Jingfeng Wu

Website: uuujf.github.io
Email: uuujf@berkeley.edu

Academic Experience

University of California, Berkeley — Berkeley, US
Postdoctoral Fellow — 2023 - Present
Advisors: Peter Bartlett and Bin Yu

Education

Johns Hopkins University — Baltimore, US
Ph.D. in *Computer Science* — 2019 - 2023
Advisor: Vladimir Braverman

Peking University — Beijing, CN
M.S. in *Applied Mathematics* — 2016 - 2019
B.S. in *Mathematics & Applied Mathematics* — 2012 - 2016

Industrial Experience

Google Research — Seattle, US
Research Intern — Summer 2022
Mentors: Wennan Zhu and Peter Kairuoz

Baidu Research — Beijing, CN
Research Intern — Winter 2018
Mentor: Haoyi Xiong

Research Interests

Deep Learning Theory, Algorithms, Machine Learning, Optimization, Statistical Learning Theory

Conference Papers (asterisk* indicates equal contributions or alphabetical order)

[1] **J. Wu**, P. L. Bartlett*, M. Telgarsky*, and B. Yu*. "Benefits of Early Stopping in Gradient Descent for Overparameterized Logistic Regression". In: *International Conference on Machine Learning (ICML)*. 2025.

[2] R. Zhang, **J. Wu**, L. Lin, and P. L. Bartlett. "Minimax Optimal Convergence of Gradient Descent in Logistic Regression via Large and Adaptive Stepsizes". In: *ICML*. 2025.

[3] Y. Cai*, K. Zhou*, **J. Wu**, S. Mei, M. Lindsey, and P. L. Bartlett. "Implicit Bias of Gradient Descent for Non-Homogeneous Deep Networks". In: *ICML*. 2025.

[4] H. Zhang, D. Morwani, N. Vyas, **J. Wu**, D. Zou, U. Ghai, D. P. Foster, and S. M. Kakade. "How Does Critical Batch Size Scale in Pre-training?" In: *International Conference on Learning Representations (ICLR)*. 2025.

[5] Y. Cai, **J. Wu**, S. Mei, M. Lindsey, and P. L. Bartlett. "Large Stepsize Gradient Descent for Non-Homogeneous Two-Layer Networks: Margin Improvement and Fast Optimization". In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2024.

[6] L. Lin, **J. Wu**, S. M. Kakade, P. L. Bartlett, and J. D. Lee. "Scaling Laws in Linear Regression: Compute, Parameters, and Data". In: *NeurIPS*. 2024.

[7] R. Zhang, **J. Wu**, and P. L. Bartlett. "In-Context Learning of a Linear Transformer Block: Benefits of the MLP Component and One-Step GD Initialization". In: *NeurIPS*. 2024.

[8] **J. Wu**, P. L. Bartlett[*], M. Telgarsky[*], and B. Yu[*]. "Large Stepsize Gradient Descent for Logistic Loss: Non-Monotonicity of the Loss Improves Optimization Efficiency". In: *Conference on Learning Theory (COLT)*. 2024.

[9] **J. Wu**, D. Zou, Z. Chen, V. Braverman, Q. Gu, and P. L. Bartlett. "How Many Pretraining Tasks Are Needed for In-Context Learning of Linear Regression?" In: *ICLR*. 2024.

[10] X. Li, Y. Deng, **J. Wu**, D. Zhou, and Q. Gu. "Risk Bounds of Accelerated SGD for Overparameterized Linear Regression". In: *ICLR*. 2024.

[11] **J. Wu**, V. Braverman, and J. D. Lee. "Implicit Bias of Gradient Descent for Logistic Regression at the Edge of Stability". In: *NeurIPS*. 2023.

[12] **J. Wu**, W. Zhu, P. Kairouz, and V. Braverman. "Private Federated Frequency Estimation: Adapting to the Hardness of the Instance". In: *NeurIPS*. 2023.

[13] H. Li[*], **J. Wu**[*], and V. Braverman. "Fixed Design Analysis of Regularization-Based Continual Learning". In: *Conference on Lifelong Learning Agents (CoLLAs)*. 2023.

[14] **J. Wu**[*], D. Zou[*], Z. Chen[*], V. Braverman, Q. Gu, and S. M. Kakade. "Finite-Sample Analysis of Learning High-Dimensional Single ReLU Neuron". In: *ICML*. 2023.

[15] **J. Wu**[*], D. Zou[*], V. Braverman, Q. Gu, and S. M. Kakade. "The Power and Limitation of Pretraining-Finetuning for Linear Regression under Covariate Shift". In: *NeurIPS*. 2022.

[16] D. Zou[*], **J. Wu**[*], V. Braverman, Q. Gu, and S. M. Kakade. "Risk Bounds of Multi-Pass SGD for Least Squares in the Interpolation Regime". In: *NeurIPS*. 2022.

[17] **J. Wu**[*], D. Zou[*], V. Braverman, Q. Gu, and S. M. Kakade. "Last Iterate Risk Bounds of SGD with Decaying Stepsize for Overparameterized Linear Regression". In: *ICML*. 2022.

[18] **J. Wu**, V. Braverman, and L. F. Yang. "Gap-dependent Unsupervised Exploration for Reinforcement Learning". In: *International Conference on Artificial Intelligence and Statistics (AISTATS)*. 2022.

[19] D. Zou[*], **J. Wu**[*], V. Braverman, Q. Gu, D. P. Foster, and S. M. Kakade. "The Benefits of Implicit Regularization from SGD in Least Squares Problems". In: *NeurIPS*. 2021.

[20] **J. Wu**, V. Braverman, and L. F. Yang. "Accommodating Picky Customers: Regret Bound and Exploration Complexity for Multi-Objective Reinforcement Learning". In: *NeurIPS*. 2021.

[21] H. Li, A. Krishnan, **J. Wu**, S. Kolouri, P. K. Pilly, and V. Braverman. "Lifelong Learning with Sketched Structural Regularization". In: *Asian Conference on Machine Learning (ACML)*. 2021.

[22] Z. Yu, C. Hu, **J. Wu**, X. Sun, V. Braverman, M. Chowdhury, Z. Liu, and X. Jin. "Programmable Packet Scheduling with a Single Queue". In: *ACM Special Interest Group on Data Communication (SIGCOMM)*. 2021.

[23] D. Zou[*], **J. Wu**[*], V. Braverman, Q. Gu, and S. M. Kakade. "Benign Overfitting of Constant-Stepsize SGD for Linear Regression". In: *COLT*. 2021.

[24] **J. Wu**, D. Zou, V. Braverman, and Q. Gu. "Direction Matters: On the Implicit Bias of Stochastic Gradient Descent with Moderate Learning Rate". In: *ICLR*. 2021.

[25] J. You, **J. Wu**, X. Jin, and M. Chowdhury. "Ship Compute or Ship Data? Why not Both?" In: *USENIX Symposium on Networked Systems Design and Implementation (NSDI)*. 2021.

[26] Z. Yu, **J. Wu**, V. Braverman, I. Stoica, and X. Jin. "Twenty Years After: Hierarchical Core-Stateless Fair Queueing." In: *NSDI*. 2021.

[27] **J. Wu**, V. Braverman, and L. F. Yang. "Obtaining Adjustable Regularization for Free via Iterate Averaging". In: *ICML*. 2020.

[28] **J. Wu**, W. Hu, H. Xiong, J. Huan, V. Braverman, and Z. Zhu. "On the Noisy Gradient Descent that Generalizes as SGD". In: *ICML*. 2020.

[29] B. Yu[*], **J. Wu**[*], J. Ma, and Z. Zhu. "Tangent-Normal Adversarial Regularization for Semi-Supervised Learning". In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019.

[30] Z. Zhu[*], **J. Wu**[*], B. Yu, L. Wu, and J. Ma. "The Anisotropic Noise in Stochastic Gradient Descent: Its Behavior of Escaping from Sharp Minima and Regularization Effects". In: *ICML*. 2019.

Journal Papers (asterisk* indicates equal contributions or alphabetical order)

[31] A. Soltoggio, E. Ben-Iwhiwhu, V. Braverman, ..., **J. Wu**, et al. "A collective AI via lifelong learning and sharing at the edge". In: *Nature Machine Intelligence* (2024).

[32] D. Zou*, **J. Wu***, V. Braverman, Q. Gu, and S. M. Kakade. "Benign Overfitting of Constant-Stepsize SGD for Linear Regression". In: *Journal of Machine Learning Research (JMLR)* (2023).

Invited Talks

"Reimagining Gradient Descent: Large Stepsize, Oscillation, and Acceleration"
- **MPI & UCLA**, *Math Machine Learning Seminar*, hosted by Guido Montufar — June 2025
- **SIAM DS25**, *Dynamical Systems for Machine Learning*, hosted by Molei Tao — May 2025
- **UCLA**, *Level Set Meeting*, hosted by Shu Liu and Stanley Osher — January 2025
- **Simons Foundation**, *MoDL Annual Meeting*, hosted by Peter Bartlett and Rene Vidal — September 2024
- **UC San Diego**, *MoDL Collaboration Meeting*, hosted by Chaoyue Liu et al. — May 2024
- **UCLA**, *Computer Science Seminar*, hosted by Quanquan Gu — March 2024
- **UC Berkeley**, *Biostatistics Seminar*, hosted by Lexin Li — February 2024
- **UC San Diego**, *Group Seminar*, hosted by Mikhail Belkin — February 2024

"A Statistical Viewpoint on Implicit Regularization: GD for Logistic Regression"
- **ICTP**, *6th Youth in High-Dimensions Conference*, hosted by Marco Mondelli et al. — July 2025
- **UC Berkeley**, *Deep Learning Theory Workshop*, hosted by Peter Bartlett et al. — February 2025

"New Insights about SGD: Stepsize, Risk Convergence, and Implicit Regularization"
- **UC Davis**, *Statistics Seminar*, hosted by Xiao Hui Tai — Octorber 2023

"Implicit Bias of Gradient Descent for Logistic Regression at the Edge of Stability"
- **TTI-Chicago**, *MoDL Collaboration Meeting*, hosted by Sam Buchanan et al. — May 2023

"The Implicit Regularization of SGD in Least Squares and Beyond"
- **Rice University**, *Algorithms and ML Seminar*, hosted by Anastasios Kyrillidis — January 2023
- **Princeton University**, *Group Seminar*, hosted by Jason Lee — December 2022
- **Georgia Tech**, *Group Seminar*, hosted by Molei Tao — November 2022
- **Google Research**, *Learning Theory Seminar*, hosted by Mehryar Mohri — August 2022
- **MPI & UCLA**, *Math Machine Learning Seminar*, hosted by Guido Montufar — June 2022

Professional Services

Oragnizer
- Deep Learning Theory Workshop at Simons Institute at UC Berkeley — February 2025

Conference Reviewer
- International Conference on Machine Learning (ICML) — 2020 - 2025
- Conference on Neural Information Processing Systems (NeurIPS) — 2020 - 2025
- International Conference on Learning Representations (ICLR) — 2021 - 2025
- International Conference on Artificial Intelligence and Statistics (AISTATS) — 2021 - 2023
- Conference on Uncertainty in Artificial Intelligence (UAI) — 2023

Conference Program Committee Member
  – AAAI Conference on Artificial Intelligence (AAAI)   2021 - 2023

Journal Reviewer
  – Journal of Machine Learning Research (JMLR)
  – IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)
  – Transactions on Machine Learning Research (TMLR)
  – SIAM Journal on Mathematics of Data Science (SIMODS)
  – IEEE Transactions on Information Theory
  – Applied Probability Journals
  – Journal of Artificial Intelligence Research (JAIR)

TEACHING

- **Teaching Assistant** for "Machine Learning: Deep Learning", Johns Hopkins University   Spring 2023

HONORS

- **Rising Stars in Data Science**, University of Chicago and UC San Diego   2023
- **MINDS Data Science Fellowship**, Johns Hopkins University   Summer 2021
- **Best Reviewers (Top 10%)**, ICML   2021