Yuxing Liu

217-419-9692 yuxing6@illinois.edu Homepage LinkedIn

Education

University of Illinois Urbana-Champaign **Aug. 2024 – Present**
Ph.D. student in Computer Science, advised by Prof. Tong Zhang *Illinois, U.S.*

Fudan University **Sep. 2020 – Jun. 2024**
B.S. in Data Science *Shanghai, China*

Research Interests

My research interests mainly lie in the intersection of machine learning and optimization. Currently, I am particularly interested in improving the efficiency of training modern machine learning models (e.g., LLMs) and exploring valid explanations for the effectiveness of practical optimizers and training settings.

Selected Publications

(* denotes equal contribution)

[1] ASGO: Adaptive Structured Gradient Optimization.
Kang An*, **Yuxing Liu***, Rui Pan, Yi Ren, Shiqian Ma, Donald Goldfarb, and Tong Zhang. [NeurIPS 2025]

[2] Adagrad under Anisotropic Smoothness.
Yuxing Liu*, Rui Pan*, and Tong Zhang. [ICLR 2025]

[3] Decentralized Convex Finite-Sum Optimization with Better Dependence on Condition Numbers.
Yuxing Liu, Lesi Chen, and Luo Luo. [ICML 2024]

[4] Accelerated Convergence of Stochastic Heavy Ball Method under Anisotropic Gradient Noise.
Rui Pan*, **Yuxing Liu***, Xiaoyu Wang, and Tong Zhang. [ICLR 2024]

Research

ASGO: Adaptive Structured Gradient Optimization **Dec. 2024 – Present**
Joint work with Kang An, Rui Pan, Ren Yi, Shiqian Ma, Donald Goldfarb and Tong Zhang *In NeurIPS 2025*

- We proposed ASGO, an optimizer that leverages the structured properties such as low-rank gradients and block-wise diagonal Hessians. We theoretically analyzed the convergence of ASGO and showed better convergence results than other matrix-preconditioned methods like Shampoo.
- We further demonstrated ASGO's effectiveness in exploiting gradient low-rankness and approximate Hessian block-wise diagonality, both typically observed in neural networks, highlighting ASGO's potential in real-world applications.
- We developed a practical implementation of ASGO with targeted modifications to enable a more efficient training. The effectiveness of ASGO and the modifications are empirically validated on Transformer pretraining tasks, showing improved performance over AdamW and Shampoo and comparable results to Muon.
- We are currently working on further improving the efficiency and performance of ASGO and verifying its performance on larger foundation models and more practical tasks.

Adagrad under Anisotropic Smoothness **Nov. 2023 – Sep. 2024**
Joint work with Rui Pan and Tong Zhang *In ICLR 2025*

- We attempted to theoretically explain the benefits of adaptive gradient methods over classical gradient methods with uniform step sizes across all coordinates.
- We analyzed AdaGrad under the anisotropic smoothness and noise assumptions. We further extended the results to more practical settings by introducing a generalized form of anisotropic smoothness.
- We discussed how the convergence results indicate the potential benefits of AdaGrad compared to classical gradient methods in terms of better dimensional dependence, which was also verified by multiple experiments.

Momentum Accelerates SGD in Large Batch Settings **Jun. 2023 – Sep. 2023**
Joint work with Rui Pan and Tong Zhang *In ICLR 2024*

- We aimed to theoretically explain the benefits of SGD with Heavy Ball Momentum (SHB) over vanilla SGD.
- We developed and applied novel analysis techniques on bounding non-commutative matrix products in order to prove convergence results when a decaying step size scheduler is employed.
- We proved that SHB achieves an overall (near)-minimax convergence rate and converges within fewer iterations than SGD on quadratic objectives under the large batch setting.

Complexity of Decentralized Optimization **Mar. 2023 – Jan. 2024**
Joint work with Lesi Chen and Luo Luo *In ICML 2024*

- We proposed a novel algorithm with better time complexity in the decentralized finite-sum optimization setting by improving the dependence on local condition number to global condition number.
- The key design of the algorithm is the novel sampling method when applying variance reduction techniques, which also led to a much smaller total first-order oracle complexity of all agents compared to existing methods.
- We proved that the time complexity results are near-optimal with respect to the condition numbers.

Working Experiences

Bytedance Doubao (Seed) **Jun. 2025 – Aug. 2025**

Student Researcher (Doubao (Seed) - Model Architecture Optimization) *California, US*

- Worked on developing and testing novel optimizers (similar to or beyond the current workhorse AdamW) for pretraining large language models.
- Developed implementations for novel optimizers for both open-source and internal projects.
- Studied the scaling behaviour of optimizers across a large range of models and training sizes.
- Conducted comprehensive analyses on the performance and efficiency of optimizers and attempted to enhance the understanding of optimizers from numerical observations.

Hong Kong University of Science and Technology **Aug. 2023 – Sep. 2023**

Undergraduate Visiting Internship Student, Advised by Prof. Tong Zhang *Hong Kong, China*

Guangdong Yuecai Holdings **Aug. 2022 – Sep. 2022**

Assistant in Apartment of Equity Investment *Guangdong, China*

Technical Skills

Programming: Python, Pytorch, MATLAB, R, C/C++