Yijun Dong

Updated August 22, 2025

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

Research Interests

My research is situated at the intersection of **randomized algorithms**, **high-dimensional probability**, and **statistical learning theory**, focusing on computation- and data-efficient algorithms for high-dimensional problems in machine learning and scientific computing. Algorithmically, I design fast and reliable randomized algorithms for dimension reduction, data selection, and model pruning. Theoretically, I work on the mathematical foundation of learning paradigms like data augmentation, knowledge distillation, and post-training alignment.

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: **Per-Gunnar Martinsson**, **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: **Effrosyni Seitaridou**, **Eric Weeks**.
- Thesis: Crystals and Liquids in Gravitationally Confined Quasi-2D Colloidal Systems.

Publications (∗ denotes equal contribution or alphabetical order)

Conference Publications

- Yijun Dong, Yicheng Li, Yunai Li, Jason D. Lee, Qi Lei. "Discrepancies are Virtue: Weak-to-Strong Generalization through Lens of Intrinsic Dimension." *International Conference on Machine Learning (ICML)*, 2025.
- Jianwei Li, Yijun Dong, Qi Lei. "Greedy Output Approximation: Towards Efficient Structured Pruning for LLMs Without Retraining." *Conference on Parsimony and Learning (CPAL)*, 2025.
- Yijun Dong*, Hoang Phan*, Xiang Pan*, Qi Lei. "Sketchy Moment Matching: Toward Fast and Provable Data Selection for Finetuning." *Conference on Neural Information Processing Systems (NeurIPS)*, 2024.
- Yijun Dong*, Kevin Miller*, Qi Lei, Rachel Ward. "Cluster-aware Semi-supervised Learning: Relational Knowledge Distillation Provably Learns Clustering." *Conference on Neural Information Processing Systems (NeurIPS)*, 2023.
- 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.

- 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.

Journal Publications

- Yijun Dong, Chao Chen, Per-Gunnar Martinsson, Katherine Pearce. "Robust Blockwise Random Pivoting: Fast and Accurate Adaptive Interpolative Decomposition." *SIAM Journal on Matrix Analysis and Applications*, 2025.
- Katherine Pearce, Chao Chen, Yijun Dong, Per-Gunnar Martinsson. "Adaptive Parallelizable Algorithms for Interpolative Decompositions via Partially Pivoted LU." *Numerical Linear Algebra with Applications*, 2025.
- Yijun Dong, Per-Gunnar Martinsson, Yuji Nakatsukasa. "Efficient Bounds and Estimates for Canonical Angles in Randomized Subspace Approximations." *SIAM Journal on Matrix Analysis and Applications*, 2024.
- Yijun Dong, Per-Gunnar Martinsson. "Simpler is better: A comparative study of randomized algorithms for computing the CUR decomposition." *Advances in Computational Mathematics*, 2023.
- 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*, 2017.

Preprints & Other Publications

- Yijun Dong*, Xiang Pan*, Hoang Phan*, Qi Lei. "Randomly Pivoted V-optimal Design: Fast Data Selection under Low Intrinsic Dimension." *Workshop on Machine Learning and Compression @ NeurIPS*, 2024.

Academic Visits

Talks (by topics)

- Discrepancies are Virtue: Weak-to-Strong Generalization through Lens of Intrinsic Dimension.
  - 2025/07 ICML 2025, Vancouver, Canada. (poster)
  - 2025/05 Flatiron Institute CCM ML Seminar, New York, NY. (invited)
  - 2025/04 John Hopkins University Postdoc Seminar, Virtual. (invited)
  - 2025/03 UT Austin Data & Algebra seminar, Virtual. (invited)
- Randomize Time Integration for Poorly Conditioned Dynamical Systems.
  - 2025/03 SIAM CSE25 minisymposium on "Scientific Machine Learning for Stable Prediction of Dynamical Systems", Fort Worth, TX.
- Sketchy Moment Matching: Toward Fast and Provable Data Selection for Finetuning.
  - 2025/01 JMM 2025 ILAS Special Session on Randomness in Numerical Linear Algebra, Seattle, WA.
  - 2024/12 NeurIPS 2024, Vancouver, Canada. (poster)
  - 2024/10 UDelaware Numerical Analysis & PDE seminar, Newark, DE. (invited)
- RBRP: Fast and Accurate Interpolation Decomposition with Adaptiveness and Randomness.

2025/06 Householder Symposium XXII, Ithaca, NY. (poster)

2024/10 SIAM MDS24 minisymposium on "Efficient Computation and Learning with Randomized Sampling and Pruning", Atlanta, GA. (poster)

2024/03 SIAM PP24 minisymposium on "Randomized Methods in Linear Solvers and Matrix Factorizations", Baltimore, MD.

- Relational Knowledge Distillation Provably Learns Clustering.

  2023/12 NeurIPS 2023, New Orleans, LA. (poster)

- Efficient Bounds and Estimates for Canonical Angles in Randomized Subspace Approximations.

  2023/11 SIAM TX-LA 2023 minisymposium on "Nonlinear Algebra in Applications", Lafayette, LA.

  2023/08 ICIAM minisymposium on "Randomized Numerical Linear Algebra", Tokyo, Japan.

  2023/03 Texas Women in Math Symposium, Austin, TX.

- Adaptively Weighted Data Augmentation Consistency Regularization for Robust Optimization under Concept Shift.

  2023/07 ICML 2023, Honolulu, HI. (poster)

  2023/04 2023 Rising Stars in Computational and Data Sciences, Austin, TX.

  2022/11 UCLA IPAM Workshop IV: Multi-Modal Imaging with Deep Learning and Modeling, Los Angeles, CA. (poster)

- Sample Efficiency of Data Augmentation Consistency Regularization.

  2023/04 IFML Workshop 2023, Seattle, WA.

  2022/10 Oden Institute CSEM Student Forum, Austin, TX.

  2022/09 SIAM MDS22, San Diego, CA. (poster)

- A comparative study of randomized algorithms for computing the CUR decomposition.

  2022/01 Jane Street Symposium 2022, Virtual.

  2021/05 SIAM LA21, Virtual. (poster)

- Forming 2D colloidal crystals with sedimented colloids.

  2018/03 American Physical Society March Meeting, Los Angeles, CA.

Other Academic Visits

2026/02 ICERM Workshop on Randomized Numerical Linear Algebra, Brown University, Providence, RI.

2025/08 IPAM Research Collaboration Workshop on Randomized Numerical Linear Algebra, UCLA, Los Angeles, CA.

2023/07 Simons Laufer Mathematical Sciences Institute (SLMath) Summer Graduate School, IBM Almaden, San Jose, CA.

Academic Service

Reviewer
- Journals: 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.
- Conferences: AISTATS, ICML, L4DC, NeurIPS.

Organizer
- SIAM MDS24 minisymposium on "Efficient Computation and Learning with Randomized Sampling and Pruning", with Yifan Chen and Qi Lei.

Teaching Experience

2023-Present **Instructor**, Courant Institute, New York University, New York, NY.
- Fall 2025: Probability and Statistics.
- Spring 2025: Introduction to Mathematical Modeling.
- Fall 2024: Computational Statistics.
- Spring 2024: Mathematics for Economics.
- Fall 2023: Discrete Mathematics.

2023/07 **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 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 Michael Novack).
- Fall 2021: Numerical Analysis: Linear Algebra (TA for Per-Gunnar Martinsson).
- Fall 2020: Differential Equations with Linear Algebra (TA for Sam Raskin).

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

Student Advising

Undergraduate Research

2025 Bowen Gong, NYU.

2024-2025 Yicheng Li, NYU Shanghai, (co-advised with Qi Lei).

2024-2025 Yunai Li, Shanghai Jiao Tong University, (co-advised with Qi Lei).

Distinctions & Fellowships

2025 AMS-Simons Travel Grant.

2023 Graduate School Summer Fellowship, UT Austin.

2023 Rising Stars in Computational and Data Sciences.

2019 NIMS Graduate Fellowship, UT Austin.

2018 Peter O'Donnell Graduate Fellowship, UT Austin.

2018 Trevor Evans Award, Emory University.

2017 Phi Beta Kappa & Sigma Pi Sigma, Emory University.

2016 SURE Summer Research Fellowship, Emory University.

2016 Dan C. Moore Mathematics Award, Emory University.

2016 Williams Baird Physics Award, Emory University.

Travel Awards (for individual conferences)

2025 Householder XXII Early Career Travel Support.

2024 SIAM Early Career Travel Award for PP24.

2022 Professional Development Award, UT Austin.

Industrial Experience

2022/06-2022/08 **Dell Technologies**, Research Intern, Austin, TX.
- Semi-supervised tabular learning with data augmentation and consistency regularization.

2021/05-2021/08 **Dell Technologies**, Research Intern, Austin, TX.
- Streaming telemetry time series compression on edge devices.

Skills

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

Language
- Chinese (native), English (fluent), Japanese (elementary).