

吴磊

北京市海淀区颐和园路 5 号, 北京大学智华楼 372
leiwu@math.pku.edu.cn • leiwu0.github.io • Google Scholar

工作经历	▪ 北京大学, 中国北京 助理教授, 数学科学学院、国际机器学习研究中心	2021.12 – 至今
	▪ 宾夕法尼亚大学, 美国费城 博士后, 统计与数据科学系 (合作导师: 苏炜杰)	2021.08 – 2021.11
	▪ 普林斯顿大学, 美国普林斯顿 博士后, 应用与计算数学项目 (PACM; 合作导师: 鄂维南)	2018.11 – 2021.07
教育经历	▪ 北京大学, 中国北京 博士, 计算数学 (导师: 鄂维南)	2012.09 – 2018.07
	▪ 南开大学, 中国天津 学士, 数学	2008.09 – 2012.06
研究方向	▪ 深度学习的科学基础 尺度律、学习动力学、逼近与泛化。	
	▪ 大规模预训练 优化器、超参、架构与数据。	
科研项目	▪ 青年科学基金项目 (B 类, 原优秀青年科学基金项目) 深度学习的数学理论; 主持, 200 万元。	2026 – 2028
	▪ 重大研究计划培育项目 基于稳定性的深度学习训练算法研究; 主持, 80 万元。	2025 – 2027
	▪ 北京大学科研启动经费 主持, 50 万元。	2022 – 2024
学术论文	同行评审论文*	
	<p>[38] Z. Wang[†], B. Li[†], J.-N. Wang[†], L. Wang[†], J. Wang[†], L. Wu, “When and How signSGD Outperform SGD: A Functional Scaling Law Perspective,” <i>NeurIPS 2026 (Spotlight)</i>.</p> <p>[37] B. Li[†], Z. Wang[†], F. Chen, S. Zhao, R. Zheng, L. Wu, “Optimal Learning-Rate Schedules under Functional Scaling Laws: Power Decay and Warmup-Stable-Decay,” <i>COLT 2026</i>.</p> <p>[36] J. Wang[†], M. Wang[†], J. Zhang, W. Wang, P. Pei, X. Cai, W. E, L. Wu, “GradPower: Powering Gradients for Faster Language Model Pre-Training,” <i>ICML 2026</i>.</p> <p>[35] J. Wang[†], B. Li[†], Z. Zhou, M. Wang, Y. Sun, J. Zhang, X. Cai, L. Wu, “Fast Catch-Up, Late Switching: Optimal Batch Size Scheduling via Functional Scaling Laws,” <i>ICLR 2026</i>.</p> <p>[34] M. Wang, L. Wu, “The Noise Geometry of Stochastic Gradient Descent,” <i>Transactions on Machine Learning Research</i>, 2026.</p> <p>[33] B. Li[†], F. Chen[†], Z. Huang[†], L. Wang[†], L. Wu, “Functional Scaling Laws in Kernel Regression: Loss Dynamics and Learning Rate Schedules,” <i>NeurIPS 2025 (Spotlight)</i>.</p> <p>[32] C. Zeng, Y. Zhang, J. Zhou, Y. Wang, Z. Wang, Y. Liu, L. Wu, D.Z. Huang, “Point Cloud Neural Operator for Parametric PDEs on Complex and Variable Geometries,” <i>Comput. Methods Appl. Mech. Eng.</i>, 2025.</p> <p>[31] J. Wang[†], M. Wang[†], Z. Zhou[†], J. Yan, W. E, L. Wu, “The Sharpness Disparity Principle in Transformers for Accelerating Language Model Pre-Training,” <i>ICML 2025</i>.</p> <p>[30] K. Zhan[†], P. Li[†], L. Wu, “Analyzing the role of permutation invariance in linear mode connectivity,” <i>AISTATS 2025</i>.</p> <p>[29] H. Chen[†], J. Long[†], L. Wu[†], “A duality framework for analyzing random feature and two-layer neural networks,” <i>Annals of Statistics</i>, 2025.</p>	

*† 表示作者按字母顺序排列或共同贡献。

- [28] M. Wang, J. Wang, H. He, Z. Wang, G. Huang, F. Xiong, Z. Li, **L. Wu**, “Improving generalization and convergence by enhancing implicit regularization,” *NeurIPS 2024*.
- [27] Z. Liu, M. Wang, H. Li, **L. Wu**, “Parameter symmetry and noise equilibrium of stochastic gradient descent,” *NeurIPS 2024*.
- [26] M.A. Mohamadi, Z. Li, **L. Wu**, D.J. Sutherland, “Why do you grok? A theoretical analysis on grokking modular addition,” *ICML 2024*.
- [25] K. Liu[†], Z. Wang[†], **L. Wu**, “The local landscape of phase retrieval under limited samples,” *IEEE Transactions on Information Theory*, 2024.
- [24] M. Wang, Z. Min, **L. Wu**, “Achieving margin maximization exponentially fast via progressive norm rescaling,” *ICML 2024*.
- [23] **L. Wu**, “Embedding inequalities for Barron-type spaces,” *Journal of Machine Learning*, 2023.
- [22] Z. Wang, **L. Wu**, “Theoretical analysis of inductive biases in deep convolutional networks,” *NeurIPS 2023*.
- [21] **L. Wu**, W.J. Su, “The implicit regularization of dynamical stability in stochastic gradient descent,” *ICML 2023*.
- [20] **L. Wu**, M. Wang, W.J. Su, “The alignment property of SGD noise and how it helps select flat minima: A stability analysis,” *NeurIPS 2022*.
- [19] C. Ma, D. Kunin, **L. Wu**, L. Ying, “Beyond the quadratic approximation: the multiscale structure of neural network loss landscapes,” *Journal of Machine Learning*, 2022.
- [18] **L. Wu**, J. Long, “A spectral-based analysis of the separation between two-layer neural networks and linear methods,” *Journal of Machine Learning Research*, 2022.
- [17] **L. Wu**, “Learning a single neuron for non-monotonic activation functions,” *AISTATS 2022*.
- [16] C. Ma[†], **L. Wu**[†], W. E, “A qualitative study of the dynamic behavior of adaptive gradient algorithms,” *Mathematical and Scientific Machine Learning (MSML)*, 2021.
- [15] W. E[†], C. Ma[†], **L. Wu**[†], “The Barron space and flow-induced function spaces for neural network models,” *Constructive Approximation*, 2021.
- [14] W. E[†], C. Ma[†], S. Wojtowytsch[†], **L. Wu**[†], “Towards a mathematical understanding of neural network-based machine learning: what we know and what we don’t,” *SIAM Transactions on Applied Mathematics*, 2020.
- [13] H. Lei, **L. Wu**, W. E, “Machine learning based non-Newtonian fluid model with molecular fidelity,” *Physical Review E*, 2020.
- [12] W. E[†], C. Ma[†], **L. Wu**[†], “Machine learning from a continuous viewpoint, I,” *Science China Mathematics*, 2020.
- [11] C. Ma[†], **L. Wu**[†], W. E, “The slow deterioration of the generalization error of the random feature model,” *Mathematical and Scientific Machine Learning (MSML)*, 2020.
- [10] W. E[†], C. Ma[†], **L. Wu**[†], “A comparative analysis of the optimization and generalization property of two-layer neural network and random feature models under gradient descent dynamics,” *Science China Mathematics*, 2020.
- [9] W. E[†], C. Ma[†], **L. Wu**[†], “The generalization error of minimum-norm solutions for over-parameterized neural networks,” *Journal of Pure and Applied Functional Analysis*, 2020.
- [8] **L. Wu**, Z. Zhu, “Towards understanding and improving the transferability of adversarial examples in deep neural networks,” *ACML*, 2020.
- [7] **L. Wu**[†], Q. Wang[†], C. Ma, “Global convergence of gradient descent for deep linear residual networks,” *NeurIPS 2019*.
- [6] W. E[†], C. Ma[†], **L. Wu**[†], “A priori estimates of the population risk for two-layer neural networks,” *Communications in Mathematical Sciences*, 2019.
- [5] Z. Zhu, J. Wu, B. Yu, **L. Wu**, J. Ma, “The anisotropic noise in stochastic gradient descent: Its behavior of escaping from minima and regularization effects,” *ICML 2019*.
- [4] **L. Wu**, C. Ma, W. E, “How SGD selects the global minima in over-parameterized learning: A dynamical stability perspective,” *NeurIPS 2018*.
- [3] Y.-A. Ma, E.B. Fox, T. Chen, **L. Wu**, “Irreversible samplers from jump and continuous Markov processes,” *Statistics and Computing*, 2018.
- [2] **L. Wu**, Z. Zhu, W. E, “Towards understanding generalization of deep learning: perspective of loss landscapes,” *Workshop on Principled Approaches to Deep Learning, ICML 2017*.

[1] H. Lei, N.A. Baker, **L. Wu**, G.K. Schenter, C.J. Mundy, A.M. Tartakovsky, “Smoothed dissipative particle dynamics model for mesoscopic multiphase flows in the presence of thermal fluctuations,” *Physical Review E*, 2016.

预印本

[9] J.-N. Wang[†], Z. Huang[†], K. Li[†], Z. Wang, **L. Wu**, “Momentum in large-batch training: Polyak enlarges the critical batch size, Nesterov improves data efficiency,” *arXiv preprint*, 2026.

[8] Z. Liu[†], R. Zheng[†], S. Zhang, C. Tian, K. Chen, Z. Zhang, **L. Wu**, “Effective Learning Rate Governs Loss Dynamics in Language Model Pretraining,” *arXiv preprint*, 2026.

[7] Y. Liu[†], Z. Wang[†], **L. Wu**[†], S. Zhang[†], “Smoothness Adaptivity in Constant-Depth Neural Networks: Optimal Rates via Smooth Activations,” *arXiv preprint*, 2026.

[6] Z. Lin[†], P. Li[†], **L. Wu**, “Exploring neural network landscapes: star-shaped and geodesic connectivity,” *arXiv preprint*, 2024.

[5] J. Long[†], X. Peng[†], **L. Wu**[†], “Optimal Rates and Saturation for Noiseless Kernel Ridge Regression,” *arXiv preprint*, 2024.

[4] H. Chen, J. Long, **L. Wu**, “The L^∞ Learnability of Reproducing Kernel Hilbert Spaces,” *arXiv preprint*, 2023.

[3] C. Ma[†], **L. Wu**[†], W. E, “The quenching-activation behavior of the gradient descent dynamics for two-layer neural network models,” *arXiv preprint*, 2020.

[2] Z. Li[†], C. Ma[†], **L. Wu**[†], “Complexity measures for neural networks with general activation functions using path-based norms,” *arXiv preprint*, 2020.

[1] W. E[†], C. Ma[†], **L. Wu**[†], “Analysis of the gradient descent algorithm for a deep neural network model with skip-connections,” *arXiv preprint*, 2019.

教学经历	北京大学	
	▪ 深度学习理论选讲	2024 – 至今
	▪ 机器学习数学导引	2023 – 至今
	▪ 高等数学（B）	2023
	▪ 数据科学导引	2022
	其他教学	
	▪ 机器学习数学导引，普林斯顿大学（共同授课）	2021
	▪ 深度学习暑期课程：	
	• 复旦大学	2026
	• 北京大学	2018、2019、2021
	• 上海交通大学	2020
学术组织	▪ 组织者，Foundations of Modern Intelligence (FoMI) Seminar（持续举办）。	
	▪ 共同组织者，PKU Workshop on Mathematical Foundations of Modern Machine Learning，2026。	
	▪ 共同组织者，the 3rd China Conference on Scientific Machine Learning (CSML)，2025。	
审稿服务	▪ 会议：NeurIPS、ICML、ICLR、AISTATS。	
	▪ 期刊：JMLR; SIAM J. Numer. Anal.; SIAM J. Math. Data Sci.; SIAM J. Math. Anal.; Constructive Approximation; J. Sci. Comput.; Ann. Appl. Probab.; Mathematical Programming。	