenges: fine-tuning requires storing not only large model parameters but also high-dimensional computational graphs and extensive reward models, leading to prohibitive memory overhead. Parameter-efficient fine-tuning (PEFT) methods alleviate part of the storage burden but fall short of addressing the optimization-specific memory costs inherent to gradient-based and RL-based fine-tuning, especially under large model sizes and long training horizons. To tackle these issues, we aim to develop memory-efficient optimization and reinforcement learning algorithms that are tailored for large-scale model fine-tuning and preference alignment tasks. Specifically, we will design lightweight policy optimization frameworks with compressed model representations, memory-adaptive optimization strategies, and efficient reward modeling techniques for both LLMs and diffusion models. Our goal is to reduce memory consumption without compromising fine-tuning quality or model performance. These algorithms will be supported by theoretical convergence guarantees and rigorously validated on large-scale benchmarks across language and vision domains, enabling scalable and resource-efficient optimization for next-generation generative models.

❖ **Multi-Agent Intelligence Systems and LLM-Based Agent Architectures.**

Reinforcement learning (RL) is a powerful framework for learning optimal control policies in complex environments. While traditional single-agent RL has been extensively studied, it does not capture the complexities of multi-agent systems where agents interact dynamically, following cooperative, competitive, or assistive patterns. Multi-agent RL extends beyond single-agent settings by modeling these interactions, with applications ranging from spectrum management in wireless networks to resource allocation in power systems. However, multi-agent RL introduces fundamental challenges such as scalability, communication efficiency, and stability under decentralized decision-making. In addition to conventional multi-agent systems, the emergence of LLM-based agent architectures—where LLMs are deployed as autonomous agents capable of reasoning, planning, and decision-making—has opened new directions for multi-agent RL research. These LLM agents must coordinate, compete, and assist each other in complex, often partially observable environments, requiring new algorithmic paradigms that blend classical RL with language-based reasoning and communication. In future work, we will systematically study these major categories of multi-agent systems, along with LLM-based agent interactions. Our focus will be on developing computation- and communication-efficient algorithms that are model-free, fully decentralized, and privacy-preserving. Specifically, for cooperative systems, we will design decentralized actor-critic algorithms for joint policy optimization; for competitive systems, we will develop policy optimization methods for finding Nash equilibria in zero-sum and general-sum games. For LLM agents, we will investigate novel RL algorithms that integrate language-based reasoning and decentralized coordination, backed by finite-time convergence analysis and complexity guarantees.

Funding Resources and Opportunities

In the future, I will actively apply for external funding from major federal funding agencies and military agencies. This includes the Career Development Programs of NSF, DOE and DARPA that support early-career faculty towards a lifetime commitment to leadership in education and research. I will also collaborate with the national labs to compete for fundings from DOE. In addition, I will actively stay connected with industries and seek for funded projects.

Publications

[1] Shaocong Ma, James Diffenderfer, Bhavya Kailkhura, et al. "Deep learning of PDE correction and mesh adaption without automatic differentiation". In: *Machine Learning* 114.61 (2025). https://doi.org/10.1007/s10994-025-06746-9.

[2] Shaocong Ma and Heng Huang. "Revisiting Zeroth-Order Optimization: Minimum-Variance Two-Point Estimators and Directionally Aligned Perturbations". In: *International Conference on Learning Representations (ICLR)*. **Spotlight**; https://openreview.net/forum?id=ywFOSIT9ik. 2025.

[3] Yi Zhou and Shaocong Ma. "Stochastic Optimization Methods for Policy Evaluation in Reinforcement Learning". In: *Foundations and Trends in Optimization* 6.3 (2024). https://www.nowpublishers.com/article/Details/OPT-045, pp. 145–192.

[4] Shaocong Ma, Ziyi Chen, Shaofeng Zou, and Yi Zhou. "Decentralized Robust V-learning for Solving Markov Games with Model Uncertainty". In: *The Journal of Machine Learning Research (JMLR)* (2023). https://www.jmlr.org/papers/v24/23-0310.html.

[5] Shaocong Ma, James Diffenderfer, Bhavya Kailkhura, and Yi Zhou. "End-to-End Mesh Optimization of a Hybrid Deep Learning Black-Box PDE Solver". In: *NeurIPS 2023 (ML4PS Workshop)* (2023). https://ml4physicalsciences.github.io/2023/files/NeurIPS_ML4PS_2023_8.pdf.

[6] Ziyi Chen, Shaocong Ma, and Yi Zhou. "Accelerated proximal alternating gradient-descent-ascent for nonconvex minimax machine learning". In: *2022 IEEE International Symposium on Information Theory (ISIT)*. https://ieeexplore.ieee.org/document/9834691. IEEE. 2022, pp. 672–677.

[7] Ziyi Chen, Shaocong Ma, and Yi Zhou. "Finding correlated equilibrium of constrained Markov game: A primal-dual approach". In: *Advances in Neural Information Processing Systems (NeurIPS)* 35 (2022). https://openreview.net/forum?id=2-CflpDkezH, pp. 25560–25572.

[8] Shaocong Ma, Ziyi Chen, Yi Zhou, Kaiyi Ji, and Yingbin Liang. "Data sampling affects the complexity of online sgd over dependent data". In: *Uncertainty in Artificial Intelligence (UAI)*. https://openreview.net/forum?id=Sb4xBLUsqx9. PMLR. 2022, pp. 1296–1305.

[9] Ziyi Chen, Shaocong Ma, and Yi Zhou. "Sample efficient stochastic policy extragradient algorithm for zero-sum markov game". In: *International Conference on Learning Representations (ICLR)*. https://openreview.net/forum?id=IvepFxYRDG. 2021.

[10] Shaocong Ma, Ziyi Chen, Yi Zhou, and Shaofeng Zou. "Greedy-GQ with Variance Reduction: Finite-time Analysis and Improved Complexity". In: *International Conference on Learning Representations (ICLR)*. https://openreview.net/forum?id=6t_dLShIUyZ. 2020.

[11] Shaocong Ma and Yi Zhou. "Understanding the impact of model incoherence on convergence of incremental SGD with random reshuffle". In: *International Conference on Machine Learning (ICML)*. https://proceedings.mlr.press/v119/ma20e.html. PMLR. 2020, pp. 6565–6574.

[12] Shaocong Ma, Yi Zhou, and Shaofeng Zou. "Variance-reduced off-policy TDC learning: Non-asymptotic convergence analysis". In: *Advances in Neural Information Processing Systems (NeurIPS)* 33 (2020). https://dl.acm.org/doi/10.5555/3495724.3496964, pp. 14796–14806.