Shaocong Ma | Research Statement

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

Gradient-based optimization algorithms have been highly successful in training modern machine learning models. However, as we move towards real-world applications, many challenges in training such a system arise, including limited GPU memory for training large language models (LLMs) with billion parameters, the absence of automatic differentiation in scientific software, and the presence of dependent or adversarial data in financial applications. To this end, my research focuses on developing optimization frameworks that consider these challenges and, in turn, designing robust and efficient algorithms with provable guarantees on real-world data. Specifically, my recent and forthcoming work can be summarized in the following three paradigms:

- *Optimization beyond Auto-Differentiation.* Auto-differentiation can be memory-consuming. Training billion-parameter LLMs requires storing large intermediate activations for backpropagation, often exceeding GPU memory. Zeroth-order optimization has recently emerged as a promising approach: instead of storing full gradients, each layer is updated using estimated gradients reconstructed from function values and layer-wise random perturbations, substantially reducing memory consumption. However, the limited accuracy of zeroth-order gradient estimators remains a key bottleneck for training large-scale machine learning models. My research addresses this challenge by developing a theoretical framework to analyze the accuracy of such estimators. In particular, my work on characterizing the variance of zeroth-order gradient estimators was recognized as a ***ICLR 2025 Spotlight*** [3]. In this work, we derived the minimum-variance condition for existing gradient estimators and proposed the *directionally aligned perturbations*, a novel technique that can adaptively selects the most important parameters to update. My subsequent work, accepted as a ***NeurIPS 2025 Spotlight*** [1], focuses on the inherent bias in existing gradient estimators. We designed a new class of gradient estimators that completely mitigates bias for zeroth-order gradient estimators. These works offer an improved method for constructing appropriate gradient estimators, making zeroth-order optimization a practical approach for fine-tuning modern machine learning models.

  Besides the memory consideration, auto-differentiation can also be unavailable. Many scientific simulators (*e.g.*, computational fluid dynamics and airfoil aerodynamics) are legacy codes that lack built-in auto-differentiation support. My work, ***collaborated with Lawrence Livermore National Laboratory***, focuses on developing an end-to-end machine learning pipeline that jointly optimizes deep graph neural network as well as parameters within external scientific simulators [5]. Our approach leverages state-of-the-art zeroth-order optimization techniques, enabling seamless integration with these simulators and accelerating scientific discovery.

- *Optimization under Data Dependence.* Classical stochastic gradient-based optimization methods typically assume *i.i.d.* data, which rarely holds in practice. In real-world examples, data streams in domains such as financial markets or robotic control usually exhibit temporal dependence. My research develops optimization methods that explicitly address this issue. One direction focuses on establishing the theoretical foundations for different types of data dependence, including nontrivial sampling schemes [13] and mixing processes [10]. Another direction aims to improve traditional approaches in these challenging settings, such as accelerating the training process through variance-reduction techniques [6, 12, 14].

- *Optimization with Adversarial Data.* Distributionally robust optimization has emerged as a critical framework for addressing adversarial data distributions, with broad applications in adversarial attacks, robust reinforcement learning, and game theory. For instance, reinforcement learning (RL) agents often operate in uncertain environments where resilience to worst-case scenarios is essential. This motivates my research on developing robust algorithms that explicitly account for adversarial conditions, thereby enhancing the reliability of RL agents through distributionally robust optimization methods [2, 4]. Furthermore, in stochastic games, competitive opponents inherently generate adversarial signals. Learning in such challenging environments motivates the development of more sophisticated equilibrium theories [7, 8, 9, 11].

These three paradigms not only capture the core challenges that motivate my research, but also align closely with the projects I have pursued in my past work. In the next section, I will highlight how my previous research has laid the foundation for addressing these challenges.

Past Research

Thrust 1: Optimization beyond Auto-Differentiation

In many modern machine learning applications, particularly those involving large language models or external physical simulators, explicit gradients of the objective function are either prohibitively expensive to compute or even unavailable. This scenario motivates the development of zeroth-order optimization algorithms that rely solely on function evaluations. However, such methods often suffer from high sample complexity due to the inaccuracy of gradient estimators.

- *Improved Accuracy for Memory-Efficient Optimization.* To address the inaccuracy of zeroth-order optimization, we explicitly analyze the variance and bias of gradient estimators that rely solely on function evaluations, deriving optimal estimators with improved accuracy. Specifically, in our work [3], we establish the minimum-variance condition for zeroth-order gradient estimators and introduce a novel stochastic perturbation technique, *Directionally Aligned Perturbations* (DAPs). DAPs adaptively enhance accuracy along critical directions while retaining the minimum variance property of classical two-point estimators. Furthermore, we introduce a trackable algorithm for deriving zeroth-order gradient estimators with DAPs. Empirical results demonstrate that our method consistently yields significant performance improvements in tasks such as mesh optimization and language model fine-tuning, particularly in settings where gradients are highly sparse. Theoretically, our method achieves optimal complexity, matching the best known lower bound. This work was selected as an ICLR 2025 Spotlight. Besides the variance, we also analyze the inherent bias in existing zeroth-order gradient estimators, showing that it is unavoidable due to the fundamental design of random smoothing. To overcome this limitation, we propose a new class of unbiased estimators [1]. By reformulating the directional derivative of the objective function as the expectation form, we derive the unbiased condition and develop a bias-free estimation technique. This approach also achieves optimal complexity under our proposed closed-form hyper-parameter setting, matching the theoretical complexity lower bound. This work was selected as a NeurIPS 2025 Spotlight. Overall, this line of research pushes the boundary of gradient-free stochastic optimization beyond the conventional analysis. By systematically characterizing both the variance and bias of zeroth-order gradient estimators, we provably obtain the optimal condition to minimize the variance and the bias, making the memory-efficient optimization without explicit gradients only theoretically sound but also practically powerful.

- *End-to-End Optimization for Scientific Software.* Deep learning has been widely applied to solving partial differential equations (PDEs) in computational fluid dynamics (CFDs). 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 auto-differentiation through its iterative numerical process, a feature not readily available in many existing solvers. In my collaborated work with Lawrence Livermore National Laboratory [5], we explore the feasibility of end-to-end training for a hybrid model that couples a black-box PDE solver with a deep graph neural network model for fluid flow prediction. 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. This line of research offers a practical pathway to integrate cutting-edge machine learning techniques with traditional scientific computing workflows. Whereas optimizing parameters within classical scientific pipelines has often been regarded as computationally infeasible, our framework demonstrates that such challenges can be overcome. By bridging black-box solvers with gradient-free training strategies, we provide a feasible route to end-to-end optimization in domains where differentiable solvers are unavailable, thereby expanding the reach of modern machine learning in scientific discovery.

Thrust 2: Optimization under Data Dependence

In many real-world machine learning scenarios, particularly in financial markets, robotics, and sequential decision-making, the assumption of *i.i.d.* data is often violated. Instead, data typically arrive in dependent streams with temporal, spatial, or structural correlations. This dependence introduces significant challenges for optimization, as classical stochastic gradient methods may fail to converge efficiently or even produce misleading updates. These issues motivate the development of new optimization frameworks that explicitly account for dependent data.

- *Stochastic Optimization Theory under Non-i.i.d. Data.* To address this non-*i.i.d.* issue, my research focuses on understanding the convergence behavior and on developing new algorithmic frameworks that improve sample and computational efficiency. In particular, we established sharp convergence guarantees for stochastic gradient descent (SGD) under random reshuffling, a non-*i.i.d.* sampling scheme, showing that it achieves smaller optimization errors compared to standard SGD [13]. We also analyzed the convergence of online SGD under scenarios where data is highly dependent [10]. We observed that data dependency negatively affects the convergence of the online SGD algorithm; however, this impact can be mitigated by increasing the batch size. This insight motivates us to exploit large batches as a tool to accelerate convergence while preserving statistical efficiency.

- *Stochastic Variance-Reduced Optimization under Markovian Data.* Another line of my research focuses on accelerating training in the challenging setting of dependent data. Temporal difference (TD) learning and Q-learning are two of the most fundamental reinforcement learning (RL) algorithms for policy evaluation and policy optimization. Both can be naturally reformulated as gradient-based optimization problems over Markovian data, giving rise to the TD with correction (TDC) algorithm and the Greedy-GQ algorithm. However, their convergence is known to suffer from high optimization variance due to the stochastic and dependent nature of samples generated by dynamic environments. In my research, we developed a two-timescale variance-reduction scheme for the classic TDC algorithm in the off-policy evaluation setting [14], and further extended this framework to the Greedy-GQ algorithm [12]. For both cases, we introduced new analytical tools based on a recursive refinement proof strategy, which enabled us to establish sharper finite-time convergence rates and improved sample complexities under linear function approximation and Markovian sampling. These results are also collected in a monograph I was invited to write [6].

Thrust 3: Optimization with Adversarial Data

In many practical machine learning scenarios, data may not only be dependent but also adversarial. That is, the data distribution can be intentionally perturbed or potentially hostile due to the presence of competitive agents, adversarial attacks, or environmental uncertainties. Such settings are particularly prevalent in reinforcement learning, where agents must remain robust against worst-case transitions, as well as in multi-agent games, where competitive opponents naturally generate adversarial signals.

- *Robust Reinforcement Learning (RL).* RL agents often operate in uncertain environments where resilience to worst-case scenarios is critical. This goal is typically modeled as the distributionally robust optimization (DRO) problem, where the data generated through the agent-environment iteration could be adversarial to our ultimate objectives. Following this direction, my research develops robust policy optimization methods that explicitly account for adversarial perturbations in transition dynamics and reward structures. In particular, we proposed robust RL algorithms that achieve strong performance in worst-case environments while satisfying safety constraints [4]. This work provides the first finite-time convergence analysis for robust RL under safety constraints. Furthermore, we introduced a structured robustness framework tailored to the directional nature of financial markets [2]. Unlike traditional ambiguity sets, which are often overly conservative, our formulation incorporates prior information about market dynamicsfor example, buying an asset is more likely to increase its price rather than decrease it. By embedding such structure, our approach offers a more precise representation of environmental uncertainty. Both theoretical analysis and empirical results validate that this framework mitigates conservatism and enables more efficient policy learning. These two works significantly enhance the reliability of RL systems in safety-critical domains, particularly in finance and autonomous control.

- *Equilibrium of Markov Games.* In stochastic and dynamic games, the presence of competitive opponents inherently generates adversarial signals, which presents a critical challenging for traditional optimization or reinforcement learning algorithms. My research develops new optimization frameworks for learning in such environments, addressing both theoretical and algorithmic challenges. On the theoretical side, we study sample-efficient equilibrium evaluation and establish convergence guarantees for stochastic games [9, 11]. On the algorithmic side, we design accelerated methods for min-max optimization, a fundamental tool for two-player zero-sum games, achieving faster convergence to equilibrium strategies [8]; in the empirical experiments of this work, we treat the Wasserstein-robust model as a zero-sum game between an attacker and a defender, which reduces to a standard min-max optimization problem. Extending this line of work to settings with uncertainty in agent-environment interactions, we proposed a fully decentralized robust RL algorithm that computes robust correlated equilibria with polynomial episode complexity [7]. This work provides the first non-asymptotic convergence guarantee for robust multi-player general-sum Markov games under environment uncertainty and was published on the top machine learning journal, Journal of Machine Learning Research (JMLR).

Future Directions

The advancement of artificial intelligence (AI) depends critically on the development of scalable optimization algorithms that can handle increasingly complex systems. While tremendous progress has been made in large-scale optimization and learning, significant challenges remain in improving efficiency, robustness, and generalization across tasks, especially for agentic systems or scientific pipelines. During my Ph.D. research, I laid the foundations for addressing these challenges by analyzing the theoretical limitations of existing approaches and proposing optimization techniques tailored to each specific domain, including physical simulations, financial applications, and LLM fine-tuning. Building upon this foundation, my future research agenda will focus on designing new optimization frameworks that can push the boundary of AI systems. These directions include:

- **Optimization for Scientific Pipelines.**

  Machine learning has demonstrated remarkable potential in solving complex problems across physics, materials science, and engineering. However, many of these domains require specialized optimization techniques to integrate prior scientific knowledge and accelerate computationally expensive simulations. For example, PDE-constrained optimization arises naturally in computational fluid dynamics, gene expression, and material discovery, but existing ML approaches fail to capture domain-specific structures such as the mesh configuration, combinatorial structure in genes, and the symmetry in chemical materials. In my collaboration with scientists at Livermore National Lab, I have explored hybrid optimization frameworks that integrate physical information with graph-based components to accelerate simulation and control [5], the external physical information plays a critical role in accelerating the convergence and improving the generalization ability. Going forward, I aim to design optimization algorithms that explicitly exploit scientific problem structure, balance exploration and exploitation in scientific settings, and provide convergence guarantees under noisy or partially observed environments. These efforts will enable machine learning to play a transformative role in advancing scientific discovery and engineering design.

- **Optimization for Agentic AI Architectures.**

  The rise of agentic AI systems, where autonomous agents interact to accomplish complex tasks, creates new challenges for optimization. Unlike static learning settings, these architectures involve dynamic and stochastic interactions among agents and tools with complicated relations including cooperation, competition, and information exchange. Designing optimization techniques for such settings requires handling high-dimensional, non-stationary, and partially observed environments. Zeroth-order optimization methods, which do not require explicit gradient information, are particularly promising for this domain as they can adapt to non-differentiable objectives and stochastic dynamics. My future work will focus on developing principled zeroth-order and hybrid optimization strategies for learning over dynamic stochastic graphs, enabling scalable training of multi-LLM-agent architectures. This research has the potential to impact a broad spectrum of AI applications, ranging from automated scientific discovery and formal proof verification to personalized legal assistance and medicine.

- **Equilibrium Theory of Multi-Agent Intelligence Systems.**

  Despite rapid advances in multi-agent learning, there is currently no unifying theoretical framework for characterizing the long-run behavior of trainable agentic systems. In particular, existing analyses are either problem-specific or limited to simplified equilibria that fail to capture the complexity of modern agent-based architectures, such as those involving LLM-driven agents. My future research will focus on developing an equilibrium theory of multi-agent intelligence, aimed at describing the stationary and asymptotic properties of interacting trainable agents. This involves extending tools from game theory, dynamical systems, and statistical physics to understand convergence, stability, and emergent cooperation/competition in large-scale agent populations. Such a theory would provide rigorous guarantees for system-wide behavior, guiding the design of robust, predictable, and trustworthy AI ecosystems that go beyond single-agent optimization.

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** and Heng Huang. "On the Optimal Construction of Unbiased Gradient Estimators for Zeroth-Order Optimization". In: *Advances in Neural Information Processing Systems (NeurIPS)* (2025). *NeurIPS 2025 Spotlight*.

[2] **Shaocong Ma** and Heng Huang. "Robust Reinforcement Learning in Finance: Modeling Market Impact with Elliptic Uncertainty Sets". In: *Advances in Neural Information Processing Systems (NeurIPS)* (2025).

[3] **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)* (2025). *ICLR 2025 Spotlight*.

[4] **Shaocong Ma**, Ziyi Chen, Yi Zhou, and Heng Huang. "Rectified Robust Policy Optimization for Model-Uncertain Constrained Reinforcement Learning without Strong Duality". In: *TMLR* (2025).

[5] **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).

[6] Yi Zhou and **Shaocong Ma**. "Stochastic Optimization Methods for Policy Evaluation in Reinforcement Learning". In: *Foundations and Trends in Optimization* 6.3 (2024), pp. 145–192.

[7] **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).

[8] 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)* (2022), pp. 672–677.

[9] 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), pp. 25560–25572.

[10] **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)* (2022), pp. 1296–1305.

[11] 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)* (2021).

[12] **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)* (2020).

[13] **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)* (2020), pp. 6565–6574.

[14] **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), pp. 14796–14806.