

Ankit Pensia

E-mail: ankitpensia94@gmail.com
Mob : +1 (608)-695-6931

Website: <https://ankitp.net>
Google Scholar, dblp

Research machine learning, robust statistics, differential privacy, high-dimensional statistics, algorithmic fairness

Employment **Simons Institute for the Theory of Computing, UC Berkeley** Aug'24 - ongoing
Research Fellow, Modern Paradigms in Generalization

Postdoctoral Fellow, Resilience Research Pod
Mentors: Shafi Goldwasser and Venkat Guruswami

IBM Research '23 - '24
Herman Goldstine Postdoctoral Fellow

Education **University of Wisconsin-Madison** '17 - '23
Ph.D., Department of Computer Sciences
Advisors: Po-Ling Loh, Varun Jog, and Ilias Diakonikolas
Title: *Efficient Statistical Inference under Sampling and Computational Constraints*
Graduate Student Research Award for the best Ph.D. thesis in the department

Indian Institute of Technology Kanpur '12 - '17
B.Tech. - M.Tech. (Dual Degree), Department of Electrical Engineering

Internship **Research Intern**, Google Research, New York Summer '22
Mentors: Dr. Pranjal Awasthi and Dr. Satyen Kale
Developed a scalable learning algorithm under fairness constraints and implemented it in Tensorflow (published in ALT 2024).

Selected Awards
Simons-Berkeley Research Fellowship for 2024
IBM Herman Goldstine Postdoctoral Fellowship
Outstanding Graduate Student Research Award, 2023 (for the best Ph.D. Thesis), University of Wisconsin-Madison Computer Sciences Department
Recognized as a top reviewer by ICML 2020, ICLR 2021, NeurIPS 2022, ICML 2024
Travel Awards: NeurIPS 2022 Scholar Award, ISIT 2018, ISIT 2019, ISIT 2022
Special CS Scholarship, UW-Madison for the year 2017-2018
Academic Excellence Award, IIT Kanpur for the year 2015-2016

Representative Publications ▶ SoS Certifiability of Subgaussian Distributions and its Algorithmic Applications with† I. Diakonikolas, S. Hopkins, and S. Tiegel
Manuscript, 2024.

We prove that for each subgaussian distribution, there is a small sum-of-squares proof that its moments are bounded. This result is surprising because it goes against the conventional wisdom of the community. Our result dramatically expands the scope of the current algorithmic toolkit to more general distributions.

▶ A Sub-Quadratic Time Algorithm for Robust Sparse Mean Estimation **ICML'24 (spotlight)**
A. Pensia
International Conference on Machine Learning, 2024.

We provide the first subquadratic time algorithm for robust sparse mean estimation, thus answering an open problem posed by Cheng and Diakonikolas. Our techniques overcome the $\Omega(d^2)$ bottleneck arising from the need to use the sample covariance matrix to detect outliers. More details are available in this talk.

- ▶ The Sample Complexity of Simple Binary Hypothesis Testing COLT'24
A. Pensia, V. Jog, and P. Loh
Conference on Learning Theory, 2024.

We revisit the fundamental statistical problem of simple binary hypothesis testing. Despite being studied since the early 1900s, the tight non-asymptotic rates of this problem were unknown (in contrast, the asymptotic error rates were very well-understood). We close this century-old gap and derive optimal non-asymptotic bounds.

- ▶ Simple Binary Hypothesis Testing under Local Differential Privacy and Communication Constraints Trans. Inf. Theory
COLT'23
A. Pensia, A. Asadi, V. Jog, and P. Loh
IEEE Transactions on Information Theory, 2024.
An extended abstract appeared at COLT 2023.

We study simple binary hypothesis testing in the presence of local differential privacy and/or communication constraints. While the very-high privacy regime ($\epsilon \ll 1$) was well-understood, practical applications operate in moderate-privacy regime ($\epsilon \gg 1$). In this regime, we show new **surprising phenomenon**, both algorithmic and statistical. For example, sometimes increasing the privacy budget does not improve statistical accuracy.

- ▶ Streaming Algorithms for High-Dimensional Robust Statistics ICML'22
with† I. Diakonikolas, D.M. Kane, and T. Pittas
International Conference on Machine Learning, 2022.

For outlier-robust algorithms, we develop the first (runtime and sample-efficient) streaming algorithm with (near-optimal) **linear memory** usage for a variety of high-dimensional tasks. All the prior robust algorithms needed to store the entire dataset in memory, thus requiring at least quadratic memory.

- ▶ Outlier Robust Mean Estimation with Subgaussian Rates via Stability NeurIPS'20
with† I. Diakonikolas and D. M. Kane
Advances in Neural Information Processing Systems, 2020.

We show that recent outlier-robust algorithms also achieve subgaussian confidence intervals for heavy-tailed distributions, and vice-versa. In particular, we identify the "stability" condition as the **bridge** between these two contamination models.

Remaining Publications (chronological)

- ▶ SoS Certificates for Sparse Singular Values and Their Applications: Robust Statistics, Subspace Distortion, and More
with† I. Diakonikolas, S. Hopkins, and S. Tiegel
Manuscript, 2024.
- ▶ Optimal Robust Estimation under Local and Global Corruptions: Stronger Adversary and Smaller Error
with† T. Pittas
Manuscript, 2024.
- ▶ Robust Regression with Covariate Filtering: Heavy Tails and Adversarial Contamination JASA
A. Pensia, V. Jog, and P. Loh
Journal of the American Statistical Association, 2024.
- ▶ Black-Box k -to-1-PCA Reductions: Theory and Applications COLT'24
with† A. Jambulapati, S. Kumar, J. Li, S. Pandey, K. Tian
Conference on Learning Theory, 2024.
- ▶ Robust Sparse Estimation for Gaussians with Optimal Error under Huber Contamination ICML'24
with† I. Diakonikolas, D.M. Kane, S. Karmalkar, and T. Pittas
International Conference on Machine Learning, 2024.

†Alphabetical ordering as per the convention in theoretical computer science.

- ▷ Semi-supervised Group DRO: Combating Sparsity with Unlabeled Data
with† P. Awasthi and S. Kale
Algorithmic Learning Theory, 2024. ALT'24
- ▷ Communication-constrained hypothesis testing:
Optimality, robustness, and reverse data processing inequalities
A. Pensia, V. Jog, and P. Loh
IEEE Transactions on Information Theory, 2024.
A shorter version appeared at ISIT 2022. Trans. Inf. Theory
ISIT'22
- ▷ A Spectral Algorithm for List-Decodable Covariance Estimation
in Relative Frobenius Norm
with† I. Diakonikolas, D.M. Kane, J. C. H. Lee, and T. Pittas
Advances in Neural Information Processing Systems, 2023. NeurIPS'23
(spotlight)
- ▷ Near-Optimal Algorithms for Gaussians with Huber Contamination:
Mean Estimation and Linear Regression
with† I. Diakonikolas, D.M. Kane, and T. Pittas
Advances in Neural Information Processing Systems, 2023. NeurIPS'23
- ▷ Nearly-Linear Time and Streaming Algorithms for Outlier-Robust PCA
with† I. Diakonikolas, D.M. Kane, and T. Pittas
International Conference on Machine Learning, 2023. ICML'23
- ▷ Gaussian Mean Testing Made Simple
with I. Diakonikolas and D. M. Kane
SIAM Symposium on Simplicity in Algorithms, 2023. SOSA'23
- ▷ Outlier-Robust Sparse Mean Estimation for Heavy-Tailed Distributions
with† I. Diakonikolas, D. M. Kane, and J. C. H. Lee
Advances in Neural Information Processing Systems, 2022 NeurIPS'22
- ▷ List-Decodable Sparse Mean Estimation via Difference-of-Pairs Filtering
with† I. Diakonikolas, D.M. Kane, S. Karmalkar, and T. Pittas
Advances in Neural Information Processing Systems, 2022. NeurIPS'22
(oral)
- ▷ Robust Sparse Mean Estimation via Sum of Squares
with† I. Diakonikolas, D.M. Kane, S. Karmalkar, and T. Pittas
Conference on Learning Theory, 2022. COLT'22
- ▷ Sharp Concentration Inequalities for the Centered Relative Entropy
with† A. Bhatt
Information and Inference: A Journal of the IMA, 2022. Information
and Inference
- ▷ Statistical Query Lower Bounds for List-Decodable Linear Regression
with† I. Diakonikolas, D. M. Kane, T. Pittas, and A. Stewart
Advances in Neural Information Processing Systems, 2021. NeurIPS'21
(spotlight)
- ▷ Estimating Location Parameters in Sample-heterogeneous Distributions
A. Pensia, V. Jog, and P. Loh
Information and Inference: A Journal of the IMA, 2021.
A shorter version appeared at ISIT 2019. Information
and Inference
- ▷ Optimal Lottery Tickets via SubsetSum:
Logarithmic Over-Parameterization is Sufficient
A. Pensia*, S. Rajput*, A. Nagle, H. Vishwakarma, and D. Papailiopoulos
Advances in Neural Information Processing Systems, 2020. NeurIPS'20
(spotlight)

† Alphabetical ordering as per the convention in theoretical computer science.

*Equal contribution.

- ▶ **Extracting Robust and Accurate Features via a Robust Information Bottleneck** **JSAIT**
 A. Pensia, V. Jog, and P. Loh
IEEE Journal on Selected Areas in Information Theory, 2020.
- ▶ **Deep Topic Models for Multi-Label Learning** **AISTATS'19**
 R. Panda*, A. Pensia*, N. Mehta, M. Zhou, and P. Rai
International Conference on Artificial Intelligence and Statistics, 2019.
- ▶ **Generalization Error Bounds for Noisy, Iterative Algorithms** **ISIT'18**
 A. Pensia, V. Jog, and P. Loh
IEEE International Symposium on Information Theory, 2018.

Academic Visits

- Simons Institute for the Theory of Computing Fall '21
 ▶ Participated in the “Computational Complexity of Statistical Inference” program
- Centre for Mathematical Sciences, University of Cambridge Spring '22

Professional Service

- Workshop co-organizer** at ICLR 2023: Pitfalls of limited data and computation for Trustworthy ML.
- Area chair:** ALT 2024
- Journal reviewer:** ◊ Annals of Statistics ◊ Annales de l’Institut Henri Poincaré ◊ IEEE Transactions on Information Theory ◊ Information and Inference: A journal of the IMA. ◊ Journal of the American Statistical Association (JASA) ◊ IEEE Journal on Selected Areas in Information Theory (JSAIT) ◊ Journal of Machine Learning Research (JMLR) ◊ Journal of the Royal Statistical Society: Series B (JRRSB) ◊ Machine Learning (Springer) ◊ SIAM Journal on Mathematics of Data Science (SIMODS) ◊ Statistics & Probability Letters
- Conference reviewer:** ◊ ALT 2021 ◊ COLT 2021, 2022, 2023, 2024 ◊ FOCS 2022, 2024 ◊ ICLR 2021 (outstanding reviewer), 2022 ◊ ICML 2020 (top 33% reviewer), 2021, 2022, 2023, 2024 (top reviewer) ◊ ISIT 2022, 2023 ◊ ITCS 2023 ◊ NeurIPS 2020, 2021, 2022 (top reviewer), 2023 ◊ STOC 2022

In total, I have completed 100+ reviews.

Mentored an undergraduate student as part of the **WISCERS** program in 2020-2021.

Mentor, Learning Theory Alliance Workshop at NeurIPS 2024

Invited Talks

- SoS Certifiability of Subgaussian Distributions and Its Algorithmic Applications
 - ▶ Simons Institute (Industry Day), UC Berkeley, 2024
 - ▶ TOCA-SV@Stanford (Graduating Bits), Stanford University, 2024
- Towards Practical Algorithms for Outlier-Robust Estimation
 - ▶ Meet the Fellows, Simons Institute, UC Berkeley, 2024
- Robust sparse estimation: An overview
 - ▶ Workshop on New Frontiers in Robust Statistics, TTIC, Chicago, 2024
- Non-asymptotic simple binary hypothesis testing
 - ▶ Conference on Learning Theory, Edmonton, 2024
 - ▶ IBM Math and TCS Council, 2023
- Simple binary hypothesis testing: Locally private and communication-efficient
 - ▶ IMS International Conference on Statistics and Data Science (ICSIDS), Portugal, 2023
 - ▶ Conference on Learning Theory, Bangalore, 2023
 - ▶ International Indian Statistical Association (IISA), Colorado, 2023

- Information Theory and Applications Workshop (ITA), San Diego, 2023
- Algorithms Seminar, Google, 2023

Hypothesis testing under communication constraints

- Conference on Information Sciences and Systems (CISS), 2022
- Institute for Foundations of Data Science (IFDS) Ideas Forum at UW-Madison, 2022

Statistical query lower bounds for list-decodable linear regression

- Reading group on “Statistical Query” at Simons Institute for the Theory of Computing, 2021

Robust estimation in high dimensions: Heavy tails and adversarial contamination

- Foundations of Data Science, Yale University, 2023
- ETH AI Center Symposium, 2022
- Cornell Young Researchers Workshop, 2021
- IFDS Workshop on Statistical Approaches to Understanding Modern ML Methods, 2021
- IFDS Seminar at UW-Madison, 2020

Estimating location parameters in entangled single-sample distributions

- IFDS Seminar at UW-Madison, 2019

Teaching Experience

Teaching assistant for ◊ Mathematical Foundations of Machine Learning ◊ Introduction to Bioinformatics ◊ Probabilistic Mobile Robotics ◊ Communication Systems