SIJIN CHEN

(+1) 609-937-9583 | chensj@princeton.edu | sjchen0.github.io | Google Scholar

EDUCATION

Princeton University   Princeton, NJ, United States
Ph.D. in Electrical and Computer Engineering   *Sep. 2023 – present*

- Dissertation advisor: Prof. Jason M. Klusowski; Cumulative GPA: 3.8/4.0
- Courses: machine learning theory, measure-theoretic probability, stochastic calculus, reinforcement learning theory

Chinese University of Hong Kong   Hong Kong
B.Sc. in Computer Science with Honours, First Class; Minor in Mathematics   *Sep. 2019 – Jun. 2023*

- Major GPA: 3.9/4.0; Cumulative GPA: 3.8/4.0
- Courses: approximation algorithms, stochastic processes, real analysis, abstract algebra

RESEARCH INTEREST

Foundational understanding of modern machine learning and large language models; optimization; statistics.

PUBLICATIONS

1. **Sijin Chen**, Omar Hagrass, and Jason M. Klusowski, *Decoding Game: On Minimax Optimality of Heuristic Text Generation Strategies*, International Conference on Learning Representations (ICLR), 2025. https://arxiv.org/abs/2410.03968

2. **Sijin Chen**, Zhize Li, and Yuejie Chi, *Escaping Saddle Points in Heterogeneous Federated Learning via Distributed SGD with Communication Compression*, International Conference on Artificial Intelligence and Statistics (AISTATS), 2024. https://proceedings.mlr.press/v238/chen24d.html

3. **Sijin Chen**, Xiwei Cheng, and Anthony Man-Cho So, *Non-Convex Joint Community Detection and Group Synchronization via Generalized Power Method*, International Conference on Artificial Intelligence and Statistics (AISTATS), 2024. https://proceedings.mlr.press/v238/chen24e.html

4. Wu Zheng, Weiliang Tang, **Sijin Chen**, Li Jiang, and Chi-Wing Fu, *CIA-SSD: Confident IoU-Aware Single-Stage Object Detector from Point Cloud*, AAAI Conference on Artificial Intelligence (AAAI), 2021. https://ojs.aaai.org/index.php/AAAI/article/view/16470

RESEARCH EXPERIENCES

Length generalization of transformers in sharper landscapes
advised by Prof. Jason M. Klusowski, Princeton University   *Jan. 2025 – present*

- Designed and conducted experiments on fundamental sequential tasks to evaluate length generalization of transformers, a key component of LLM reasoning abilities
- Identified abrupt increase of landscape sharpness accompanying the emergence of length generalization, in contrast to the classical learning theory that sharp landscape hinders generalization
- Explained the phenomenon as a consequence of feature learning in transformers, currently working on the theoretical justifications by deriving approximations to landscape Hessian using random matrix theory

Game-theoretic foundations of language model decoding strategies
advised by Prof. Jason M. Klusowski, Princeton University   *Jun. 2024 – Oct. 2024*

- Developed rigorous theory behind the heuristic design of truncation-based sampling methods for text generation
- Modeled text generation as a zero-sum game on log-likelihood maximization under adversarial shift of true distribution
- Established the implicit regularization induced by the adversary, deriving that truncation-based sampling is a first-order approximation to the minimax optimal solution
- Devised a decoding method corresponding to our optimality theory which consistently achieved superior performance in open-ended text generation tasks in Llama-2-7B, GPT-J-6B, and GPT-2 models

Second-order stationarity of compressed SGD for heterogeneous distributed learning
advised by Prof. Yuejie Chi, Carnegie Mellon University   *May 2022 – Oct. 2023*

- Designed an error-feedback mechanism for gradient compression in distributed learning with heterogeneous data
- Proved convergence of the algorithm to second-order stationary points by escaping saddle points, using martingale concentration arguments under non-convex smooth landscape
- Implemented a PyTorch optimizer based on the proposed algorithm, achieving empirical advantages over state-of-the-art methods in heterogeneous distributed learning

Provably fast methods for generalized community detection
advised by Prof. Anthony Man-Cho So, Chinese University of Hong Kong   *Jun. 2021 – Dec. 2021*

- Proposed a generalized power method (GPM) to solve a community detection problem with extra node information
- Established an estimation error bound for spectral initialization using random matrix and random graph arguments

- Proved a linear convergence guarantee for GPM, significantly reducing the time complexity from the existing semi-definite relaxation approach, while improving the information-theoretic limit from pure detection problems

3D computer vision via deep learning

advised by <u>Prof. Philip Chi-Wing Fu</u>, Chinese University of Hong Kong *Jun. 2020 – Nov. 2020*

- Cooperated with researchers to design 3D convolutional neural network models for autonomous driving scenes
- Proposed data augmentation methods for performance improvement, validated their efficiency for model training on the benchmark dataset KITTI

ACADEMIC ACTIVITIES

Presentations

The 27th International Conference on Artificial Intelligence and Statistics (AISTATS) *May 2024*

Reviewing

The 13th International Conference on Learning Representations (ICLR) *Oct. 2024 – Dec. 2024*

Teaching

SML310: Research Projects in Data Science *Fall 2024*

ORF570: Special Topics in Statistics and Operations Research: Transformers and Large Language Models *Spring 2025*

HONORS AND AWARDS

Princeton University SEAS Travel Grant	*Mar. 2025*
Gordon Y. S. Wu Fellowship in Engineering	*Aug. 2023*
Hong Kong Government Scholarship	*Dec. 2022*
VTech Group of Companies Scholarship	*Feb. 2022*
Best Project Award of Summer Research Internship	*Sep. 2020*

SKILLS

Languages: Mandarin Chinese (native), Cantonese (proficient)

Programming: Python, PyTorch, MATLAB, R, C/C++

Computing: Linux, Slurm cluster, parallel computing

Hobbies: soccer, volleyball, piano performance, classical music