

Distributed Stochastic Nested Optimization for Emerging Machine Learning Models: Algorithm and Theory

Hongchang Gao

Department of Computer and Information Sciences
Temple University, PA, USA
hongchang.gao@temple.edu

Abstract

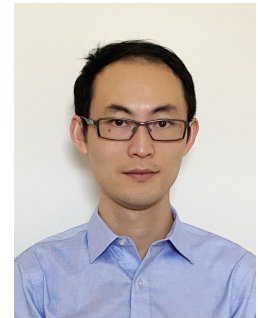
Traditional machine learning models can be formulated as the expected risk minimization (ERM) problem: $\min_{\mathbf{w} \in \mathbb{R}^d} \mathbb{E}_{\xi}[\ell(\mathbf{w}; \xi)]$, where $\mathbf{w} \in \mathbb{R}^d$ denotes the model parameter, ξ represents training samples, $\ell(\cdot)$ is the loss function. Numerous optimization algorithms, such as stochastic gradient descent (SGD), have been developed to solve the ERM problem. However, a wide range of emerging machine learning models are beyond this class of optimization problems, such as model-agnostic meta-learning (Finn, Abbeel, and Levine 2017). Of particular interest of my research is the *stochastic nested optimization (SNO)* problem, whose objective function has a nested structure. Specifically, I have been focusing on two instances of this kind of problem: *stochastic compositional optimization (SCO)* problems, which cover meta-learning, area-under-the-precision-recall-curve optimization, contrastive self-supervised learning, etc., and *stochastic bilevel optimization (SBO)* problems, which can be applied to meta-learning, hyperparameter optimization, neural network architecture search, etc.

With the emergence of large-scale distributed data, such as the user data generated on mobile devices or intelligent hardware, it is imperative to develop distributed optimization algorithms for SNO (**Distributed SNO**). A significant challenge for optimizing distributed SNO problems lies in that the stochastic (hyper-)gradient is a biased estimation of the full gradient. Thus, existing distributed optimization algorithms when applied to them suffer from slow convergence rates. In this talk, I will discuss my recent works about distributed SCO (Gao and Huang 2021; Gao, Li, and Huang 2022) and distributed SBO (Gao, Gu, and Thai 2022; Gao 2022) under both centralized and decentralized settings, including algorithmic details about reducing the bias of stochastic gradient, theoretical convergence rate, and practical machine learning applications, and then highlight challenges for future research.

Biography

Dr. Hongchang Gao is an assistant professor in the department of computer and information sciences at Temple University. Dr. Gao received his Ph.D. degree in

Electrical and Computer Engineering from University of Pittsburgh in 2020. His expertise lies in the intersection of machine learning and optimization, as well as their applications in data mining. In particular, his recent research focuses on distributed optimization and federated learning. For instance, he developed the first decentralized optimization algorithm for distributed SCO problems with theoretical guarantees. Dr. Gao has published his research in prestigious AI journals and conference proceedings, e.g., ICML, NeurIPS, AAAI, IJCAI, KDD, etc. Currently, he is serving as the Associate Editor of the Journal of Combinatorial Optimization. He also has been serving multiple top machine learning and data mining conferences and journals as program committee members and organizing committee members.



References

- Finn, C.; Abbeel, P.; and Levine, S. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning*, 1126–1135. PMLR.
- Gao, H. 2022. On the Convergence of Momentum-Based Algorithms for Federated Stochastic Bilevel Optimization Problems. *arXiv preprint arXiv:2204.13299*.
- Gao, H.; Gu, B.; and Thai, M. T. 2022. Stochastic Bilevel Distributed Optimization over a Network. *arXiv preprint arXiv:2206.15025*.
- Gao, H.; and Huang, H. 2021. Fast Training Method for Stochastic Compositional Optimization Problems. *Advances in Neural Information Processing Systems*, 34: 25334–25345.
- Gao, H.; Li, J.; and Huang, H. 2022. On the Convergence of Local Stochastic Compositional Gradient Descent with Momentum. In *International Conference on Machine Learning*, 7017–7035. PMLR.