Yijun Dong

Curriculum Vitae

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

Research Interests

Randomized Numerical Linear Algebra, Theoretical Machine Learning. 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 **Assistant Professor/Courant Instructor (Postdoc)**, *Courant Institute of Mathematical Sciences*, New York University, New York, NY, US.

Education

2018-2023 **Ph.D. in Computational Science, Engineering, and Mathematics**, *Oden Institute*, University of Texas at Austin, Austin, TX, US.

- Advisors: Prof. Per-Gunnar Martinsson, Prof. Rachel Ward
- Thesis: Randomized Dimension Reduction with Statistical Guarantees

2014-2018 **B.S. in Applied Mathematics & Engineering Science**, *Magna Cum Laude*, Emory University, Atlanta, GA, US.

- Advisors: Prof. Effrosyni Seitaridou, Prof. Eric Weeks
- Thesis: Crystals and Liquids in Gravitationally Confined Quasi-2D Colloidal Systems

Awards/Honors/Fellowships

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

Preprints (∗ for equal contribution)

1. **Yijun Dong**, Chao Chen, Per-Gunnar Martinsson, Katherine Pearce. "Robust Blockwise Random Pivoting: Fast and Accurate Adaptive Interpolative Decomposition". *arXiv preprint arXiv:2309.16002*, 2023.
2. **Yijun Dong**, Per-Gunnar Martinsson, Yuji Nakatsukasa. "Efficient Bounds and Estimates for Canonical Angles in Randomized Subspace Approximations". *arXiv preprint arXiv:2211.04676*, 2022.

3. Katherine J Pearce, Chao Chen, **Yijun Dong**, Per-Gunnar Martinsson. "Adaptive Parallelizable Algorithms for Interpolative Decompositions via Partially Pivoted LU". *arXiv preprint arXiv:2310.09417*, 2023.

Publications (∗ for equal contribution)

1. **Yijun Dong***, Kevin Miller*, Qi Lei, Rachel Ward. "Cluster-aware Semi-supervised Learning: Relational Knowledge Distillation Provably Learns Clustering". *Thirty-seventh Conference on Neural Information Processing Systems (NeurIPS)*, 2023.
2. **Yijun Dong***, Yuege Xie*, Rachel Ward. "Adaptively Weighted Data Augmentation Consistency Regularization for Robust Optimization under Concept Shift". *International Conference on Machine Learning (ICML)*, 2023.
3. Shuo Yang*, **Yijun Dong***, Rachel Ward, Inderjit S Dhillon, Sujay Sanghavi, Qi Lei. "Sample Efficiency of Data Augmentation Consistency Regularization". *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2023.
4. **Yijun Dong**, Per-Gunnar Martinsson. "Simpler is better: A comparative study of randomized algorithms for computing the CUR decomposition". *Advances in Computational Mathematics*, 2023.
5. Chen Cheng*, **Yijun Dong***, Matthew Dorian*, Farhan Kamili*, Effrosyni Seitaridou. "Quantifying Biofilm Formation of *Sinorhizobium meliloti* Bacterial Strains in Microfluidic Platforms by Measuring the Diffusion Coefficient of Polystyrene Beads". *Open Journal of Biophysics*, 7, no. 3 (2017): 157-173.

Selected Talks

1. SIAM Conference on Parallel Processing for Scientific Computing (PP24), Minisymposium on Randomized Methods in Linear Solvers and Matrix Factorizations (Baltimore, Maryland, Mar 2024): "Robust Blockwise Random Pivoting: Fast and Accurate Interpolation Decomposition with Adaptiveness and Randomness".
2. Thirty-seventh Conference on Neural Information Processing Systems (NeurIPS) (New Orleans, LA, Dec 2023): "Cluster-aware Semi-supervised Learning: Relational Knowledge Distillation Provably Learns Clustering". (poster)
3. 10th ICIAM Conference, Minisymposium on Randomized Numerical Linear Algebra (Tokyo, Japan, Aug 2023): "Efficient Bounds and Estimates for Canonical Angles in Randomized Subspace Approximations".
4. International Conference on Machine Learning (ICML) 2023 (Honolulu, HI, Jul 2023): "Adaptively Weighted Data Augmentation Consistency Regularization for Robust Optimization under Concept Shift". (poster)
5. 2023 Rising Stars in Computational and Data Sciences (Austin, TX, Apr 2023): "Adaptively Weighted Data Augmentation Consistency Regularization".
6. IPAM Workshop IV: Multi-Modal Imaging with Deep Learning and Modeling (CMSWS4) (Los Angeles, CA, Nov 2022): "AdaWAC: Adaptively Weighted Augmentation Consistency Regularization for Volumetric Medical Image Segmentation". (poster)
7. SIAM Conference on Mathematics of Data Science (MDS22) (San Diego, CA, Sep 2022): "Sample Efficiency of Data Augmentation Consistency Regularization". (poster)
8. Jane Street Symposium 2022 (New York, NY, Jan 2022): "Revitalize Classical Algorithms with Randomization: Efficient Low-rank Approximations with

Statistical Guarantees".

9. SIAM Conference on Applied Linear Algebra (LA21) (Virtual, May 2021): "A Randomized CUR Decomposition via Partially Pivoted LU Factorization". (poster)
10. American Physical Society March Meeting (Los Angeles, CA, March 2018): "Forming 2D colloidal crystals with sedimented colloids".

Service

Journal reviewer: SIAM Journal on Matrix Analysis and Applications (20'), IMA Journal of Numerical Analysis (22'), BIT Numerical Mathematics (22'), Calcolo (23'), Annals of Applied Probability (23'), Journal of Computational Mathematics and Data Science (23')

Conference reviewer: AISTATS (23')

Teaching Experience

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

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**, Department of Mathematics & Oden Institute, 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**, Department of Physics, Oxford College of Emory University, Oxford, GA.
- Introduction to Physics, Modern Physics

Industrial Experience

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

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

Skills

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

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