

QIUSHUI XU

☎ (+1) 814-321-6058 ✉ xqsisme@gmail.com

EDUCATION

Penn State University

Aug 2021 – Expected 2026

Ph.D. in Industrial Engineering - Operations Research

State College, PA

- **Research interests:** convex optimization, nonlinear optimization, feature engineering, Large Language Models (LLMs), reinforcement learning (RL), recommendation algorithm & system

Lanzhou University

Aug 2016 – Jun 2020

B.S. in Mathematics, top 1 out of 183

Gansu, China

EXPERIENCE

Microsoft Research Asia, Machine Learning Group

May 2024 – Aug 2024

Research Scientist Intern (Advisor: Lei Song)

- Conducted in-depth analysis of **GPT-Pretrained Decision Transformer** by generating heatmaps of key, query, and value matrices, identifying a novel phenomenon termed '**Markov heads**'.
- Theoretically and empirically demonstrated that Markov heads exhibit a strong focus on the final input token under random embedding initializations, with this behavior persisting after extensive fine-tuning.
- Evaluated the importance of Markov and Non-Markov heads using **Mixture of Attention**, and introduced **GPT-DTMA** to enable adaptive learning across environments without retraining from scratch.
- Executed experiments on **MuJoCo Locomotion** and **Maze tasks**, achieving superior performance in both short- and long-term settings compared to baseline models.

Baidu Big Data Lab, Inc.

Feb 2021 – Apr 2021

Research Scientist Intern (Advisor: Haori Xiong)

- Contributed to a technical report on high-dimensional inverse covariance matrix computation methods.
- Employed **Python** and **Matplotlib** to conduct comprehensive experiments and perform detailed analysis of results, enabling data-driven insights and informed decision-making.
- Acquired proficiency in optimization techniques, including **linear programming** and **quadratic approximation**.

Penn State University

Aug. 2021 – Present

PhD Research Assistant (Advisor: Necdet Serhat Aybat)

- Designed and implemented high-efficiency algorithms to tackle large-scale deep learning challenges.
- Proposed innovative **tuning-free accelerated first-order algorithms** for deterministic and stochastic scenarios.
- Integrated adaptive learning rates to achieve outstanding theoretical and practical performance, while pioneering backtracking conditions and establishing convergence results for minimax problems.
- Executed experiments on distributed robust optimization for neural networks using **PyTorch**, achieving state-of-the-art performance surpassing adaptive algorithms like **Adam**.

Eular Intelligence Technology Co., Ltd

Sep 2019 – Dec 2019

Data Analyst Intern

- Utilized **Pandas** and **SQL** to collect, clean, and label data across multiple provinces.
- Resolved data missing issues and implemented data-driven strategies, boosting dataset accuracy to **95%**.
- Applied **deep neural networks** to analyze classified data, delivering valuable insights and feedback for analysis.
- Optimized the **knowledge graph platform**, supporting enhanced functionality for hundreds of public companies.

SELECTED PUBLICATIONS

Enhancing Cross-domain Pre-Trained Decision Transformers with Adaptive Attention

Wenhao Zhao, **Qiushui Xu**, Linjie Xu, Lei Song, Jinyu Wang, Chunlai Zhou, Jiang Bian

Submitted to *Proceedings of the AAAI Conference on Artificial Intelligence*

A Stochastic GDA Method With Backtracking For Solving Nonconvex Concave Minimax Problems

Qiushui Xu, Xuan Zhang, Necdet Serhat Aybat and Mert Gurbuzbalaban

Submitted to *Journal of Machine Learning Research*

TECHNICAL SKILLS

R, Python, C++, MATLAB, SQL, TensorFlow, PyTorch, Pandas, Excel