Lorenzo Beretta

IBM Research, Cambridge, MA

• lorenzo2beretta@gmail.com • lorenzo2beretta.github.io

1 Research Interests

Theory of algorithms, AI safety, high-dimensional geometry, high-dimensional probability, sublinear algorithms, optimal transport.

2 Education

- **Ph.D. in Computer Science University of Copenhagen** Oct 2020 – Apr 2024
  Advised by Mikkel Thorup & Mikkel Abrahamsen
- **Scuola Normale Superiore (Honors Program)** 2015 – 2020
  Over five years, I completed two extra years of coursework, earning the equivalent of a master's in mathematics
- **M.S. Computer Science, University of Pisa** Sep 2018 – Oct 2020
  Grade 110/110 cum Laude
- **B.S. Mathematics, University of Pisa** Sep 2015 – Oct 2018
  Grade 110/110 cum Laude

3 Appointments

- **Goldstine Fellow at IBM Research** Sep 2025 – present
- **Postdoc at University of California, Santa Cruz** Sep 2024 – Sep 2025
  Hosted by Vaggos Chatziafratis

4 Awards

- **Herman Goldstine Postdoctoral Fellowship at IBM** 2025
- **Best Student Paper Award, SODA 2022** 2022
  for "Better Sum Estimation via Weighted Sampling", also invited to ***HALG 2022*** *and* ***TALG special issue***
- **EU TALENT Doctoral Fellowship Grant** 2020
  I was granted three-year PhD fellowship by the Horizon 2020 Marie Skłodowska-Curie program
- **Undergraduate Honors Program: Scuola Normale Superiore** 2015
  I was admitted to Scuola Normale Superiore, top math program in Italy (≈ 15 students admitted per year)

5 AI Safety Related Projects

- **Data Attribution Sketching (work in progress)**
  with Sam Gunn and Noah Golowich
  We design an algorithm for efficient data attribution in LLMs. Down-weighting fine-tuning datapoints and viewing the output of the LLM as a function of the down-weights, this function is well-approximated by a low-rank, low-degree polynomial, which we learn efficiently using sketching techniques.

- **Sketched Lanczos Uncertainty Score**

  with Marco Miani and Søren Hauberg, published at NeurIPS 2024 (C8)

  We design a low-memory uncertainty score for neural networks by combining the Lanczos algorithm with sketching to compress the Fisher information matrix into a compact summary. The resulting score has memory footprint sublinear in the number of parameters and achieves state-of-the-art out-of-distribution detection.

6 Publications

In theoretical computer science, authors are sorted alphabetically. All publications on this list follow this convention apart for C8, where the starred authors share first authorship as is customary in machine learning.

Preprints

- **Optimal Injectivity of the SparseStack**
  Lorenzo Beretta, Simone Bombari, Margalit Glasgow, Raphael Arkady Meyer, Kijun Shin, David P. Woodruff
  `https://arxiv.org/abs/2611.12867`

- **A Tight Analysis of Khatri-Rao Oblivious Subspace Embeddings**
  Lorenzo Beretta, Cameron Musco
  `https://arxiv.org/abs/2608.28094`

Peer Reviewed Conference Proceedings

- **Faster Estimation of the Average Degree of a Graph Using Random Edges and Structural Queries**
  Lorenzo Beretta, Deeparnab Chakrabarty, C. Seshadhri
  Symposium on Discrete Algorithms (SODA), 2026 C12

- **Feature Selection and Junta Testing are Statistically Equivalent**
  Lorenzo Beretta, Nathaniel Harms, Caleb Koch
  Symposium on Discrete Algorithms (SODA), 2026 C11

- **Approximating High-Dimensional Earth Mover's Distance as Fast as Closest Pair**
  Lorenzo Beretta, Vincent Cohen-Addad, Rajesh Jayaram, Erik Waingarten
  Symposium on Foundations of Computer Science (FOCS), 2025 C10

- **New Statistical and Computational Results for Learning Junta Distributions**
  Lorenzo Beretta
  International Conference on Randomization and Computation (RANDOM), 2025 C9

- **Sketched Lanczos uncertainty score: a low-memory summary of the Fisher information**
  Marco Miani*, Lorenzo Beretta*, Søren Hauberg
  Advances in Neural Information Processing Systems (NeurIPS), 2024 C8

- **Online sorting and online TSP: randomized, stochastic, and high-dimensional**
  Mikkel Abrahamsen, Ioana Bercea, Lorenzo Beretta, Jonas Klausen, László Kozma
  European Symposium on Algorithms (ESA), 2024 C7

- **Approximate Earth Mover's Distance in Truly-Subquadratic Time**
  Lorenzo Beretta, Aviad Rubinstein
  Symposium on Theory of Computing (STOC), 2024 C6

- **Multi-Swap k-Means++**
  Lorenzo Beretta, Vincent Cohen-Addad, Silvio Lattanzi, Nikos Parotsidis
  Advances in Neural Information Processing Systems (NeurIPS), 2023 C5

- **Locally Uniform Hashing**
  Ioana Bercea, Lorenzo Beretta, Jakob Bæk Tejs Houen, Jonas Klausen, Mikkel Thorup
  Symposium on Foundations of Computer Science (FOCS), 2023 C4

- **Online Sorting and Translational Packing of Convex Polygons**
  Anders Aamand, Mikkel Abrahamsen, Lorenzo Beretta, Linda Kleist
  Symposium on Discrete Algorithms (SODA), 2023 C3
- **Better Sum Estimation via Weighted Sampling**
  Lorenzo Beretta, Jakub Tetek
  Symposium on Discrete Algorithms (SODA), 2022
  Best Student Paper Award
  invited to HALG 2022 and TALG special issue C2
- **Online Packing to Minimize Area or Perimeter**
  Mikkel Abrahamsen, Lorenzo Beretta
  International Symposium on Computational Geometry (SoCG), 2021 C1

Journal Publications

Better Sum Estimation via Weighted Sampling
Lorenzo Beretta, Jakub Tetek
Transactions on Algorithms (TALG) J2

An Optimal Algorithm to Find Champions of Tournament Graphs
Lorenzo Beretta, Franco Maria Nardini, Roberto Trani, Rossano Venturini
IEEE Transactions on Knowledge and Data Engineering (TKDE) J1

7 Academic Service

- **Conference PC**
  ESA (2026)
- **Conference Review**
  ESA (2022, 2023, 2024), FOCS (2023, 2026), ICALP (2025), ISAAC (2024), ITCS (2024), NeurIPS (2025), SoCG (2026), SODA (2026, 2027), SOSA (2024, 2026, 2027), STOC (2025), SWAT (2024)
- **Journal Review**
  Journal of Computational Geometry (JoCG)

8 Internships

- **Stanford University** Feb 2023 – Aug 2023
  I worked with Aviad Rubinstein on sublinear-time matching algorithms
- **Google Research** Oct 2022 – Dec 2022
  I worked on designing, analyzing and implementing clustering algorithms
- **Quantitative Investment Internship, Société Générale** Aug 2019 – Oct 2019
  I worked on implementing efficient algorithms for optimal execution

9 Teaching and Leadership

- **Teaching Assistant, Copenhagen University** 2021 – 2022
  I served twice as T.A. for the "Randomized Algorithms" graduate course at Copenhagen University
- **Scout Leader Volunteer** 2017 – 2020
  I volunteered as a Scout leader, managing a troop of about 30 teenagers