Yijun Dong

Curriculum Vitae

✉ *yd1319@nyu.edu*
dyjdongyijun.github.io

Research Interests

Randomized Numerical Linear Algebra, Learning Theory.
I am interested in the computational and sample efficiency of algorithms in machine learning and scientific computing. From the computational efficiency perspective, my work is centered on matrix sketching and randomized low-rank decompositions like SVD and CUR. From the sample efficiency perspective, my work focuses on the generalization and distributional robustness of learning algorithms in data-limited settings.

Employment

2023-Present **New York University, Courant Institute**, New York, NY, US,
Assistant Professor/Courant Instructor.

Education

2018-2023 **University of Texas at Austin, Oden Institute**, Austin, TX, US,
Ph.D. in Computational Science, Engineering, and Mathematics.
- Advisors: Prof. Per-Gunnar Martinsson, Prof. Rachel Ward.
- Thesis: Randomized Dimension Reduction with Statistical Guarantees.

2014-2018 **Emory University**, Atlanta, GA, US,
B.S. in Applied Mathematics & Engineering Science, Magna Cum Laude.
- Advisors: Prof. Effrosyni Seitaridou, Prof. Eric Weeks.
- Thesis: Crystals and Liquids in Gravitationally Confined Quasi-2D Colloidal Systems.

Preprints (∗ denotes equal contribution)

1. Greedy Output Approximation: Towards Efficient Structured Pruning for LLMs Without Retraining. Jianwei Li, Yijun Dong, Qi Lei. *arXiv:* 2407.19126, 2024.
2. Robust Blockwise Random Pivoting: Fast and Accurate Adaptive Interpolative Decomposition. Yijun Dong, Chao Chen, Per-Gunnar Martinsson, Katherine Pearce. *arXiv:* 2309.16002, 2023.
3. Adaptive Parallelizable Algorithms for Interpolative Decompositions via Partially Pivoted LU. Katherine Pearce, Chao Chen, Yijun Dong, Per-Gunnar Martinsson. *arXiv:* 2310.09417, 2023.

Conference Publications (∗ denotes equal contribution)

1. Sketchy Moment Matching: Toward Fast and Provable Data Selection for Fine-tuning. Yijun Dong*, Hoang Phan*, Xiang Pan*, Qi Lei. *Conference on Neural Information Processing Systems (NeurIPS)*, 2024. (to appear)
2. Cluster-aware Semi-supervised Learning: Relational Knowledge Distillation Provably Learns Clustering. Yijun Dong*, Kevin Miller*, Qi Lei, Rachel Ward. *Conference on Neural Information Processing Systems (NeurIPS)*, 2023.
3. Adaptively Weighted Data Augmentation Consistency Regularization for Robust

Optimization under Concept Shift. Yijun Dong*, Yuege Xie*, Rachel Ward. *International Conference on Machine Learning (ICML)*, 2023.

4. Sample Efficiency of Data Augmentation Consistency Regularization. Shuo Yang*, Yijun Dong*, Rachel Ward, Inderjit S Dhillon, Sujay Sanghavi, Qi Lei. *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2023.

Journal Publications (∗ denotes equal contribution)

1. Efficient Bounds and Estimates for Canonical Angles in Randomized Subspace Approximations. Yijun Dong, Per-Gunnar Martinsson, Yuji Nakatsukasa. *SIAM Journal on Matrix Analysis and Applications*, 2024. (to appear)
2. Simpler is better: A comparative study of randomized algorithms for computing the CUR decomposition. Yijun Dong, Per-Gunnar Martinsson. *Advances in Computational Mathematics*, 2023.
3. Quantifying Biofilm Formation of *Sinorhizobium meliloti* Bacterial Strains in Microfluidic Platforms by Measuring the Diffusion Coefficient of Polystyrene Beads. Chen Cheng*, Yijun Dong*, Matthew Dorian*, Farhan Kamili*, Effrosyni Seitaridou. *Open Journal of Biophysics*, 2017.

Teaching Experience

2023-Present **Instructor**, Courant Institute, New York University, New York, NY.
- Fall 2024: Computational Statistics.
- Spring 2024: Mathematics for Economics.
- Fall 2023: Discrete Mathematics.

Jul 2023 **Teaching Assistant**, Simons Laufer Mathematical Sciences Institute (SLMath) Summer Graduate School, IBM Almaden, San Jose, CA.
- Mathematics of Big Data: Sketching and (Multi-) Linear Algebra (TA for Drs. Kenneth Clarkson, Lior Horesh, Misha Kilmer, Tamara Kolda, and Shashanka Ubaru).

2020-2022 **Teaching Assistant**, UT Austin, Austin, TX.
- Fall 2022: Differential Equations with Linear Algebra (TA for Dr. Michael Novack).
- Fall 2021: Numerical Analysis: Linear Algebra (TA for Prof. Per-Gunnar Martinsson).
- Fall 2020: Differential Equations with Linear Algebra (TA for Prof. Sam Raskin).

2015-2016 **Student Tutor**, Emory University, Oxford, GA.
- Introduction to Physics, Modern Physics.

Service

Journal review Annals of Applied Probability, BIT Numerical Mathematics, Calcolo, IEEE Transactions on Signal Processing, IMA Journal of Numerical Analysis, Journal of Computational Mathematics and Data Science, SIAM Journal on Matrix Analysis and Applications, SIAM Journal on Scientific Computing.

Conference review AISTATS (2023), NeurIPS (2024).

Organization
- SIAM MDS24 minisymposium on “Efficient Computation and Learning with Randomized Sampling and Pruning”, with Yifan Chen, Qi Lei.

Talks

Robust Blockwise Random Pivoting: Fast and Accurate Interpolation Decomposition with Adaptiveness and Randomness
- SIAM Conference on Parallel Processing for Scientific Computing (PP24) minisymposium on "Randomized Methods in Linear Solvers and Matrix Factorizations", Baltimore, Maryland, Mar 2024.

Cluster-aware Semi-supervised Learning: Relational Knowledge Distillation Provably Learns Clustering
- Conference on Neural Information Processing Systems (NeurIPS 2023), New Orleans, Louisiana, Dec 2023. (poster)

Efficient Bounds and Estimates for Canonical Angles in Randomized Subspace Approximations
- 6th SIAM Texas-Louisiana Sectional Meeting (SIAM TX-LA 2023) minisymposium on "Nonlinear Algebra in Applications", Lafayette, Louisiana, Nov 2023.
- The International Council for Industrial and Applied Mathematics (ICIAM) minisymposium on "Randomized Numerical Linear Algebra", Tokyo, Japan, Aug 2023.
- Texas Women in Math Symposium (TWIMS2023), Austin, Texas, Mar 2023.

Adaptively Weighted Data Augmentation Consistency Regularization for Robust Optimization under Concept Shift
- International Conference on Machine Learning (ICML 2023), Honolulu, Hawaii, Jul 2023. (poster)
- 2023 Rising Stars in Computational and Data Sciences, Austin, Texas, Apr 2023.
- IPAM Workshop IV: Multi-Modal Imaging with Deep Learning and Modeling (CM-SWS4), Los Angeles, California, Nov 2022. (poster)

Sample Efficiency of Data Augmentation Consistency Regularization
- IFML Workshop 2023, Seattle, Washington, Apr 2023.
- Oden Institute CSEM Student Forum, Austin, Texas, Oct 2022.
- SIAM Conference on Mathematics of Data Science (MDS22), San Diego, California, Sep 2022. (poster)

Simpler is better: A comparative study of randomized algorithms for computing the CUR decomposition
- Jane Street Symposium 2022, Virtual, Jan 2022.
- SIAM Conference on Applied Linear Algebra (LA21), Virtual, May 2021. (poster)

Forming 2D colloidal crystals with sedimented colloids
- American Physical Society March Meeting, Los Angeles, California, Mar 2018.

Industrial Experience

Jun-Aug 2022 **Dell Technologies**, Research Intern, Austin, TX.
- Semi-supervised tabular learning with data augmentation and consistency regularization

May-Aug 2021 **Dell Technologies**, Research Intern, Austin, TX.
- Streaming telemetry time series compression on edge devices

Fellowships and Awards

2023	Graduate School Summer Fellowship	UT Austin
2023	Rising Stars in Computational and Data Sciences	UT Austin
2019-2020	NIMS Graduate Fellowship	UT Austin
2018-2019	Peter O'Donnell Graduate Fellowship	UT Austin

2018 Trevor Evans Award — Emory University

Awarded to top graduate of Emory Department of Mathematics

Skills

Programming
- Proficient: Bash, Git, LaTeX, MATLAB, Python.
- Prior knowledge: C++, IDL, Java, Julia.

Language
- Chinese (native), English (proficient), Japanese