Quan Xiao | Curriculum Vitae

110 8th Street, Troy, NY, 12180

Homepage | xiaoq5@rpi.edu | quanx1808@gmail.com
Linkedin | Google Scholar | JenniferquanXiao

EDUCATION

Rensselaer Polytechnic Institute — **Troy, NY**
Ph.D. of Computer and Systems Engineering (GPA: 4.0/4.0) — *Sep. 2021 - (Expected) May 2026*
Advisor: Tianyi Chen

University of Science and Technology of China — **Hefei, Anhui, China**
Bachelor of Statistics — *Sep. 2016 - July. 2020*

RESEARCH INTEREST

My research explores optimization and machine learning, with a focus on **bilevel optimization** for developing provably efficient algorithms that advance **generative AI models (e.g., LLMs, diffusion models) and next-generation AI computing (e.g., analog circuit devices)**. I aim to bridge theory and practice, designing efficient, scalable and hardware-aware optimization methods for a wide range of machine learning applications.

HONORS AND AWARDS

IEEE Signal Processing Society Scholarship	*2024*
Belsky Award for Computational Sciences and Engineering in RPI	*2023*
Neural Information Processing Systems Scholar Award	*2023*
International Conference on Machine Learning Participation Grant	*2022*
Successful Participant in Mathematical Contest In Modeling	*2019*
Excellent student scholarship of the University of Science and Technology of China	*2017 - 2019*

SELECTED PUBLICATIONS (* equal contribution)

[9] **Quan Xiao**, and Tianyi Chen. Unlocking Global Optimality in Bilevel Optimization: A Pilot Study. *Proc. of International Conference on Learning Representations*, 2025

[8] Han Shen, **Quan Xiao**, and Tianyi Chen. On Penalty-based Bilevel Gradient Descent Method. *Mathematical Programming (Series A)*, 2024.

[7] Liuyuan Jiang, **Quan Xiao**, Victor M Tenorio, Fernando Real-Rojas, Antonio Marques, Tianyi Chen. A Primal-Dual-Assisted Penalty Approach to Bilevel Optimization with Coupled Constraints. *Proc. of Neural Information Processing Systems*, 2024

[6] Jiajia Yu, **Quan Xiao**, Tianyi Chen, and Rongjie Lai. A Bilevel Optimization Method for Inverse Mean-Field Games. *Inverse Problems*, 2024.

[5] **Quan Xiao**, Songtao Lu, and Tianyi Chen. A Generalized Alternating Method for Bilevel Learning under the Polyak-Łojasiewicz Condition. *Proc. of Neural Information Processing Systems*, 2023

[4] **Quan Xiao**, Qing Ling, and Tianyi Chen. Lazy Queries Can Reduce Variance in Zeroth-order Optimization. *IEEE Transactions on Signal Processing*, 2023, vol. 71, pp. 3695-3709.

[3] **Quan Xiao**, Han Shen, Wotao Yin and Tianyi Chen. Alternating Projected SGD for Equality-constrained Bilevel Optimization. *Proc. of International Conference on Artificial Intelligence and Statistics*, 2023.

[2] *Momin Abbas, ***Quan Xiao**, *Lisha Chen, Pin-Yu Chen and Tianyi Chen. Sharp-MAML: Sharpness-Aware Model-Agnostic Meta Learning. *Proc. of International Conference on Machine Learning*, 2022.

[1] Tianyi Chen, Yuejiao Sun, **Quan Xiao**, and Wotao Yin. A Single-Timescale method for stochastic bilevel optimization. *Proc. of International Conference on Artificial Intelligence and Statistics*, 2022. **(Oral, top 2%)**

SLECTED PREPRINTS

[2] **Quan Xiao**, Hui Yuan, A F M Saif, Gaowen Liu, Ramana Rao Kompella, Mengdi Wang, Tianyi Chen. A First-order Generative Bilevel Optimization Framework for Diffusion Models, 2025. Submitted to ICML 2025.

[1] Zhaoxian Wu, **Quan Xiao**, Tayfun Gokmen, Hsinyu Tsai, Kaoutar El Maghraoui, and Tianyi Chen. Pipeline Gradient-based Model Training on Analog In-memory Accelerators. arXiv preprint arxiv:2410.15155, 2024. Submitted to ICML 2025.

PROFESSIONAL EXPERIENCE

- **IBM Thomas J. Watson Research Center**, Yorktown, NY.
  Research intern with mentor: Dr. Debarun Bhattacharjya. May. 2024 - Aug. 2024
  Project: Uncertainty quantification and selective generation in large language model.
  - Quantified **uncertainty in LLM and developed a selective generation approach based on confidence estimation during the fine-tuning stage**. This method was successfully applied to IBM's product of LLM-based SQL code generation, optimizing performance by selectively generating outputs with higher confidence. Part of the results were submitted to UAI 2025.

- **IBM Thomas J. Watson Research Center**, Yorktown, NY.
  Research intern with mentor: Dr. Songtao Lu. May. 2022 - Aug. 2022
  Project: Efficient method for bilevel optimization beyond strong convexity.
  - Developed an efficient method with optimal iteration complexity for bilevel optimization problem with non-strongly-convex lower-level problem. The paper was published by NeurIPS 2023.

INVITED TALK

- **Global Convergence in Bilevel Optimization: A Pilot Study.**
  Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA. Nov. 2024
- **Efficient Methods for Bilevel Optimization with Polyak-Lojasiewicz Lower-level Problems.**
  Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA. Nov. 2023
- **Solving A Class of Bilevel Optimization Without Lower-level Convexity.**
  SIAM Conference on Optimization (OP23), Seattle, WA. June. 2023

SERVICES

- Reviewer of Proc. of AAAI ('25), AISTATS ('22 – '25), NeurIPS ('23 – '24), ICLR ('24 – '25), ICML ('24 – '25).
- Reviewer of journal of IEEE Transactions on Signal Processing, IEEE Transactions on Networking.
- Reviewer and Technical Program Committee member of Cross-Community Federated Learning Workshop at MLSys 2022.