

# MINGZE WANG

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

[mingzewang@stu.pku.edu.cn](mailto: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]
  - **Expressivity**: Explore the expressive power of Transformers through the lens of approximation theory [9][12]; the expressivity of state-space models.
  - **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 optimization algorithms which can (i) converge faster [10][13]; (ii) generalize better [6][10].
- **Transformer and Large Language Models**: theory and algorithm. [8][10][12][13]
  - **Expressivity**: The expressive power and mechanisms of Transformer [8][12]; the mechanisms of in-context learning [12]; the expressivity of state-space models.
  - **Algorithm Design**: Design faster optimizers for training LLMs [10][13]; design more efficient model architectures; design more efficient strategy for data selection.
- **Non-convex and Convex Optimization**: theory and algorithm. [2][4][6][10][11][12][13][14]
  - **Convex Optimization in ML**. [6]
  - **Non-convex Optimization in ML**. [2][4][10][11][12][13][14]
  - **Algorithm Design**: Design faster optimizers for training neural networks [10][13]; accelerate the convergence for the problems with specific structure [6].
- **Computer vision and Natural language processing**: algorithm and application [7][10][13].

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”).

	Optimization & training dynamics	Generalization & implicit bias	Expressivity & approximation power
Theory	<ul style="list-style-type: none"> <li>fully-connected networks work [2][4]</li> <li>transformer models work [12]</li> </ul>	<ul style="list-style-type: none"> <li>flatness bias work [3][5][9][10][11]</li> <li>margin bias work [4][6]</li> </ul>	<ul style="list-style-type: none"> <li>transformer models work [8][12]</li> </ul>
Algorithm	<ul style="list-style-type: none"> <li>faster convergence work [10][13]</li> </ul>	<ul style="list-style-type: none"> <li>better generalization work [6][10]</li> </ul>	<ul style="list-style-type: none"> <li>more efficient models work [15]</li> </ul>

- Works [1]~[14] have been published or preprinted.
- Work [15] are on the preparation.

## PUBLICATIONS & PREPRINTS

14. Tongtian Zhu, Tianyu Zhang, Mingze Wang, Zhanpeng Zhou, Can Wang. **A Single Global Merging Suffices: Recovering Centralized Learning Performance in Decentralized Learning.**  
submitted to **ICML 2025**. 2025.
13. Jinbo Wang\*, Mingze Wang\*<sup>†</sup>, Zhanpeng Zhou\*, Junchi Yan, Weinan E, Lei Wu. **The Sharpness Disparity Principle in Transformers for Accelerating Language Model Pre-Training.**  
submitted to **ICML 2025**. *arXiv preprint: 2502.19002, 1-23. 2025.*
12. Mingze Wang<sup>†</sup>, Ruoxi Yu, Weinan E, Lei Wu. **How Transformers Get Rich: Approximation and Dynamics Analysis.**  
submitted to **ICML 2025**. *arXiv preprint: 2410.11474, 1-47. 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<sup>†</sup>, 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<sup>†</sup>, 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<sup>†</sup>, 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**<sup>†</sup>, 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**<sup>†</sup>, Chao Ma. **Generalization Error Bounds for Deep Neural Networks Trained by SGD.** Under review. *arXiv preprint: 2206.03299*, 1-32. 2022.

\* indicates equal contribution; † means project lead.

## 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**).

## 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. 2024.05  
 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** Beijing, China  
 Teaching assistant: Deep Learning Theory, taught by Prof. Zhiyuan Li (TTIC) Summer School 2023.  
 Teaching assistant: Calculus (A) Fall 2021  
 Teaching assistant: Calculus (B) Fall 2022, 2023, 2024; Spring 2022, 2023, 2024

## EXPERIENCE

**Meituan, LLM group** Beijing, China  
 Algorithm Intern 2025.01 - Present  
 Work on designing stable and faster optimization algorithms for LLM pretraining.

**Institute for Advanced Algorithms Research, LLM group** Shanghai, China  
 Algorithm Intern 2023.12 - 2024.08  
 Work on designing faster optimizers for LLM pretraining.

**Moqi Technology** Beijing, China  
 Algorithm Intern 2021.09 - 2022.06