Hengyu Fu

Tel: (+86) 15680881723, E-mail: hengyuf@berkeley.edu

EDUCATION

UNIVERSITY OF CALIFORNIA, BERKELEY Sep. 2025 - Jul. 2030 (Expected)
Electrical Engineering and Computer Sciences

PEKING UNIVERSITY Sep. 2021 - Jul. 2025
School of Mathematical Sciences: Major: ***Mathematics****, Double Major:* ***Economics***

- **Academic Achievement:** Overall GPA: 3.9/4.0 (Top 5 of 230)
- **Selected Coursework:** Mathematical Analysis (I) 97.5, Mathematical Analysis (II) (H) 96, Geometry(I) 98, Advanced Algebra (I) 92.5, Advanced Algebra (II) 94.5, Probability Theory 99, Abstract Algebra 97, Statistical Thinking 92, Applied Stochastic Processes 97, Mathematical Statistics 100, Optimization Methods 93, Game Theory 98, Applied Regression Analysis 99, Bayesian Theory and Computation 95, Introduction to Biostatistics 99.
- **Honors and Awards:**
  First Prize in National Olympiad in Informatics in Provinces (2017), Gold Medal in China Mathematics Olympics Final (2020, top 60, national training team) , Peking University 2021 & 2024 "Jin Qin" Scholarship.

PUBLICATIONS

- **Hengyu Fu**, Zehao Dou, Jiawei Guo, Mengdi Wang, Minshuo Chen, "Diffusion Transformer Captures Spatial-Temporal Dependencies: A Theory for Gaussian Process Data", **ICLR 2025**, https://arxiv.org/abs/2407.16134.
- **Hengyu Fu***, Tianyu Guo*, Yu Bai, Song Mei, "What can a Single Attention Layer Learn? A Study Through the Random Features Lens", **NeurIPS 2023**, https://arxiv.org/abs/2307.11353.
- **Hengyu Fu**, Zhuoran Yang, Mengdi Wang, Minshuo Chen, "Unveil Conditional Diffusion Models with Classifier-free Guidance: A Sharp Statistical Theory", https://arxiv.org/abs/2403.11968.
- **Hengyu Fu**, Zihao Wang, Eshaan Nichani, Jason D. Lee, "Learning Hierarchical Polynomials of Multiple Nonlinear Features", **ICLR 2025,** https://arxiv.org/abs/2411.17201.
- Ruicheng Ao*, **Hengyu Fu***, David Simchi-Levi, "Two-stage Online Reusable Resource Allocation: Reservation, Overbooking and Confirmation Call", http://arxiv.org/abs/2410.15245.
  (*Equal Contribution).

WORKING EXPERIENCE

Machine Learning Researcher Mar. 2025 – May. 2025
Company: Wizard Quant, Beijing

- Developed Transformers-based models for stock price prediction.

RESEARCH EXPERIENCE

Non-Linear Feature Learning Capabilities of Three-Layer Neural Networks Jun. 2024 – Sep. 2024
Advisor: Prof. Jason D. Lee, Electrical and Computer Engineering and Computer Science, Princeton University.

- Demonstrated that three-layer neural networks efficiently capture the underlying non-linear features in the target functions with the first step of gradient descent.
- Proved that through layer-wise training, three-layer neural networks can learn hierachical polynomials of non-linear features with improved sample complexity compared to kernel methods.
- Empirically verified our theory through numerical experiments.

Learning Sequential Data with Transformers in Diffusion Models Mar. 2024 – Jun. 2024

Advisor: Prof. Mengdi Wang, Electrical and Computer Engineering and the Center for Statistics and Machine Learning, Princeton University & Prof. Minshuo Chen, Industrial Engineering and Management Sciences Northwestern University.

- Proved that Transformers are able to unroll algorithms that approximate the score function of sequential data distribution and capture the inner special-temporal dependencies with a mild sample complexity.
- Conducted experiments on semi-synthetic data to show that Diffusion Transformers (DiT) indeed learn the score function in the same way we proposed.

Statistical Theory of Conditional Diffusion Models Jul. 2023 – Mar. 2024

Advisor: Prof. Mengdi Wang, Electrical and Computer Engineering and the Center for Statistics and Machine Learning, Princeton University. & Prof. Zhuoran Yang, Statistics and Data Science, Yale University.

- Established approximation and estimation guarantees on the conditional score function with neural networks throughout the diffusion process under regularity assumptions on initial density function.
- Extended the estimation theory on classifier-free diffusion guidance and provided theoretical guarantee of conditional diffusion models in solving practical problems such as reward conditioned data generation, inverse problem, etc.

Learnability of the Attention Layers in Transformers Mar. 2023 - Jul. 2023

Advisor: Prof. Song Mei, Department of Statistics, University of California, Berkeley.

- Studied theoretically on the learning and generalization of a single multi-head attention layer in the random feature setting.
- Found that the random feature attention with non-zero mean weight initialization of the query-key matrix outperforms the zero-mean counterpart in learning certain target functions.
- Conducted experiments on simulated data to substantiate our theoretical findings and further illustrate the interplay between the sample size and the complexity of the target function.

SKILLS & OTHERS

- **Computer Skills:** Python (&PyTorch), MATLAB, C++, LaTeX
- **Language:** Chinese (Native), English (Fluent)
- **Conference Reviewing:** NeurIPS 2024, ICLR 2025, AISTATS 2025, ICML 2025