Lei, Qi

Website: https://cecilialeiqi.github.io/
Google Scholar: https://scholar.google.com/citations?user=kGOgaowAAAAJ&hl=en
Email: qilei@princeton.edu

Research Interests

My research aims to bridge the theoretical and empirical boundary of modern machine learning algorithms and in particular AI safety, with a focus on data privacy, distributionally robust algorithms, sample- and parameter-efficient learning.

Professional Experience

New York University, NY, United States September 2022 -
- *Assistant Professor in Mathematics and Data Science*, Courant Institute of Mathematical Sciences and the Center for Data Science

Princeton University, NJ, United States July 2020 - August 2022
- *Associate Research Scholar (CIFellow)*, Electrical and Computer Engineering Department September 2021 - Present
- *Postdoc Research Associate (CIFellow)*, Electrical and Computer Engineering Department July 2020 - September 2021
- Mentor: Jason D. Lee

Institute for Advanced Study, Princeton, NJ, United States
September 2019 - July 2020
- Visiting Graduate Student for the "Special Year on Optimization, Statistics, and Theoretical Machine Learning"

Simons Institute, Berkeley, CA, United States May 2019 - August 2019
- Research Fellow for the Foundations of Deep Learning Program

Education

University of Texas at Austin, TX, United States August 2014 - May 2020
- *Ph.D., Oden Institute for Computational Sciences and Engineering*
- Advisors: Alexandros G. Dimakis and Inderjit S. Dhillon

Zhejiang University, Zhejiang, China Sep 2010 - June 2014
- *B.S., School of Mathematics (with honors)* (GPA 3.92/4.0, rank 1^{st})

Grants and Award (Faculty)

- NYU Research Catalyst Prize, 'Towards Data-Driven Resilient Supply Chains: A Statistical Foundation of Learning Robust Inventory Control Policies'" Co-PI, 15k NYU,2023
- Whitehead Fellowship for Junior Faculty in Biomedical and Biological Sciences,"Provably Privacy-preserving Learning Pipelines for Sensitive Health Data", PI, 30k NYU,2023
- Award #DE-SC0024721, "Learning reduced models under extreme data conditions for design and rapid decision-making in complex systems", Co-PI, 3.97M U.S. Department of Energy, Office of Science Energy Earthshot Initiative, 2023

Selected Grants and Awards (pre-faculty)

- Computing Innovation Fellowship CRA, 2020-2022
- Simons-Berkeley Research Fellowship Simons Institute, 2019
- The National Initiative for Modeling and Simulation Research Fellowship UT Austin, 2014-2018

- Outstanding Dissertation Award — Oden Institute, 2021
- Rising Star for Machine Learning — University of Maryland, 2021
- Rising Star for EECS — UIUC, 2019 & MIT, 2021
- Rising Star for Computational and Data Science — UT Austin, 2020

Thesis

Qi Lei, "Provably effective algorithms for min-max optimization" May 2020
Received the 2021 Outstanding Dissertation Award, Oden Institute

Publications
(∗ indicates α-β order)

1. Sheng Liu, Zihan Wang, **Qi Lei**, "Data Reconstruction Attacks and Defenses: A Systematic Evaluation", *to appear at AISTATS 2025*
2. Tao Wen, Zihan Wang, Quan Zhang, **Qi Lei**, "Elastic Representation: Mitigating Spurious Correlations for Group Robustness", *to appear at AISTATS 2025*
3. Ziliang Samuel Zhong, Xiang Pan, **Qi Lei**, "Bridging Domains with Approximately Shared Features", *to appear at AISTATS 2025*
4. Jianwei Li, Yijun Dong, **Qi Lei**, "Greedy Output Approximation: Towards Efficient Structured Pruning for LLMs Without Retraining", *to appear at CPAL 2025*
5. Tao Wen, Elynn Chen, Yuzhou Chen, **Qi Lei**, "Bridging Domain Adaptation and Graph Neural Networks: A Tensor-Based Framework for Effective Label Propagation", *to appear at CPAL 2025*
6. Qi Zhang, Yifei Wang, Jingyi Cui, Xiang Pan, **Qi Lei**, Stefanie Jegelka, Yisen Wang, "Beyond Interpretability: The Gains of Feature Monosemanticity on Model Robustness", *to appear at ICLR 2025*
7. Yijun Dong, Hoang Phan, Xiang Pan, **Qi Lei**, "Sketchy Moment Matching: Toward Fast and Provable Data Selection for Finetuning", *NeurIPS 2024*
8. Qian Yu, Yining Wang, Baihe Huang, **Qi Lei**, Jason D Lee, "Stochastic Zeroth-Order Optimization under Strongly Convexity and Lipschitz Hessian: Minimax Sample Complexity", *NeurIPS 2024*
9. Hoang Phan, Andrew G Wilson, **Qi Lei**, "Controllable Prompt Tuning For Balancing Group Distributional Robustness", *International Conference of Machine Learning (ICML), 2024*
10. Hong J Jeon, Jason D Lee, **Qi Lei**, Benjamin Van Roy, "An Information-Theoretic Analysis of In-Context Learning", *International Conference of Machine Learning (ICML), 2024*
11. Qian Yu, Yining Wang, Baihe Huang, **Qi Lei**, Jason D Lee, "Sample Complexity for Quadratic Bandits: Hessian Dependent Bounds and Optimal Algorithms", *Advances in Neural Information Processing Systems 36, 2023*
12. Yijun Dong, Kevin Miller, **Qi Lei**, Rachel Ward, "Cluster-aware Semi-supervised Learning: Relational Knowledge Distillation Provably Learns Clustering", *Advances in Neural Information Processing Systems 36, 2023*
13. Jianwei Li, **Qi Lei**, Wei Cheng, Dongkuan Xu, "Towards Robust Pruning: An Adaptive Knowledge-Retention Pruning Strategy for Language Models", *EMNLP conference, 2023: 1229-1247*
14. Jianwei Li, Weizhi Gao, **Qi Lei**, Dongkuan Xu, "Breaking through Deterministic Barriers: Randomized Pruning Mask Generation and Selection", *EMNLP findings, 2023: 11407-11423*
15. Tianci Liu, Tong Yang, Quan Zhang, **Qi Lei**, "Optimization for Amortized Inverse Problems", *International Conference of Machine Learning (ICML), 2023*

16. Zihan Wang, Jason Lee, **Qi Lei**, "Reconstructing Training Data from Model Gradient, Provably", *International Conference on Artificial Intelligence and Statistics (AISTATS), 2023: 6595-6612*
17. Shuo Yang, Yijun Dong, Rachel Ward, Inderjit Dhillon, Sujay Sanghavi, **Qi Lei**, "Sample Efficiency of Data Augmentation Consistency Regularization", *International Conference on Artificial Intelligence and Statistics (AISTATS), 2023: 3825-3853*
18. Kurtland Chua, **Qi Lei**, Jason Lee, "Provable Hierarchy-Based Meta-Reinforcement Learning", *International Conference on Artificial Intelligence and Statistics (AISTATS), 2023: 10918-10967*
19. Qian Yu, Yining Wang, Baihe Huang, **Qi Lei**, Jason Lee, "Optimal Sample Complexity Bounds for Non-convex Optimization under Kurdyka-Lojasiewicz Condition", *International Conference on Artificial Intelligence and Statistics (AISTATS), 2023: 6806-6821*
20. Minhao Cheng, **Qi Lei**, Pin-Yu Chen, Inderjit Dhillon, Cho-Jui Hsieh, "Cat: Customized adversarial training for improved robustness", *International Joint Conference on Artificial Intelligence (IJCAI), 2022: 673-679*
21. Jason D. Lee*, **Qi Lei***, Nikunj Saunshi*, Jiacheng Zhuo*, "Predicting What You Already Know Helps: Provable Self-Supervised Learning", *Neural Information Processing Systems (NeurIPS), 2021: 309-323*
22. Baihe Huang*, Kaixuan Huang*, Sham M. Kakade*, Jason D. Lee*, **Qi Lei***, Runzhe Wang*, Jiaqi Yang*, "Optimal Gradient-based Algorithms for Non-concave Bandit Optimization", *Neural Information Processing Systems (NeurIPS), 2021: 29101-29115*
23. Kurtland Chua, **Qi Lei**, Jason D. Lee, "How Fine-Tuning Allows for Effective Meta-Learning", *Neural Information Processing Systems (NeurIPS), 2021: 8871-8884*
24. Baihe Huang*, Kaixuan Huang*, Sham M. Kakade*, Jason D. Lee*, **Qi Lei***, Runzhe Wang*, Jiaqi Yang*, "Going Beyond Linear RL: Sample Efficient Neural Function Approximation", *Neural Information Processing Systems (NeurIPS), 2021: 8968-8983*
25. **Qi Lei**, Wei Hu, Jason D. Lee. "Near-Optimal Linear Regression under Distribution Shift", *International Conference of Machine Learning (ICML), 2021: 6164-6174*
26. Tianle Cai*, Ruiqi Gao*, Jason D Lee*, **Qi Lei***. "A Theory of Label Propagation for Subpopulation Shift", *International Conference of Machine Learning (ICML), 2021: 1170-1182*
27. Jay Whang, **Qi Lei**, Alexandros G. Dimakis, "Solving Inverse Problems with a Flow-based Noise Model", *International Conference of Machine Learning (ICML), 2021: 11146-11157*
28. Simon S. Du*, Wei Hu*, Sham M. Kakade*, Jason D. Lee*, **Qi Lei***, "Few-Shot Learning via Learning the Representation, Provably", *International Conference on Learning Representations (ICLR), 2021*
29. **Qi Lei***, Sai Ganesh Nagarajan*, Ioannis Panageas*, Xiao Wang*, "Last iterate convergence in no-regret learning: constrained min-max optimization for convex-concave landscapes", *International Conference on Artificial Intelligence and Statistics (AISTATS), 2021: 1441-1449*

30. Xiao Wang, **Qi Lei**, Ioannis Panageas, "Fast Convergence of Langevin Dynamics on Manifold: Geodesics meet Log-Sobolev", *Neural Information Processing Systems (NeurIPS), 2020*

31. Jiacheng Zhuo, **Qi Lei**, Alexandros G. Dimakis, Constantine Caramanis. "Communication-Efficient Asynchronous Stochastic Frank-Wolfe over Nuclear-norm Ball", *International Conference on Artificial Intelligence and Statistics (AISTATS), 2020: 1464-1474*

32. **Qi Lei**, Jason Lee, Alexandros G. Dimakis, Contantinos Daskalakis. "SGD Learns One-Layer Networks in WGANs", *International Conference of Machine Learning (ICML), 2020: 5799-5808*

33. **Qi Lei**, Jiacheng Zhuo, Constantine Caramanis, Inderjit S. Dhillon, Alexandros G. Dimakis, "Primal-Dual Block Generalized Frank-Wolfe", *Neural Information Processing Systems (NeurIPS), 2019: 13866-13875*

34. **Qi Lei**, Ajil Jalal, Inderjit S. Dhillon, Alexandros G. Dimakis, "Inverting Deep Generative models, One layer at a time", *Neural Information Processing Systems (NeurIPS), 2019: 13910-13919*

35. **Qi Lei**, Jinfeng Yi, Roman Vaculin, Lingfei Wu, Inderjit S. Dhillon, "Similarity Preserving Representation Learning for Time Series Analysis", *International Joint Conference on Artificial Intelligence (IJCAI), 2019: 2845-2851*

36. **Qi Lei**, Lingfei Wu, Pin-Yu Chen, Alexandros G. Dimakis, Inderjit S. Dhillon, Michael Witbrock, "Discrete Adversarial Attacks and Submodular Optimization with Applications to Text Classification", *Systems and Machine Learning (MLSys), 2019* (**covered by Nature News)**

37. Jinfeng Yi, **Qi Lei**, Wesley M Gifford, Ji Liu, Junchi Yan, Bowen Zhou, "Fast Unsupervised Location Category Inference from Highly Inaccurate Mobility Data", *SIAM International Conference on Data Mining 2019: 55-63*

38. Zhewei Yao, Amir Gholami, **Qi Lei**, Kurt Keutzer, Michael W. Mahoney, "Hessian-based Analysis of Large Batch Training and Robustness to Adversaries", *Neural Information Processing Systems (NIPS), 2018: 4954-4964*

39. Jiong Zhang, **Qi Lei**, Inderjit S. Dhillon. "Stabilizing Gradients for Deep Neural Networks via Efficient SVD Parameterization", *International Conference of Machine Learning (ICML), 2018: 5801-5809*

40. Jinfeng Yi, **Qi Lei**, Junchi Yan, Wei Sun, "Session expert: A lightweight conference session recommender system", *IEEE International Conference on Big Data (Big Data), 2018: 1677-1682*

41. Lingfei Wu, Ian En-Hsu Yen, Jinfeng Yi, Fangli Xu, **Qi Lei**, Michael Witbrock. "Random Warping Series: A Random Features Method for Time-Series Embedding", *International Conference on Artificial Intelligence and Statistics (AISTATS), 2018: 793-802*

42. Hsiang-fu Yu, Cho-Jui Hsieh, **Qi Lei**, Inderjit S. Dhillon. "A Greedy Approach for Budgeted Maximum Inner Product Search", *Neural Information Processing Systems (NIPS), 2017: 5453-5462*

43. **Qi Lei**, Enxu Yen, Chao-yuan Wu, Inderjit S. Dhillon, Pradeep Ravikumar, "Doubly Greedy Primal-Dual Coordinate Methods on Sparse Empirical Risk Minimization", *International Conference of Machine Learning (ICML), 2017: 2034-2042*

44. Rashish Tandon, **Qi Lei**, Alexandros G. Dimakis, Nikos Karampatziakis, "Gradient Coding: Avoiding Stragglers in Distributed Learning", *International Conference of Machine Learning (ICML), 2017: 3368-3376*

45. **Qi Lei**, Kai Zhong, Inderjit S. Dhillon, "Coordinate-wise Power Method", *Neural Information Processing System (NIPS), 2016: 2056-2064*
46. Arnaud Vandaele, Nicolas Gillis, **Qi Lei**, Kai Zhong, Inderjit S. Dhillon, "Coordinate Descent Methods for Symmetric Nonnegative Matrix Factorization", *IEEE Transactions on Signal Processing*, 64.21 (2016): 5571-5584
47. Maria R. D'Orsogna, **Qi Lei**, Tom Chou, "First assembly times and equilibration in stochastic coagulation-fragmentation", *Journal of Chemical Physics*, 2015: 143.1, 014112
48. Jiazhou Chen, **Qi Lei**, Yongwei Miao, Qunsheng Peng, "Vectorization of Line Drawing Image based on Junction Analysis", *Science China Information Sciences*, 2014:1-14
49. Jiazhou Chen, **Qi Lei**, Fan Zhong, Qunsheng Peng, "Interactive Tensor Field Design Based on Line Singularities", *Proceedings of the 13th International CAD /Graphics*, 2013

Workshop Articles

1. Yijun Dong, Xiang Pan, Hoang Phan, **Qi Lei**, "Randomly Pivoted V-optimal Design: Fast Data Selection under Low Intrinsic Dimension", *ML and Compression-NeurIPS 2024*
2. Jianwei Li, Sheng Liu, **Qi Lei**, "'Beyond Gradient and Priors in Privacy Attacks: Leveraging Pooler Layer Inputs of Language Models in Federated Learning", *FL@FM-NeurIPS 2023*
3. Chun-Yin Huang, **Qi Lei**, Xiaoxiao Li, "Efficient Medical Image Assessment via Self-supervised Learning", *DALI@MICCAI Workshop, 2022: 102-111* (**with Best Paper Honorable Mention**)
4. Kaixuan Huang*, Sham M. Kakade*, Jason D. Lee*, **Qi Lei***, "A Short Note on the Relationship of Information Gain and Eluder Dimension", *ICML 2021 Workshop on Reinforcement Learning Theory*

Patents

- "Method and System for General and Efficient Time Series Representation Learning via Dynamic Time Warping."
  Q. Lei, J. Yi, R. Vaculin, and W. Sun
- "Real-Time Cold Start Recommendation and Rationale within a Dialog System".
  Q. Lei, J. Yi, R. Vaculin, M. Pietro

Teaching

Center for Data Science, New York University
- Linear Algebra and Optimization: *Instructor* Fall 2024
- Modern Topics in Statistical Learning Theory: *Instructor* Spring 2023,2024

Courant Institute Mathematics Department, New York University
- Mathematical Statistics: *Instructor* Fall 2023, 2024
- Probability and Statistics: *Instructor* Fall 2022

Department of Electrical and Computer Engineering, Princeton
- Theory of Deep Learning: Representation and Weakly Supervised Learning: *Teaching Assistant* Fall 2020

Department of Electrical and Computer Engineering, UT Austin
- Scalable Machine Learning: *Teaching Assistant* Fall 2019

Oden Institute for Computational Engineering and Sciences, UT Austin

- Mathematical Methods in Applied Engineering and Sciences: *Instructor Intern* Fall 2015

Industry Experiences

Facebook/Photo&Video Search June 2018 - September 2018

- Explored offline/online evaluation gaps by estimating expected number of clicks based on historical logging data.

Amazon/A9 Product Search May 2017 - August 2017

- Inline search suggestions: used deep learning methods for NLP user search tasks.

Amazon Web Services (AWS Deep Learning Team) January 2017 - April 2017

- Documentations for MXNet: a deep learning framework designed for both efficiency and flexibility.

IBM Thomas J. Watson Research Center May 2016 - October 2016

- Clients' propensity prediction of trading options Partnered with one of the largest American financial companies on a challenge problem of predicting its clients' propensity of trading options
- Create World of Watson Session recommendation system: `https://myibm.ibm.com/events/wow/watson/`

Service

Co-organizing the Math and Data seminar, New York University, since 2022

Co-organizing mini-symposium on "Efficient computation and learning with randomized sampling and pruning", at SIAM MDS 2024

Co-organizing the workshop on meta-learning at NeurIPS 2022

Student mentor, Oden Institute, 2018-2020

Conference Reviewer: MLSys, COLT, STOC, NeurIPS, ICML, ICLR, AISTATS, AAAI, and more

Journal Reviewer: MOR, JMLR, Annuls of Statistics, EJMS, JSAIT, TNNLS, TKDE, ISIT, IT, Computer Speech & Language, and more