Jingfeng Wu

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

Academic Experience

University of California, Berkeley — Berkeley, CA, USA
Postdoctoral Fellow, *Simons Institute for the Theory of Computing* — 2023 - Present
Advisors: Peter Bartlett and Bin Yu

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

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

Industrial Experience

Google Research — Seattle, WA, USA
Research Intern — Summer 2022
Mentors: Wennan Zhu and Peter Kairouz

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

Research Interests

Deep Learning Theory, Optimization, Statistical Learning, Algorithms

Preprint (* indicates equal contribution or alphabetical order)

[1] **J. Wu**, P. L. Bartlett*, J. D. Lee*, S. M. Kakade*, and B. Yu*. *Risk Comparisons in Linear Regression: Implicit Regularization Dominates Explicit Regularization*. arXiv:2509.17251. 2025.

Conference Papers (* indicates equal contribution or alphabetical order)

[2] **J. Wu***, P. Marion*, and P. L. Bartlett. "Large Stepsizes Accelerate Gradient Descent for Regularized Logistic Regression". In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2025.

[3] L. Lin, **J. Wu**, and P. L. Bartlett. "Improved Scaling Laws in Linear Regression via Data Reuse". In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2025.

[4] **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.

[5] 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: *International Conference on Machine Learning (ICML)*. 2025.

[6] 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: *International Conference on Machine Learning (ICML)*. 2025.

[7] 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.

[8] 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.

[9] L. Lin, **J. Wu**, S. M. Kakade, P. L. Bartlett, and J. D. Lee. "Scaling Laws in Linear Regression: Compute, Parameters, and Data". In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2024.

[10] 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: *Advances in Neural Information Processing Systems (NeurIPS)*. 2024.

[11] **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.

[12] **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: *International Conference on Learning Representations (ICLR)*. 2024.

[13] X. Li, Y. Deng, **J. Wu**, D. Zhou, and Q. Gu. "Risk Bounds of Accelerated SGD for Overparameterized Linear Regression". In: *International Conference on Learning Representations (ICLR)*. 2024.

[14] **J. Wu**, V. Braverman, and J. D. Lee. "Implicit Bias of Gradient Descent for Logistic Regression at the Edge of Stability". In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2023.

[15] **J. Wu**, W. Zhu, P. Kairouz, and V. Braverman. "Private Federated Frequency Estimation: Adapting to the Hardness of the Instance". In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2023.

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

[17] **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: *International Conference on Machine Learning (ICML)*. 2023.

[18] **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: *Advances in Neural Information Processing Systems (NeurIPS)*. 2022.

[19] 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: *Advances in Neural Information Processing Systems (NeurIPS)*. 2022.

[20] **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: *International Conference on Machine Learning (ICML)*. 2022.

[21] **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.

[22] 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: *Advances in Neural Information Processing Systems (NeurIPS)*. 2021.

[23] **J. Wu**, V. Braverman, and L. F. Yang. "Accommodating Picky Customers: Regret Bound and Exploration Complexity for Multi-Objective Reinforcement Learning". In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2021.

[24] 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.

[25] 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.

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

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

[28] 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.

[29] Z. Yu, **J. Wu**, V. Braverman, I. Stoica, and X. Jin. "Twenty Years After: Hierarchical Core-Stateless Fair Queueing." In: *USENIX Symposium on Networked Systems Design and Implementation (NSDI)*. 2021.

[30] **J. Wu**, V. Braverman, and L. F. Yang. "Obtaining Adjustable Regularization for Free via Iterate Averaging". In: *International Conference on Machine Learning (ICML)*. 2020.

[31] **J. Wu**, W. Hu, H. Xiong, J. Huan, V. Braverman, and Z. Zhu. "On the Noisy Gradient Descent that Generalizes as SGD". In: *International Conference on Machine Learning (ICML)*. 2020.

[32] 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.

[33] 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: *International Conference on Machine Learning (ICML)*. 2019.

Journal Papers (* indicates equal contribution or alphabetical order)

[34] 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).

[35] 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).

Recent Invited Talks

"A Statistical View on Implicit Regularization: GD Dominates Ridge"

Columbia, *Machine Learning & AI Seminar* (host: Daniel Hsu) October 2025

NYU, *Math & Data Seminar* (host: Matus Telgarsky) October 2025

Yale, *Statistics & Data Science Seminar* (hosts: Theodor Misiakiewicz and Omar Montasser) September 2025

"Reimagining Gradient Descent: Large Stepsize, Oscillation, and Acceleration"

Harvard, *Talk at Kempner Institute* (host: Sham Kakade) October 2025

MIT, *Talk at Laboratory for Information and Decision Systems* (host: Pablo Parrilo) September 2025

JHU, *CS Theory Seminar* (host: Vladimir Braverman) September 2025

UPenn, *Group Meeting* (host: Jason Altschuler) September 2025

MPI & UCLA, *Math Machine Learning Seminar* (host: Guido Montufar) June 2025

SIAM DS25, *Dynamical Systems for Machine Learning* (host: Molei Tao) May 2025

UCLA, *Level Set Meeting* (hosts: Shu Liu and Stanley Osher) January 2025

Simons Foundation, *MoDL Annual Meeting* (hosts: Peter Bartlett and Rene Vidal) September 2024

UC San Diego, *MoDL Collaboration Meeting* (host: Mikhail Belkin et al.) May 2024

UCLA, *Computer Science Seminar* (host: Quanquan Gu) March 2024

UC Berkeley, *Biostatistics Seminar* (host: Lexin Li) February 2024

"A Statistical View on Implicit Regularization: GD for Logistic Regression"

ICTP, *6th Youth in High-Dimensions Conference* (hosts: Marco Mondelli et al.) July 2025

UC Berkeley, *Deep Learning Theory Workshop* (hosts: Peter Bartlett et al.) February 2025

Professional Services

Organizer

Deep Learning Theory Workshop, Simons Institute, 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 - 2026
ACM-SIAM Symposium on Discrete Algorithms (SODA), subreviewer 2026
International Conference on Artificial Intelligence and Statistics (AISTATS) 2021 - 2023
Conference on Uncertainty in Artificial Intelligence (UAI) 2023
AAAI Conference on Artificial Intelligence (AAAI), PC member reviewer 2021 - 2023

Journal Reviewer

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

Teaching

NeurIPS 2025 Tutorial, Leading Speaker December 2025
Tutorial: "Theoretical Insights on Training Instability in Deep Learning"

Johns Hopkins University, Teaching Assistant Spring 2023
Course: "Machine Learning: Deep Learning"

Honors

Rising Stars in Data Science 2023
University of Chicago & UC San Diego

MINDS Data Science Fellowship Summer 2021
Johns Hopkins University

Best Reviewers (Top 10%) 2021
ICML 2021

References

Peter L. Bartlett (peter@berkeley.edu) University of California, Berkeley

Vladimir Braverman (vova@cs.jhu.edu) Johns Hopkins University

Jason D. Lee (jasondlee@berkeley.edu) University of California, Berkeley

Sham M. Kakade (sham@seas.harvard.edu) Harvard University

Matus Telgarsky (mjt10041@nyu.edu) New York University

Bin Yu (binyu@berkeley.edu) University of California, Berkeley