Research Statement

Zhize Li
School of Computing and Information Systems, Singapore Management University
Tel: (65) 6828-0923; Email: zhize@smu.edu.sg
20 December 2024

Background

My research interests span a wide variety of topics in **optimization**, **federated learning**, **machine learning** and **data science**. Machine learning problems are usually modeled as an optimization problem, ranging from convex (e.g., linear/lasso/logistic regression, support vector machine) to nonconvex optimization (e.g., matrix completion/recovery, deep neural networks). During my PhD, I mainly aim to provide efficient optimization algorithms with better theoretical guarantees for machine learning problems (e.g., [6, 8, 15, 9, 10, 3, 7] in Figure 3). In particular, my thesis "Simple and Fast Optimization Methods for Machine Learning" won the **Tsinghua Outstanding Doctoral Dissertation Award** in 2019.

With the proliferation of mobile and edge devices, Federated Learning (FL) has recently emerged as a disruptive paradigm for training large-scale machine learning models over a vast amount of geographically distributed and heterogeneous devices (see Figure 1). However, there are several challenges in FL such as **data privacy**, **data heterogeneity**, **communication efficiency**, **system resiliency**, etc. After my PhD graduation (2019 – Now), I have also developed an intense interests and contributed to the foundations of FL (e.g., [16, 4, 18, 14, 11, 19, 5] in Figure 2).

My research mainly aims to **propose simple and efficient algorithms** with **better theoretical analyses/guarantees** for optimization and FL problems, and **formulate new problems** (and develop new algorithms for them) **characterized the challenges** from important real-world applications. In the following sections, I will summarize my research work organized in two parts: 1) federated learning; 2) optimization for machine learning.

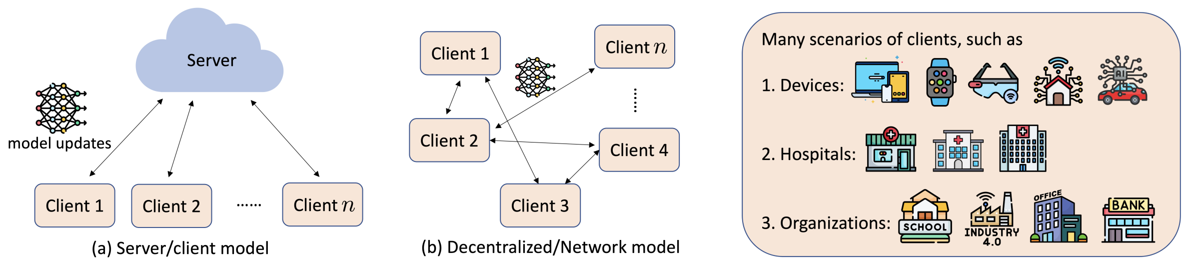

Figure 1: Federated Learning (FL) framework

Federated Learning

As demonstrated in Figure 1, the proliferation of multi-agent environments in emerging real-world applications has attracted significant attention on the distributed/federated learning problems. There are many challenges in FL and I will describe my contributions addressing three of them (i.e., **privacy**, **communication efficiency**, and **resiliency**) as follows:

Privacy. Data privacy is very important and related to everyone. Although FL may appear to protect data privacy via storing data locally and only sharing the model updates (e.g., gradient information), the training process of FL can nonetheless reveal sensitive information [20]. Recently in [16], we propose a unified framework for private FL with compression and achieve the state-of-the-art trade-offs in terms of private-utility-communication.

Communication efficiency. The communication of messages across a network forms one of the main bottlenecks of the FL training system. One principal approach is to use compression. In [4, 18], we provide the state-of-the-art results for nonconvex problems using compression. However, the compression usually leads to more communication rounds. Thus in [14, 11], we borrow the acceleration idea from optimization community and combine with compression to further improve the communication complexity.

Resiliency. Data samples collected from different clients/agents/parties can be highly unbalanced and heterogeneous. In [19], we remove the strong bounded dissimilarity assumption (thus allows arbitrary heterogeneity) and also achieve faster convergence result. In [5], we provide the first coreset framework via distributed importance sampling for communication-efficient vertical federated learning.

Figure 2: Challenges in FL and our contributions

Optimization for Machine Learning

The optimization problems usually are formulated as $\min_{\boldsymbol{x}\in\mathbb{R}^d} \left\{\Phi(\boldsymbol{x}) := f(\boldsymbol{x}) + h(\boldsymbol{x})\right\}$, where $\boldsymbol{x}$ is the machine learning (ML) model parameters, $f(\boldsymbol{x})$ denotes the loss function (convex or nonconvex), and $h(\boldsymbol{x})$ denotes the regularization term. In particular, the finite-sum form $f(\boldsymbol{x}) := \frac{1}{n}\sum_{i=1}^n f_i(\boldsymbol{x})$ captures the empirical risk minimization in ML, where there are n data samples and f_i denotes the loss on data i. Due to the increasing service of modern machine/deep learning models, this optimization problem has been extensively studied in recent years. Now I describe my contributions solving three kinds of ML optimization problems as follows:

Large-scale machine learning problems. To deal with ML problems with big data, it is important to develop efficient optimization algorithms to accelerate the ML training convergence and obtain better guarantees. In [6], we provide the first direct accelerated stochastic method which improves the convergence **by a factor of** $\sqrt{n}$, where n is the dataset size. Hence the improvement is significant especially for large-scale problems. In the follow-up [8], we further improve our convergence result to **optimal** for *convex* large-scale problems. In [15], we give **optimal** result for the *nonconvex* setting.

Nonsmooth nonconvex problems. To handle the nonsmoothness and nonconvexity in ML problems, in [9], we provide a simple proximal stochastic gradient with novel convergence analysis to improve the proximal batch gradient method **by a factor of** $\sqrt{b}$ (where b is the minibatch size) while maintaining the same proximal complexity, which solves the open problem posed by [17]. In the follow-up [10], we further improve our result to **optimal** for nonsmooth nonconvex problems.

Escaping saddle points. To avoid bad unstable saddle points and find approximate local minima for ML problems, in [3], we provide the first simple and stabilized variance-reduced gradient method with random perturbations. In the follow-up [7], we further improve our result **from** $n^{2/3}$ **to** $\sqrt{n}$ via recursive gradient, where n is the dataset size.

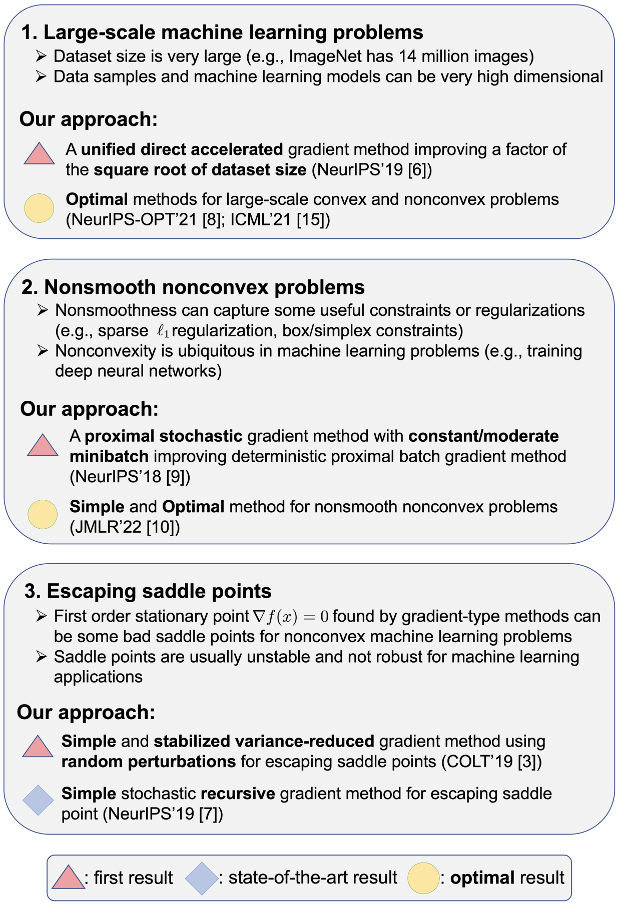

Figure 3: ML optimization and our contributions

References

[1] Wei Cao, Jian Li, Yufei Tao, and **Zhize Li**. On top-k selection in multi-armed bandits and hidden bipartite graphs. In *Advances in Neural Information Processing Systems (NIPS)*, 2015.

[2] Rong Ge, Rohith Kuditipudi, **Zhize Li**, and Xiang Wang. Learning two-layer neural networks with symmetric inputs. In *International Conference on Learning Representations (ICLR)*, 2019.

[3] Rong Ge, **Zhize Li**, Weiyao Wang, and Xiang Wang. Stabilized SVRG: Simple variance reduction for nonconvex optimization. In *Conference on Learning Theory (COLT)*, 2019.

[4] Eduard Gorbunov, Konstantin Burlachenko, **Zhize Li**, and Peter Richtárik. MARINA: Faster non-convex distributed learning with compression. In *International Conference on Machine Learning (ICML)*, 2021.

[5] Lingxiao Huang, **Zhize Li**, Jialin Sun, and Haoyu Zhao. Coresets for vertical federated learning: Regularized linear regression and K-means clustering. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

[6] Guanghui Lan, **Zhize Li**, and Yi Zhou. A unified variance-reduced accelerated gradient method for convex optimization. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[7] **Zhize Li**. SSRGD: Simple stochastic recursive gradient descent for escaping saddle points. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[8] **Zhize Li**. ANITA: An optimal loopless accelerated variance-reduced gradient method. *arXiv preprint arXiv:2103.11333*, 2021.

[9] **Zhize Li** and Jian Li. A simple proximal stochastic gradient method for nonsmooth nonconvex optimization. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2018.

[10] **Zhize Li** and Jian Li. Simple and optimal stochastic gradient methods for nonsmooth nonconvex optimization. *Journal of Machine Learning Research (JMLR)*, 23(239):1–61, 2022.

[11] **Zhize Li** and Peter Richtárik. CANITA: Faster rates for distributed convex optimization with communication compression. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

[12] **Zhize Li**, Le Zhang, Zhixuan Fang, and Jian Li. A two-stage mechanism for ordinal peer assessment. In *International Symposium on Algorithmic Game Theory (SAGT)*, 2018.

[13] **Zhize Li**, Tianyi Zhang, Shuyu Cheng, Jun Zhu, and Jian Li. Stochastic gradient Hamiltonian Monte Carlo with variance reduction for Bayesian inference. *Machine Learning*, 108(8):1701–1727, 2019.

[14] **Zhize Li**, Dmitry Kovalev, Xun Qian, and Peter Richtárik. Acceleration for compressed gradient descent in distributed and federated optimization. In *International Conference on Machine Learning (ICML)*, 2020.

[15] **Zhize Li**, Hongyan Bao, Xiangliang Zhang, and Peter Richtárik. PAGE: A simple and optimal probabilistic gradient estimator for nonconvex optimization. In *International Conference on Machine Learning (ICML)*, 2021.

[16] **Zhize Li**, Haoyu Zhao, Boyue Li, and Yuejie Chi. SoteriaFL: A unified framework for private federated learning with communication compression. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

[17] Sashank J Reddi, Suvrit Sra, Barnabás Póczos, and Alexander J Smola. Proximal stochastic methods for nonsmooth nonconvex finite-sum optimization. In *Advances in Neural Information Processing Systems (NIPS)*, 2016.

[18] Peter Richtárik, Igor Sokolov, Elnur Gasanov, Ilyas Fatkhullin, **Zhize Li**, and Eduard Gorbunov. 3PC: Three point compressors for communication-efficient distributed training and a better theory for lazy aggregation. In *International Conference on Machine Learning (ICML)*, 2022.

[19] Haoyu Zhao, Boyue Li, **Zhize Li**, Peter Richtárik, and Yuejie Chi. BEER: Fast $O(1/T)$ rate for decentralized nonconvex optimization with communication compression. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

[20] Ligeng Zhu, Zhijian Liu, and Song Han. Deep leakage from gradients. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.