

MINGZE WANG

210, Jingyuan Building #6, [Peking University](#), Beijing, China, 100084

mingzewang@stu.pku.edu.cn

SHORT BIO

I am a fourth-year Ph.D candidate in Computational Mathematics, Peking University. I am very fortunate to be advised by [Prof. Weinan E](#). Prior to that, I received my B.S. degree in Pure and Applied Mathematics (ranking 1/111 for the first three years during my undergraduate study) from Zhejiang University in 2021. My homepage is <https://wmz9.github.io/>.

EDUCATION

Peking University

Ph.D Candidate, *Computational Mathematics*
School of Mathematical Sciences
Advisor: Prof. Weinan E.

Beijing, China
2021.09 - Present

Zhejiang University

Bachelor of Science, *Pure and Applied Mathematics*
School of Mathematical Sciences
Academic ranking: 1/111, Comprehensive ranking: 1/111, Major GPA: 4.84/5 (95.5/100).

Hangzhou, China
2017.09 - 2021.06

RESEARCH INTERESTS

I am broadly interested in theory, algorithm and application of machine learning. I am also interested in non-convex and convex optimization. Recently, I am also dedicated to use theory to design algorithms elegantly. Specifically, my recent research topics are

- **Deep learning theory**: theory and theory-inspired algorithm [1][2][3][4][5][6][8][9][10][11][12][13][15][16][17][18][19]
 - **Expressivity**: Explore the expressive power of Transformers through the lens of approximation theory [8][12]; the expressivity of mixture-of-experts (MoE) [16].
 - **Optimization**: Why can optimization algorithms converge to global minima when training neural networks [2][4][12]?
 - **Implicit Bias**: Why can optimization algorithms converge to global minima with favorable generalization ability when training neural networks? Flat-minima-bias [3][5][9][10][11]; max-margin-bias aspects [4][6].
 - **Generalization**: How to measure the generalization ability of neural networks [1].
 - **Algorithm Design**: For machine learning problems, design new provable optimization algorithms which can (i) converge faster / more stably [10][13][17][18][19]; (ii) generalize better [6][10].
- **Transformer and LLMs**: theory and algorithm, especially in LLM pre-training. [8][10][12][13][15][16][17][18][19]
 - **Expressivity**: The expressive power and mechanisms of Transformer [8][12]; the expressivity of Mixture-of-experts (MoE) [16]; the mechanisms of in-context learning [12].
 - **Algorithm Design**: Design provable faster/stabler optimizers for training LLMs [10][13][17][18][19]; design more efficient model architectures;
- **Non-convex and Convex Optimization**: theory and algorithm. [2][4][6][10][11][12][13][14][17][18][19]
 - **Convex Optimization in ML**. [6]
 - **Non-convex Optimization in ML**. [2][4][10][11][12][13][14][17][18][19]
 - **Algorithm Design**: Design provable faster / more stable optimizers for training neural networks [10][13][17][18][19] accelerate the convergence for the problems with specific structure [6].

Now, I am supported by the **Young Scientists (Ph.D) Fund of the National Natural Science Foundation of China (¥300,000)** (“Analyzing and Improving the Adam Optimizer for Foundation Model Training”).

	Expressivity & approximation power	Optimization & training dynamics	Generalization & implicit bias
Theory	<ul style="list-style-type: none"> transformer models work [8][12] mixture-of-experts models work [16] 	<ul style="list-style-type: none"> fully-connected networks work [2][4] transformer models work [12] 	<ul style="list-style-type: none"> flatness bias work [3][5][9][10][11] margin bias work [4][6]
Algorithm	<ul style="list-style-type: none"> more efficient models work [20] 	<ul style="list-style-type: none"> faster / stable convergence work [10][13][17][18] [19] 	<ul style="list-style-type: none"> better generalization work [6][10]

- Works [1]~[18] have been published, preprinted, or submitted.
- Works [19][20] is in preparation.

PUBLICATIONS & PREPRINTS

* indicates equal contribution; † means project lead.

19. Mingze Wang[†] et al., Conserved Quantities in Language Model Pre-Training: Theory and Applications. (In preparation)
18. Mingze Wang[†], Jinbo Wang, Jiaqi Zhang, Peng Pei, Wei Wang, Xunliang Cai, Weinan E, Lei Wu, **Grad-Power: Powering Gradients for Faster Language Model Pre-Training.** *arXiv preprint: 2505.24275, 1-22.* (submitted to **NeurIPS 2025**). 2025.
17. Shengtao Guo*, Mingze Wang*, Jinbo Wang, Lei Wu. **A Mechanistic Study of Transformer Training Instability under Mixed Precision.** (submitted to **NeurIPS 2025**). 2025.
16. Mingze Wang[†], Weinan E. **On the Expressive Power of Mixture-of-Experts for Structured Complex Tasks.** *arXiv preprint: 2505.24205, 1-18.* (submitted to **NeurIPS 2025**). 2025.
15. Tongcheng Zhang, Zhanpeng Zhou, Mingze Wang, Andi Han, Wei Huang, Taiji Suzuki, Junchi Yan. **On the Learning Dynamics of Two-layer ReLU Networks with Label Noise SGD.** *ICML 2025 Workshop on High-dimensional Learning Dynamics (ICML 2025 Workshop HiLD)*, 2025.
14. Tongtian Zhu, Tianyu Zhang, Mingze Wang, Zhanpeng Zhou, Can Wang. **A Single Global Merging Suffices: Recovering Centralized Learning Performance in Decentralized Learning.** *ICLR 2025 Workshop on Weight Space Learning (ICLR 2025 Workshop WSL)*. 2025.
13. Jinbo Wang*, Mingze Wang*[†], Zhanpeng Zhou*, Junchi Yan, Weinan E, Lei Wu. **The Sharpness Disparity Principle in Transformers for Accelerating Language Model Pre-Training.** *International Conference on Machine Learning (ICML 2025)*, 1-23. 2025.
12. Mingze Wang[†], Ruoxi Yu, Weinan E, Lei Wu. **How Transformers Get Rich: Approximation and Dynamics Analysis.** *arXiv preprint: 2410.11474, 1-47.* *ICML 2025 Workshop on High-dimensional Learning Dynamics (ICML 2025 Workshop HiLD)*, 2024.
11. Zhanpeng Zhou*, Mingze Wang*, Yuchen Mao, Bingrui Li, Junchi Yan. **Sharpness-Aware Minimization Efficiently Selects Flatter Minima Late in Training.** *International Conference on Learning Representations (ICLR 2025, Spotlight (Top 5.1%))*, 1-31. 2024.
10. Mingze Wang[†], Jinbo Wang, Haotian He, Zilin Wang, Guanhua Huang, Feiyu Xiong, Zhiyu Li, Weinan E, Lei Wu. **Improving Generalization and Convergence by Enhancing Implicit Regularization.** *Conference on Neural Information Processing Systems (NeurIPS 2024)*, 1-44. 2024.
9. Liu Ziyin, Mingze Wang, Hongchao Li, Lei Wu. **Loss Symmetry and Noise Equilibrium of Stochastic Gradient Descent.** *Conference on Neural Information Processing Systems (NeurIPS 2024)*, 1-26. 2024.

8. **Mingze Wang**, Weinan E. **Understanding the Expressive Power and Mechanisms of Transformer for Sequence Modeling.**
Conference on Neural Information Processing Systems (NeurIPS 2024), 1-76. 2024.
7. Guanhua Huang, Yuchen Zhang, Zhe Li, Yongjian You, **Mingze Wang**, Zhouwang Yang. **Are AI-Generated Text Detectors Robust to Adversarial Perturbations?**
Annual Meeting of the Association for Computational Linguistics (ACL 2024), 1-20. 2024.
6. **Mingze Wang**[†], Zeping Min, Lei Wu. **Achieving Margin Maximization Exponentially Fast via Progressive Norm Rescaling.**
International Conference on Machine Learning (ICML 2024), 1-38. 2023.
5. **Mingze Wang**, Lei Wu. **A Theoretical Analysis of Noise Geometry in Stochastic Gradient Descent.**
NeurIPS 2023 Workshop on Mathematics of Modern Machine Learning (NeurIPS 2023 Workshop M3L).
arXiv preprint: 2310.00692, 1-30. 2023.
4. **Mingze Wang**[†], Chao Ma. **Understanding Multi-phase Optimization Dynamics and Rich Nonlinear Behaviors of ReLU Networks.**
Conference on Neural Information Processing Systems (NeurIPS 2023, Spotlight (Top 3.5%)), 1-94. 2023.
3. Lei Wu, **Mingze Wang**, Weijie J. Su. **The alignment property of SGD noise and how it helps select flat minima: A stability analysis.**
Conference on Neural Information Processing Systems (NeurIPS 2022), 1-25. 2022.
2. **Mingze Wang**[†], Chao Ma. **Early Stage Convergence and Global Convergence of Training Mildly Parameterized Neural Networks.**
Conference on Neural Information Processing Systems (NeurIPS 2022), 1-73. 2022.
1. **Mingze Wang**[†], Chao Ma. **Generalization Error Bounds for Deep Neural Networks Trained by SGD.** Under review. *arXiv preprint: 2206.03299*, 1-32. 2022.

SERVICE

Conference: Conference on Neural Information Processing Systems (**NeurIPS**); International Conference on Machine Learning (**ICML**); International Conference on Learning Representations (**ICLR**); Artificial Intelligence and Statistics (**AISTATS**).

Journal: Journal of Machine Learning Research (**JMLR**); Transactions on Pattern Analysis and Machine Intelligence (**TPAMI**); Pattern Recognition (**PR**); Transactions on Machine Learning Research (**TMLR**); Journal of Machine Learning (**JML**); Communications in Mathematical Sciences (**CMS**).

SELECTED AWARDS & HONOURS

Young Scientists (Ph.D) Fund of the National Natural Science Foundation of China (300,000 RMB).	2024.12
National Scholarship (top 0.2% in the nation; 30,000 RMB), The Ministry of Education.	2024.09
Principal Scholarship (70,000 RMB), Peking University.	2024, 2025.05
BICMR Mathematical Award for Graduate Students (top 1%; 110,000 RMB), Peking University.	2023.11
Schlumberge Scholarship (30,000 RMB), Peking University.	2022.10
PKU Academic Innovation Award (top 1%), Peking University.	2022.10
Outstanding Graduate of Zhejiang Province (top 5%); Outstanding Graduate of ZJU	2021.05
National Scholarship (top 0.2% in the nation)	2019.10
First Class Scholarship of ZJU (top 3%)	2019, 2020.10
Zhejiang Provincial Government Scholarship	2018.10
First Prize of Mathematical Contest in Modeling of ZJU (top 1%)	2020.06
Meritourious Award in The Mathematical Contest in Modeling	2020.02
National Second Prize of Chinese Undergraduate Mathematical Contest in Modeling (top 2.5%)	2019.10

TEACHING

Peking University

Teaching assistant: Deep Learning Theory, taught by [Prof. Zhiyuan Li](#) (TTIC)

Teaching assistant: Calculus (A)

Teaching assistant: Calculus (B)

Beijing, China

Summer School 2023.

Fall 2021

Fall 2022, 2023, 2024; Spring 2022, 2023, 2024

EXPERIENCE

LLM pre-training:

- **Meituan, LLM pretraining group**

Algorithm Intern

Work on designing stable and faster optimization algorithms for LLM pretraining [18][19].

Beijing, China

2025.01 - 2025.05

- **Institute for Advanced Algorithms Research, LLM group**

Algorithm Intern

Work on designing faster optimizers for LLM pretraining [10][13].

Shanghai, China

2023.12 - 2024.08

- **Xiaohongshu, LLM group (Expected)**

Algorithm Intern (*Redstar top intern* program)

Will work on exploring better scaling laws for LLM pretraining.

Beijing, China

Expected (2025)

LLM post-training:

- **Tencent, LLM group (Expected)**

Algorithm Intern (*Qingyun top intern* program)

Will work on designing better RLHF algorithms.

Beijing, China

Expected (2025)

Quantitative trading:

- **Wizard Quant (Expected)**

Algorithm Intern (*top intern* program)

Will work on designing better machine learning models in quantitative trading.

Shanghai, China

Expected (2025)

Others:

- **Moqi Technology**

Algorithm Intern

Work on privacy-preserving computer vision.

Beijing, China

2021.09 - 2022.06