Su Jiaji

Level 7, Block S16, 6 Science Drive 2
National University of Singapore, Singapore 117546

sujiaji.github.io
su_jiaji@foxmail.com

Education

Jan 2024 | **Doctor of Philosophy in Statistics**
Department of Statistics & Data Science
National University of Singapore, Singapore
Thesis Advisor: Dr. YAO Zhigang

Jun 2018 | **Bachelor of Natural Science in Statistics**
School of Mathematical Science
Zhejiang University, Hangzhou, China

Professional Experience

Since Oct 2023 | **Research Fellow** | Advisor: Dr. YAO Zhigang
Department of Statistics & Data Science
National University of Singapore

Aug 2022 – Oct 2023 | **Student Research Assistant** | Advisor: Dr. YAO Zhigang
Department of Statistics & Data Science
National University of Singapore

Selected Ongoing Projects

Since Jun 2025 | **Manifold Fitting for Biological Sequence Embeddings**
- Extends our manifold fitting framework to high-dimensional embeddings produced by Transformer-based DNA/RNA language models.
- Identifies low-dimensional structures that coherently organise sequences sharing biological functions, providing geometric insight into the "sequence language space."
- Incorporates the fitted manifolds into classification, clustering, and other downstream tasks, enhancing accuracy and interpretability over raw embeddings.

Since Jan 2024 | **Low-dimension Structure Analysis Based on UK Biobank**
- Applies manifold-fitting methods to large-scale biomedical data, with a focus on the UK Biobank.
- Aims to uncover low-dimensional structures within high-dimensional population health data.
- Leverages these structures to improve population stratification and refine medical risk assessments.

Preprints & Publications

*Equal Contribution; #Correspondence

Accepted:

*Li, B., *Su, J., *Lin, R., #Yau, S.-T., & #Yao, Z. (2025). Manifold fitting reveals metabolomic heterogeneity and disease associations in UK biobank populations. *Proceedings of the National Academy of Sciences*, *122*(22), e2500001122.

#*Yao, Z., *Su, J., & #*Yau, S.-T. (2024). Manifold fitting with cycleGAN. *Proceedings of the National Academy of Sciences*, *121*(5), e2311436121.

Su, J., #Yao, Z., Li, C., & Zhang, Y. (2023). A statistical approach to estimating adsorption-isotherm parameters in gradient-elution preparative liquid chromatography. *The Annals of Applied Statistics*, *17*(4), 3476–3499.

Preprints:

Su, J. & #Yao, Z. (2025). Principal decomposition with nested submanifolds. *arXiv preprint arXiv:2502.10010*. (*Submitted to Biometrika*).

#Yao, Z., Su, J., Li, B., & Yau, S.-T. (2023). Manifold fitting. *arXiv preprint arXiv:2304.07680*.

Talks & Presentations

May 2025 Beijing	YMSC Statistical Seminar *Yau Mathematical Sciences Center, Tsinghua University*
Jan 2025 Shanghai & Sanya	The Second Symposium of Geometry and Statistics in China *Shanghai Institute for Mathematics and Interdisciplinary Sciences & The Tsinghua Sanya International Mathematics Forum*
Oct 2024 Singapore	Interactions of Statistics and Geometry (ISAG) II *Institute for Mathematical Sciences, National University of Singapore*
Jun 2024 Shanghai	Summer Seminar Series: Statistics + Geometry + X *Shanghai Institute for Mathematics and Interdisciplinary Sciences*
Jul 2023 Beijing	Talk: 'Estimating Adsorption-isotherm Parameters' *School of Mathematics and Statistics, Beijing Institute of Technology*

Visiting

Jun 2024 – Jul 2024	Visiting Researcher Shanghai Institute for Mathematics and Interdisciplinary Sciences, China
Dec 2024 – Jan 2025	Visiting Researcher Shanghai Institute for Mathematics and Interdisciplinary Sciences, China

Teaching

Teaching assistant at National University of Singapore:

Semester	**Courses**	
2021–22 Sem 2	ST2334	Probability and Statistics
	DSA1101	Introduction to Data Science
	ST2137	Statistical Computing and Programming
2021–22 Sem 1	ST5213	Advanced Categorical Data Analysis
	DSA1101	Introduction to Data Science
2020–21 Sem 2	ST2334	Probability and Statistics
	ST3232	Design and Analysis of Experiments
2020–21 Sem 1	ST2334	Probability and Statistics
	ST5225	Statistical Analysis of Networks
2019–20 Sem 2	ST2131	Probability
	ST3239	Survey Methodology
2019–20 Sem 1	ST4231	Computer Intensive Statistical Methods
2018–19 Sem 2	ST2131	Probability