

EDUCATION

University of Science and Technology of China (USTC)

Hefei, Anhui, China

School of the Gifted Young

Bachelor of Information and Computational Science

Sep 2021 - Jun 2025

• *GPA*: 3.71/4.3 (Overall/Scale) | *Ranking*: 20/95

RESEARCH INTERESTS

Keywords: Statistical Machine Learning; Foundation Model; Large Language Model; AI/ML for Science.

My research focuses on machine learning theory, with an emphasis on the statistical and computational capabilities of foundation models and generative AI. My goal is to develop new strong models and methods for important scientific and societal problems, with robust (i.e., low-assumption and realistic) theoretical guarantees to ensure their empirical performance.

With this aim, my research interests span machine learning, optimization, statistics, and computational complexity theory.

PUBLICATIONS

Google Scholar | arXiv | *Equal contribution. ^{$\alpha\beta$} Alphabetical authorship.

- [6] Provably Efficient Training for Conditional Diffusion Transformers,
Jerry Yao-Chieh Hu*, **Maojiang Su***, Junze Yin, Yufa Zhou, Shang Wu, Alex Daniel Reneau, Zhao Song, Han Liu,
Submitted to ICML 2025, ArXiv preprint forthcoming, 2025.
- [5] Transformers are Deep Optimizers: Provable In-Context Learning for Deep Model Training,
Weimin Wu*, **Maojiang Su***, Jerry Yao-Chieh Hu*, Zhao Song, Han Liu ,
Submitted to ICML 2025, Preprint <https://arxiv.org/abs/2411.16549>, 2024.
- [4] SPARQ: Outlier-Efficient SpeechLM with Fast Adaptation and Robust Quantization,
Shang Wu*, Yen-Ju Lu*, Haozheng Luo*, **Maojiang Su**, Jerry Yao-Chieh Hu, Jiayi Wang, Jing Liu, Najim Dehak,
Jesus Villalba, Han Liu,
Submitted to ICML 2025, Openreview, <https://openreview.net/forum?id=Z2uhdwOrn0>, 2024.
- [3] Making Genomic Foundation Models more Foundational Requires Outlier Removal: A Case Study on DNABERT-2,
Haozheng Luo*, Chenghao Qiu*, **Maojiang Su**, Zhihan Zhou, Jerry Yao-Chieh Hu, Zoe Mehta, Guo Ye, Han Liu,
Submitted to ICML 2025, ArXiv preprint forthcoming, 2024.
- [2] Computational Limits of Low-Rank Adaptation (LoRA) for Transformer-Based Models,
Jerry Yao-Chieh Hu, **Maojiang Su**, En-Jui Kuo, Zhao Song, Han Liu,
Accepted by ICLR 2025, Preprint <https://arxiv.org/abs/2406.03136>, 2024.
- [1] Boolean operations on generalized polygons with arcs,
Xiaolong Feng, **Maojiang Su**, Weihua Tong, Falai Chen,
Accepted by Journal of Computer-Aided Design and Computer Graphics, ArXiv preprint forthcoming, 2024.

RESEARCH EXPERIENCES

Statistical Machine Learning Group, Northwestern University

Research Intern, with Prof. **Han Liu**

Feb 2024 - Present

Computational Limits of Low-Rank Adaptation (LoRA) for Transformer-Based Models

- Explored the computational limits of Low-Rank Adaptation (LoRA) update for fine-tuning transformer-based models.
- Identified a phase transition behavior in LoRA adaptation on transformer-based model.
- Proved the existence of nearly linear algorithms for gradient computation by controlling the LoRA update term by term.
- *Paper accepted by ICLR 2025.*

Provably Efficient Training for Conditional Diffusion Transformers

- Analyzed the computational feasibility of efficient training algorithms for conditional latent diffusion transformers.
- Proved the existence of an almost linear approximation for attention gradients with respect to inputs.

- Introduced a unified approximation framework to derive efficient training methods for all conditional DiT architectures.
- Relaxed the previous requirement that the latent encoder/decoder must be pretrained and frozen during model training.
- *Paper submitted to ICML 2025.*

Transformers are Deep Optimizers: Provable In-Context Learning for Deep Model Training

- Investigated the strong in-context learning capabilities of transformers model.
- Provided an construction of a pretrained transformer capable of simulating gradient descent steps on deep NN.
- Provided theoretical guarantees for the approximation within any given error.
- Extended our analysis to the more practical setting using Softmax-based transformers.
- *Paper submitted to ICML 2025.*

Outlier-Free SpeechLM with Fast Adaptation and Robust Quantization

- Proposed SpARQ, to tackle the outlier problem of Speech and Language multi-modal Models.
- Investigated how outliers stemming from cross-modal (speech and text) low-rank adaptation and post-training quantization stages affect the performance of the current Speech and Language multi-modal Models.
- Provided the expressive guarantee of Low-Rank Adaption for modified transformer model.
- *Paper submitted to ICML 2025.*

Making Genomic Foundation Models more Foundational Requires Outlier Removal: A Case Study on DNABERT-2

- Introduced an efficient genomic foundation model optimized for accessibility and adaptability.
- Provided the expressive guarantee of Low-Rank Adaption for modified transformer model.
- Identified the conditions for the existence of low-rank adapters.
- *Paper submitted to ICML 2025.*

Graphics & Geometric Computing Laboratory, University of Science and Technology of China

Thesis Research, with Prof. Falai Chen

Sept 2023 - Aug 2024

Boolean operations on generalized polygons with arcs

- Extended the Vatti algorithm to enable plane Boolean operations on polygons with arcs.
- Implemented and compared multiple algorithms and encapsulated the extended Vatti algorithm into a robust program.
- *Paper accepted by Journal of Computer-Aided Design and Computer Graphics.*

HONORS

• Anhui Province University Student Mathematics Competition Second Prize	2021
• USTC Outstanding Freshman Scholarship (Top 25%)	2021
• USTC Outstanding Student Scholarship (Bronze, Top 20%)	2022

SKILLS

- **Programming Languages:** C/C++, Python, \LaTeX .
- **Tool:** Pytorch, Tensorflow, Github.
- **Mathematical Skills:** Probability, Statistics, Linear Algebra, ODEs, PDEs, Mathematical Analysis, Complex Analysis.
- **Languages:** TOEFL score: 101 (R:27, L:28, S:22, W:24)