XIANG LI

website: https://lx10077.github.io
email: lx10077@upenn.edu

EDUCATION

Peking University, Beijing China *Sep. 2018 - Jul. 2023*
School of Mathematical Sciences
PhD.
Advisor: Prof. Zhihua Zhang

Peking University, Beijing China *Sep. 2014 - Jul. 2018*
School of Mathematical Sciences
B.S.
Main degree: Statistics / Dual degree: Economics

RESEARCH INTERESTS

My research interests span statistics, optimization, machine learning theory, and applied probability. My work focuses on the following areas:

- **Statistical Foundations for LLMs**: Explore the statistical and algorithmic principles underlying large language models (LLMs), with a focus on designing efficient watermarking techniques and understanding their theoretical limits.
- **Trustworthy Machine Learning**: Investigate the trade-offs among accuracy, privacy, and robustness, and develop scalable algorithms with provable guarantees for real-world applications.
- **Online Decision Making**: Design computationally efficient algorithms with rigorous uncertainty quantification for streaming data settings, including queueing systems, reinforcement learning, and multi-armed bandits.
- **Decentralized Optimization**: Develop communication-efficient optimization algorithms with provable convergence guarantees in large-scale decentralized settings, such as federated learning.

EXPERIENCE

- Aug. 2023 - Now, Post-doctoral researcher at University of Pennsylvania, working with Prof. Weijie Su and Prof. Qi Long.
- Feb. 2022 - Feb. 2023, Visiting Phd. student in Department of Statistical Science at University of Toronto, work with Prof. Qiang Sun.
- 28 Jun. - 10 Jul. 2020, The Machine Learning Summer Schools by the Max Planck Institute for Intelligent Systems, Tübingen, Germany (MLSS 2020, virtual).
- Jun. 2018 - Nov. 2018, Intern at Face ++ (Megvii), Algorithm Group, work with Shuchang Zhou.
- Mar. 2018 - May. 2018, Intern at ToSimple, Algorithm Research Group, work with Naiyan Wang.

PUBLICATIONS

* indicates equal contribution; ** indicates alphabetical order.

Journal papers

[J1] **A Statistical Framework of Watermarks for Large Language Models: Pivot, Detection Efficiency and Optimal Rules** [pdf] [arxiv]
Xiang Li, Feng Ruan, Huiyuan Wang, Qi Long, Weijie J. Su
Annals of Statistics, 53 (1): 322-351, Feb. 2025

[J2] **Debiasing Watermarks for Large Language Models via Maximal Coupling** [arxiv]
Yangxinyu Xie, **Xiang Li**, Tanwi Mallick, Weijie J. Su, Ruixun Zhang
Major revision, *Journal of the American Statistical Association*

[J3] **Stochastic Approximation MCMC, Online Inference, and Applications in Optimization of Queueing Systems** [arxiv]
Xiang Li*, Jiadong Liang*, Xinyun Chen, Zhihua Zhang
Reject and resubmit, *Operation Research*

[J4] **A Random Projection Approach to Personalized Federated Learning: Enhancing Communication Efficiency, Robustness, and Fairness** [pdf]
Yuze Han**, **Xiang Li****, Shiyun Lin**, Zhihua Zhang**
Journal of Machine Learning Research, 25 (380): 1-88, Dec. 2024

[J5] **Variance-aware Decision Making with Linear Function Approximation with Heavy-tailed Rewards** [pdf] [arxiv]
Xiang Li, Qiang Sun
Transactions on Machine Learning Research, Apr. 2024

[J6] **FedPower: Privacy-Preserving Distributed Eigenspace Estimation** [pdf] [arxiv]
Xiao Guo*, **Xiang Li***, Xiangyu Chang, Shusen Wang, Zhihua Zhang
Machine Learning, 113 (11): 8427-8458, Sep. 2024

Conference papers

[C1] **A Statistical Analysis of Polyak-Ruppert-Averaged Q-Learning** [pdf]
Xiang Li, Wenhao Yang, Jiadong Liang, Zhihua Zhang, Michael I. Jordan
International Conference on Artificial Intelligence and Statistics (AISTATS) 2023

[C2] **Statistical Analysis of Karcher Means for Random Restricted PSD Matrices** [pdf]
Hengchao Chen, **Xiang Li**, Qiang Sun
International Conference on Artificial Intelligence and Statistics (AISTATS) 2023

[C3] **Statistical Estimation and Online Inference via Local SGD** [pdf]
Xiang Li, Jiadong Liang, Xiangyu Chang, Zhihua Zhang
Conference on Learning Theory (COLT) 2022

[C4] **Asymptotic Behaviors of Projected Stochastic Approximation: A Jump Diffusion Perspective** [pdf]
Jiadong Liang, Yuze Han, **Xiang Li**, Zhihua Zhang
Neural Information Processing Systems (NeurIPS) 2022, Spotlights

[C5] **Personalized Federated Learning towards Communication Efficiency, Robustness and Fairness** [pdf]
Shiyun Lin*, Yuze Han*, **Xiang Li**, Zhihua Zhang
Neural Information Processing Systems (NeurIPS) 2022

[C6] **Communication-Efficient Distributed SVD via Local Power Iterations** [pdf]
Xiang Li, Shusen Wang, Kun Chen, Zhihua Zhang
International Conference on Machine Learning (ICML) 2021

[C7] **Finding the Near Optimal Policy via Reductive Regularization in MDPs** [pdf]
Wenhao Yang, **Xiang Li**, Guangzeng Xie, Zhihua Zhang
Workshop on Reinforcement Learning Theory, ICML 2021

[C8] **On the Convergence of FedAvg on Non-IID Data** [pdf]
Xiang Li*, Kaixuan Huang*, Wenhao Yang*, Shusen Wang, Zhihua Zhang
International Conference on Learning Representations (ICLR) 2020, Oral presentation

[C9] **Do Subsampled Newton Methods Work for High-Dimensional Data?** [pdf]
Xiang Li, Shusen Wang, Zhihua Zhang
AAAI Conference on Artificial Intelligence (AAAI) 2020

[C10] **A Regularized Approach to Sparse Optimal Policy in Reinforcement Learning** [pdf]
Wenhao Yang*, **Xiang Li***, Zhihua Zhang
Neural Information Processing Systems (NeurIPS) 2019

Preprints

[P1] **Robust Detection of Watermarks for Large Language Models Under Human Edits** [arxiv]
Xiang Li, Feng Ruan, Huiyuan Wang, Qi Long, Weijie J. Su
Submitted

[P2] **Optimal Estimation of Watermark Proportions in Hybrid AI-Human Texts**
Xiang Li, Garrett Wen, Weiqing He, Jiayuan Wu, Qi Long, Weijie J. Su
Submitted

[P3] **Decentralized Federated Learning with f-Differential Privacy**
Chendi Wang, **Xiang Li**, Buxin Su, Qi Long, Weijie J. Su
Submitted

[P4] **Corruption-Robust Variance-aware Algorithms for Generalized Linear Bandits under Heavy-tailed Rewards**
Qingyuan Yu, Euijin Baek, **Xiang Li**, Qiang Sun
Submitted

[P5] **Uncertainty Quantification of Data Shapley via Statistical Inference** [pdf]
Mengmeng Wu, Zhihong Liu, **Xiang Li**, Ruoxi Jia, Xiangyu Chang
Submitted

[P6] **Finite-Time Decoupled Convergence in Nonlinear Two-Time-Scale Stochastic Approximation** [arxiv]
Yuze Han**, **Xiang Li****, Zhihua Zhang**
Submitted

[P7] **Decoupled Functional Central Limit Theorems for Two-Time-Scale Stochastic Approximation** [arxiv]
Yuze Han, **Xiang Li**, Jiadong Liang, Zhihua Zhang
Submitted

[P8] **Online Statistical Inference for Nonlinear Stochastic Approximation with Markovian Data** [arxiv]
Xiang Li, Jiadong Liang, Zhihua Zhang
Submitted

[P9] **Privacy-Preserving Community Detection for Locally Distributed Multiple Networks** [arxiv]

Xiao Guo, **Xiang Li**, Xiangyu Chang, Shujie Ma
Submitted

[P10] **Asymptotic Behaviors and Phase Transitions in Projected Stochastic Approximation: A Jump Diffusion Approach** [arxiv]
Jiadong Liang, Yuze Han, **Xiang Li**, Zhihua Zhang
Preprint

[P11] **Communication-Efficient Local Decentralized SGD Methods** [arxiv]
Xiang Li, Wenhao Yang, Shusen Wang, Zhihua Zhang
Preprint

PRESENTATIONS

- *Optimal Robust Detection for Gumbel-Max Watermarks Under Contamination*
  - Jul. 2024, IMS-China International Conference on Statistics and Probability, Yinchuan.
  - Sep. 2024, Penn conference on Big Data in Biomedical & Population Health Sciences.
  - Sep. 2024, Warren & ASSET Center Research Mixer, UPenn.
- *A Statistical Framework of Watermarks for Large Language Models: Pivot, Detection Efficiency and Optimal Rules*
  - May 2024, the NUS IMS workshop on statistical machine learning for high dimensional data, Singapore.
  - Jul. 2024, the 2nd joint conference on statistics and data science in China (JCSDS), Kunming.
  - Nov. 2024, the conference on statistical learning and data science (SLDS), Newport Beach, California.
- *Complete Asymptotic Analysis for Projected Stochastic Approximation and Debiased Variants*
  - Sep. 2023, the Allerton conference at the University of Illinois, Allerton Park & Retreat Center.
- *Polyak-Ruppert-Averaged Q-Learning is Statistically Efficient*
  - Jul. 2022, the RL workshop in Shanghai University of Finance and Economics.
- *Statistical Estimation and Inference via Local SGD in Federated Learning*
  - Nov. 2021, the 14-th China-R conference.
- *A Regularized Approach to Sparse Optimal Policy in Reinforcement Learning*
  - Jun. 2019, the machine learning workshop at Peking University.

ACADEMIC SERVICE

- Conference review: NeurIPS, ICML, ICLR, UAI, AISTATS, IJCAI.
- Journal review: Journal of Machine Learning Research (JMLR), Transactions on Machine Learning Research (TMLR), Journal of the American Statistical Association (JASA), Annals of Applied Probability (AOAP), Operational Research (OR), IEEE Transactions on Automatic Control, IEEE Open Journal of Signal Processing, IEEE Journal on Selected Areas in Communications, Transactions on Parallel and Distributed Systems, World Wide Web (WWW).
- Event organization:

– Aug. 2024, Session Chair for "Theoretical and Algorithmic Developments in Federated Learning" at the Modeling and Optimization: Theory and Applications (MOPTA) conference.

TEACHING EXPERIENCES

- *Reinforcement Learning: Theory and Algorithms*, Fall 2019, PKU, Teaching Assistant
- *Linear Algebra*, Spring 2021, PKU, Teaching Assistant

SELECTED AWARDS

- Outstanding PhD Graduates, Peking University *2023*
- President Scholarship for Excellent Ph.D Student, Peking University *2018, 2020-2021*
- National Scholarship, China *2017, 2019*
- Travel Award *NeurIPS 2019, AAAI 2020*
- Outstanding Graduates, School of Mathematical Sciences, Peking University *2018*
- First prize in National Undergraduate Mathematical Modeling Contest, China *2016*