
Michael William Mahoney

Tel: 650-284-8472

Email: mmahoney@stat.berkeley.edu

Web: <http://www.stat.berkeley.edu/~mmahoney/>

Research Interests

- Algorithmic and statistical aspects of modern large-scale data analysis
- Design and analysis of algorithms for matrix, graph, and regression problems
- Statistical data analysis in large-scale scientific and Internet applications

Appointments and Research Experience

- | | |
|--------------|--|
| 2013-present | Associate Adjunct Professor, Department of Statistics
Faculty Member, AMP Lab, RISE Lab, Dept. of EECS
Senior Fellow, Berkeley Institute of Data Science
Director, TRIPODS FODA Institute (2017-)
University of California, Berkeley, CA, USA
Senior Research Scientist, International Computer Science Institute
External Collaborator, Lawrence Berkeley National Laboratory
Lead Data Scientist, View Labs, Inc. (Industry, 2013-2017) |
| 2008-2013 | Senior Research Scientist, Department of Mathematics
Stanford University, Stanford, CA, USA <ul style="list-style-type: none">• Design, analysis, and implementation of randomized algorithms for very large-scale matrix, graph, and regression problems• Theory and practice of implicit regularization and implicit optimization methods in scalable approximation algorithms• Statistical properties of scalable approximation algorithms• Applications to community detection, clustering, and information dynamics in large social and information networks• Applications to DNA microarray, single nucleotide polymorphism, medical imaging, and astronomical data |
| 2005-2008 | Senior Research Scientist
Yahoo Research, Sunnyvale, CA, USA <ul style="list-style-type: none">• Algorithms for matrix computations, graph partitioning, regression, and statistical data analysis• Applications to the analysis of large-scale scientific and Internet data |

- Large-scale web analytics, machine learning, and query log analysis
- 2002-2005 J. W. Gibbs Assistant Professor, Department of Mathematics
Research Affiliate, Department of Computer Science
Yale University, New Haven, CT, USA
- Randomized and approximation algorithms for large-scale matrix problems
 - Application of randomized and approximate algorithmic methods for structuring and understanding hyperspectral medical data sets
 - Application to DNA microarray and DNA single nucleotide polymorphism (SNP) data
- 2000-2002 Postdoctoral Research Associate
Columbia University, New York, NY, USA and
Yale University, New Haven, CT, USA
- Theoretical statistical mechanics, applications of complex network theory, and the analysis of algorithms of interest in chemical physics
- 1997-2000 Graduate Research Associate
Yale University, New Haven, CT, USA
- Research in computational statistical mechanics, computational chemical physics, and computational biophysics
 - Collaborated in the maintenance and administration of a PC-based computing cluster, installing and maintaining operating system, compiler, and queuing software for efficient use of computing resources
 - Extended the BOSS program to perform molecular dynamics, Monte Carlo calculations on polarizable molecular liquids, and quantum statistical mechanical path integral Monte Carlo calculations

Education

- 2000 Ph.D. Physics,
Yale University, New Haven, CT, USA
Dissertation: “The Computational Statistical Mechanics of Simple Models of Liquid Water”
- 1995 B.S. Mathematics,
University of Maryland, College Park, MD, USA

1994 B.S., Physics, Philosophy
Summa Cum Laude, Phi Beta Kappa, May 1994
University of Notre Dame, South Bend, IN, USA

Postdocs, Students and Interns, Advised or Co-advised

- N. Benjamin Erichson (postdoc, UCB, 2018-present)
- Zhewei Yao (UCB, Ph.D. expected 2021)
- Feynman Liang (UCB, Ph.D. expected 2021)
- Amir Gholaminejad (postdoc, UCB, 2017-present, joint w. K. Keutzer)
- Kai Rothauge (postdoc, UCB, 2017-present)
- Stefan Ivo Palombo (UCB, undergraduate thesis, 2017-2018)
- Shusen Wang (postdoc, UCB, 2016-2018; now at Stevens Institute)
- Peng Xu (Stanford, Ph.D. 2018, joint with C. Re; now at Amazon)
- Kimon Fountoulakis (postdoc, UCB, 2015-2018; now at Waterloo)
- Julian Shun (Miller postdoc, UCB 2015-2107; now at MIT)
- Fred Roosta (postdoc, UCB 2015-2107; now at Univ. of Queensland)
- Alex Gittens (postdoc, UCB 2015-2107; now at RPI)
- Liping Jing (visiting researcher, ICSI 2016, joint w. G. Friedland)
- Aditya Devarakonda (UCB, Ph.D. 2018, joint w. J. Demmel; now at APL)
- Di Wang (UCB, Ph.D. 2017, joint w. S. Rao; now at Georgia Tech)
- Simon Du (UCB, undergraduate thesis, 2015, joint w. M. Gu; now at CMU)
- Jiyan Yang (Stanford, Ph.D. 2016; now at Facebook)
- Aaron Adcock (Stanford, Ph.D. 2015; now at Facebook)
- Xiangrui Meng (Stanford, Ph.D. 2015; now at Databricks)
- Lorenzo Orecchia (Intern at Yahoo, 2008; then at UC Berkeley, now at BU)
- Hari Narayanan (Intern at Yahoo, 2007; then at Chicago, now at UW)
- Jure Leskovec (Intern at Yahoo, 2007; then at CMU, now at Stanford)
- Jeff Phillips (Intern at Yahoo, 2007; then at Duke, now at Utah)
- Lek-Heng Lim (Intern at Yahoo, 2006; then at Stanford, now at Chicago)
- Boulos Harb (Intern at Yahoo, 2006; then at Pennsylvania, now at Google)
- James Campbell (Yale, undergraduate thesis, 2005)

Publications

1. “HAWQ: Hessian AWare Quantization of Neural Networks with Mixed-Precision,” Z. Dong, Z. Yao, A. Gholami, M. W. Mahoney, K. Keutzer, *Technical Report, Preprint: arXiv:1905.03696* (2019).
2. “JumpReLU: A Retrofit Defense Strategy for Adversarial Attacks,” N. B. Erichson, Z. Yao, and M. W. Mahoney, *Technical Report, Preprint: arXiv:1904.03750* (2019).

3. “OverSketched Newton: Fast Convex Optimization for Serverless Systems,” V. Gupta, S. Kadhe, T. Courtade, M. W. Mahoney, and K. Ramchandran, *Technical Report, Preprint: arXiv:1903.08857* (2019).
4. “Inefficiency of K-FAC for Large Batch Size Training,” L. Ma, G. Montague, J. Ye, Z. Yao, A. Gholami, K. Keutzer, and M. W. Mahoney, *Technical Report, Preprint: arXiv:1903.06237* (2019).
5. “Sub-Sampled Newton Methods,” F. Roosta-Khorasani and M. W. Mahoney, *Mathematical Programming*, 174(1-2): 293-326 (2019).
6. “Shallow Learning for Fluid Flow Reconstruction with Limited Sensors and Limited Data,” N. B. Erichson, L. Mathelin, Z. Yao, S. L. Brunton, M. W. Mahoney, and J. N. Kutz, *Technical Report, Preprint: arXiv:1902.07358* (2019).
7. “Minimax experimental design: Bridging the gap between statistical and worst-case approaches to least squares regression,” M. Derezhinski, K. L. Clarkson, M. W. Mahoney, and M. K. Warmuth, *Technical Report, Preprint: arXiv:1902.00995* (2019); Accepted for publication, *Proc. COLT 2019*.
8. “Heavy-Tailed Universality Predicts Trends in Test Accuracies for Very Large Pre-Trained Deep Neural Networks,” C. H. Martin and M. W. Mahoney, *Technical Report, Preprint: arXiv:1901.08278* (2019).
9. “Traditional and Heavy-Tailed Self Regularization in Neural Network Models,” C. H. Martin and M. W. Mahoney, *Technical Report, Preprint: arXiv:1901.08276* (2019); Accepted for publication, *Proc. ICML 2019*.
10. “Trust Region Based Adversarial Attack on Neural Networks,” Z. Yao, A. Gholami, P. Xu, K. Keutzer, and M. W. Mahoney, *Technical Report, Preprint: arXiv:1812.06371* (2018); Accepted for publication, *Proc. CVPR 2019*.
11. “Parameter Re-Initialization through Cyclical Batch Size Schedules,” N. Mu, Z. Yao, A. Gholami, K. Keutzer, and M. W. Mahoney, *Technical Report, Preprint: arXiv:1812.01216* (2018); Presented in the *Systems for Machine Learning Workshop at the 2018 NeurIPS Conference*.
12. “On the Computational Inefficiency of Large Batch Sizes for Stochastic Gradient Descent,” N. Golmant, N. Vemuri, Z. Yao, V. Feinberg, A. Gholami, K. Rothauge, M. W. Mahoney, and J. Gonzalez, *Technical Report, Preprint: arXiv:1811.12941* (2018).
13. *The Mathematics of Data*, M. W. Mahoney, J. C. Duchi, and A. C. Gilbert, Eds. AMS, IAS/PCMI, and SIAM (2018).

14. "A Short Introduction to Local Graph Clustering Methods and Software," K. Fountoulakis, D. F. Gleich, and M. W. Mahoney, *Technical Report, Preprint: arXiv:1810.07324* (2018); *Absts. of the 7th Intl. Conference on Complex Networks and Their Applications*.
15. "Implicit Self-Regularization in Deep Neural Networks: Evidence from Random Matrix Theory and Implications for Learning," C. H. Martin and M. W. Mahoney, *Technical Report, Preprint: arXiv:1810.01075* (2018); Journal version submitted for publication.
16. "Large batch size training of neural networks with adversarial training and second-order information," Z. Yao, A. Gholami, K. Keutzer, and M. W. Mahoney, *Technical Report, Preprint: arXiv:1810.01021* (2018).
17. "Newton-MR: Newton's Method Without Smoothness or Convexity," F. Roosta, Y. Liu, P. Xu, and M. W. Mahoney, *Technical Report, Preprint: arXiv:1810.00303* (2018).
18. "Distributed Second-order Convex Optimization," C.-H. Fang, S. B. Kylasa, F. Roosta-Khorasani, M. W. Mahoney, and A. Grama, *Technical Report, Preprint: arXiv:1807.07132* (2018).
19. "Alchemist: An Apache Spark \Leftrightarrow MPI Interface," A. Gittens, K. Rothauge, M. W. Mahoney, S. Wang, L. Gerhardt, Prabhat, J. Kottalam, M. Ringenborg, and K. Maschhoff, *Technical Report, Preprint: arXiv:1806.01270* (2018); *Concurrency and Computation: Practice and Experience (Special Issue of the Cray User Group, CUG 2018)*, e5026 (2018).
20. "Accelerating Large-Scale Data Analysis by Offloading to High-Performance Computing Libraries using Alchemist," A. Gittens, K. Rothauge, S. Wang, M. W. Mahoney, L. Gerhardt, Prabhat, J. Kottalam, M. Ringenborg, and K. Maschhoff, *Technical Report, Preprint: arXiv:1805.11800* (2018); *Proc. of the 24th Annual SIGKDD*, 293-301 (2018).
21. "Group Collaborative Representation for Image Set Classification," B. Liu, L. Jing, J. Li, J. Yu, A. Gittens, and M. W. Mahoney, *International Journal of Computer Vision*, 1-26 (2018).
22. "Error Estimation for Randomized Least-Squares Algorithms via the Bootstrap," M. E. Lopes, S. Wang, and M. W. Mahoney, *Technical Report, Preprint: arXiv:1803.08021* (2018); *Proc. of the 35th ICML Conference* 3223-3232 (2018); Journal version submitted for publication.
23. "GPU Accelerated Sub-Sampled Newton's Method," S. B. Kylasa, F. Roosta-Khorasani, M. W. Mahoney, and A. Grama, *Technical Report, Preprint: arXiv:1802.09113* (2018); *Proc. 2019 SDM*, 702-710 (2019).

24. “Hessian-based Analysis of Large Batch Training and Robustness to Adversaries,” Z. Yao, A. Gholami, Q. Lei, K. Keutzer, and M. W. Mahoney, *Technical Report, Preprint: arXiv:1802.08241* (2018); *Proc. of the 2018 NeurIPS Conference*, 4954-4964 (2018).
25. “Inexact Non-Convex Newton-Type Methods,” Z. Yao, P. Xu, F. Roosta-Khorasani, and M. W. Mahoney, *Technical Report, Preprint: arXiv:1802.06925* (2018); Journal version submitted for publication.
26. “Out-of-sample extension of graph adjacency spectral embedding,” K. Levin, F. Roosta-Khorasani, M. W. Mahoney, and C. E. Priebe, *Technical Report, Preprint: arXiv:1802.06307* (2018); *Proc. of the 35th ICML Conference* 2981-2990 (2018); Journal version submitted for publication.
27. “Lectures on Randomized Numerical Linear Algebra,” P. Drineas and M. W. Mahoney, *Technical Report, Preprint: arXiv:1712.08880* (2017); In: *Lectures of the 2016 PCMI Summer School on Mathematics of Data*.
28. “Avoiding Synchronization in First-Order Methods for Sparse Convex Optimization,” A. Devarakonda, K. Fountoulakis, J. Demmel, and M. W. Mahoney, *Technical Report, Preprint: arXiv:1712.06047* (2017); *Proc. of the IPDPS 2018 Conference*, 409-418 (2018).
29. “Rethinking generalization requires revisiting old ideas: statistical mechanics approaches and complex learning behavior,” C. H. Martin and M. W. Mahoney, *Technical Report, Preprint: arXiv:1710.09553* (2017); Journal version submitted for publication.
30. “LASAGNE: Locality And Structure Aware Graph Node Embedding,” E. Faerman, F. Borutta, K. Fountoulakis, and M. W. Mahoney, *Technical Report, Preprint: arXiv:1710.06520* (2017); *Proc. 2018 International Conference on Web Intelligence*, 246-253 (2018). (Awarded the Best Student Paper Award.)
31. “A Berkeley View of Systems Challenges for AI,” I. Stoica, D. Song, R. A. Popa, D. A. Patterson, M. W. Mahoney, R. H. Katz, A. D. Joseph, M. Jordan, J. M. Hellerstein, J. Gonzalez, K. Goldberg, A. Ghodsi, D. E. Culler, and P. Abbeel, *Technical Report No. UCB/EECS-2017-159*, October 2017.
32. “GIANT: Globally Improved Approximate Newton Method for Distributed Optimization,” S. Wang, F. Roosta-Khorasani, P. Xu, and M. W. Mahoney, *Technical Report, Preprint: arXiv:1709.03528* (2017); *Proc. of the 2018 NeurIPS Conference*, 2338-2348 (2018).

33. "Second-Order Optimization for Non-Convex Machine Learning: An Empirical Study," P. Xu, F. Roosta-Khorasani, and M. W. Mahoney, *Technical Report, Preprint: arXiv:1708.07827* (2017).
34. "Newton-Type Methods for Non-Convex Optimization Under Inexact Hessian Information," P. Xu, F. Roosta-Khorasani, and M. W. Mahoney, *Technical Report, Preprint: arXiv:1708.07164* (2017); Accepted for publication, *Mathematical Programming*.
35. "A Bootstrap Method for Error Estimation in Randomized Matrix Multiplication," M. E. Lopes, S. Wang, and M. W. Mahoney, *Technical Report, Preprint: arXiv:1708.01945* (2017); *J. Machine Learning Research*, 20(39): 1–40 (2019).
36. "Capacity releasing diffusions for speed and locality," D. Wang, K. Fountoulakis, M. Henzinger, M. W. Mahoney, and S. Rao, *Technical Report, Preprint: arXiv:1706.05826* (2017); *Proc. of the 34th ICML Conference* 3598-3607 (2017).
37. "Scalable Kernel K-Means Clustering with Nystrom Approximation: Relative-Error Bounds," S. Wang, A. Gittens, and M. W. Mahoney, *Technical Report, Preprint: arXiv:1706.02803* (2017); *J. Machine Learning Research*, 20(12): 1-49 (2019).
38. "Union of Intersections (UoI) for Interpretable Data Driven Discovery and Prediction," K. E. Bouchard, A. F. Bujan, F. Roosta-Khorasani, S. Ubaru, Prabhat, A. M. Snijders, J.-H. Mao, E. F. Chang, M. W. Mahoney, S. Bhattacharyya, *Technical Report, Preprint: arXiv:1705.07585* (2017); *Proc. of the 2017 NIPS Conference*, 1078-1086 (2017).
39. "Skip-Gram - Zipf + Uniform = Vector Additivity," A. Gittens, D. Achlioptas, and M. W. Mahoney, *Proc. of the 55th ACL Meeting* 69-76 (2017).
40. "Principles and Applications of Science of Information [Scanning the Issue]," T. Courtade, A. Grama, M. W. Mahoney, and T. Weissman, *Proceedings of the IEEE*, 105(2): 183-188 (2017).
41. "Social Discrete Choice Models," D. Zhang, K. Fountoulakis, J. Cao, M. Yin, M. W. Mahoney, and A. Pozdnoukhov, *Technical Report, Preprint: arXiv:1703.07520* (2017).
42. "Sketched Ridge Regression: Optimization Perspective, Statistical Perspective, and Model Averaging," S. Wang, A. Gittens, and M. W. Mahoney, *Technical Report, Preprint: arXiv:1702.04837* (2017); *Proc. of the 34th ICML Conference* 3608-3616 (2017); *J. Machine Learning Research*, 18(218): 1-50 (2018).
43. "Avoiding communication in primal and dual block coordinate descent methods," A. Devarakonda, K. Fountoulakis, J. Demmel, and M. W. Mahoney, *Technical*

- Report, Preprint: arXiv:1612.04003 (2016); SIAM J. Scientific Computing, 41(1), C1-C27 (2019).*
44. "Feature-distributed sparse regression: a screen-and-clean approach," J. Yang, M. W. Mahoney, M. A. Saunders, and Y. Sun, *Proc. of the 2016 NIPS Conference*, 2711-2719 (2016).
 45. "Multi-label learning with semantic embeddings," L. Jing, M. Cheng, L. Yang, A. Gittens, M. W. Mahoney, *ICLR 2017 OpenReview.net*.
 46. "Mapping the Similarities of Spectra: Global and Locally-biased Approaches to SDSS Galaxy Data," D. Lawlor, T. Budavari, and M. W. Mahoney, *Technical Report, Preprint: arXiv:1609.03932 (2016); The Astrophysical Journal, 833:1, 26 (2016).*
 47. "Lecture Notes on Spectral Graph Methods," M. W. Mahoney, *Technical Report, Preprint: arXiv:1608.04845 (2016).*
 48. "Lecture Notes on Randomized Linear Algebra," M. W. Mahoney, *Technical Report, Preprint: arXiv:1608.04481 (2016).*
 49. "An optimization approach to locally-biased graph algorithms," K. Fountoulakis, D. F. Gleich, M. W. Mahoney, *Technical Report, Preprint: arXiv:1607.04940 (2016); Proceedings of the IEEE, 105(2): 256-272 (2017).*
 50. "DCAR: A Discriminative and Compact Audio Representation to Improve Event Detection," L. Jing, B. Liu, J. Choi, A. Janin, J. Bernd, M. W. Mahoney, and G. Friedland, *Technical Report, Preprint: arXiv:1607.04378 (2016); Proc. of the 2016 ACM Multimedia Conference 57-61 (2016); IEEE Transactions on Multimedia, 19(12): 2637-2650 (2017).*
 51. "Matrix Factorization at Scale: a Comparison of Scientific Data Analytics in Spark and C+MPI Using Three Case Studies," A. Gittens, A. Devarakonda, E. Racah, M. Ringenburt, L. Gerhardt, J. Kottaalam, J. Liu, K. Maschhoff, S. Canon, J. Chhugani, P. Sharma, J. Yang, J. Demmel, J. Harrell, V. Krishnamurthy, M. W. Mahoney, and Prabhat, *Technical Report, Preprint: arXiv:1607.01335 (2016); Proc. 2016 IEEE BigData, 204-213 (2016).*
 52. "Sub-sampled Newton Methods with Non-uniform Sampling," P. Xu, J. Yang, F. Roosta-Khorasani, C. Re, and M. W. Mahoney, *Technical Report, Preprint: arXiv:1607.00559 (2016); Proc. of the 2016 NIPS Conference, 3000-3008 (2016).*
 53. "Approximating the Solution to Mixed Packing and Covering LPs in $O(1/\epsilon^3)$ parallel time," M. W. Mahoney, S. Rao, D. Wang, and P. Zhang, *Proc. of the 43rd ICALP Conference, 52:1-52:14 (2016).*

54. "A Simple and Strongly-Local Flow-Based Method for Cut Improvement," N. Veldt, D. F. Gleich, and M. W. Mahoney, *Technical Report, Preprint*: arXiv:1605.08490 (2016); *Proc. of the 33rd ICML Conference 1938-1947* (2016).
55. "RandNLA: Randomized Numerical Linear Algebra," P. Drineas and M. W. Mahoney, *Communications of the ACM*, 59: 80-90 (2016).
56. "FLAG n' FLARE: Fast Linearly-Coupled Adaptive Gradient Methods," X. Cheng, F. Roosta-Khorasani, P. L. Bartlett, and M. W. Mahoney, *Technical Report, Preprint*: arXiv:1605.08108 (2016); *Proc. of the 21st AISTATS Conference*, PMLR 84:404-414 (2018).
57. "Parallel Local Graph Clustering," J. Shun, F. Roosta-Khorasani, K. Fountoulakis, and M. W. Mahoney, *Technical Report, Preprint*: arXiv:1604.07515 (2016); *Proceedings of the VLDB Endowment*, 9(12) 1041-1052 (2016).
58. "A multi-platform evaluation of the randomized CX low-rank matrix factorization in Spark," A. Gittens, J. Kottalam, J. Yang, M. F. Ringenburt, J. Chhugani, E. Racah, M. Singh, Y. Yao, C. Fischer, O. Ruebel, B. Bowen, N. G. Lewis, M. W. Mahoney, V. Krishnamurthy, and Prabhat, *Proc. 5th International Workshop on Parallel and Distributed Computing for Large Scale Machine Learning and Big Data Analytics*, at IPDPS, 2016.
59. "Mining Large Graphs," D. F. Gleich and M. W. Mahoney, In *Handbook of Big Data*. pp. 191-220, edited by P. Buhlmann, P. Drineas, M. Kane, and M. van de Laan, Chapman and Hall/CRC Press, 2016.
60. "Structural properties underlying high-quality Randomized Numerical Linear Algebra algorithms," M. W. Mahoney and P. Drineas, In *Handbook of Big Data*. pp. 137-154, edited by P. Buhlmann, P. Drineas, M. Kane, and M. van de Laan, Chapman and Hall/CRC Press, 2016.
61. "Variational Perspective on Local Graph Clustering," K. Fountoulakis, X. Cheng, J. Shun, F. Roosta-Khorasani and M. W. Mahoney, *Technical Report, Preprint*: arXiv:1602.01886 (2016); *Mathematical Programming*, 174(1-2): 553-573 (2019).
62. "Sub-Sampled Newton Methods II: Local Convergence Rates," F. Roosta-Khorasani and M. W. Mahoney, *Technical Report, Preprint*: arXiv:1601.04738 (2016); Accepted for publication, *Mathematical Programming*, with SSN I.
63. "Sub-Sampled Newton Methods I: Globally Convergent Algorithms," F. Roosta-Khorasani and M. W. Mahoney, *Technical Report, Preprint*: arXiv:1601.04737 (2016); Accepted for publication, *Mathematical Programming*, with SSN II.

64. "RandNLA, Pythons, and the CUR for Your Data Problems: Reporting from G2S3 2015 in Delphi," E. Gallopoulos, P. Drineas, I. Ipsen, and M. W. Mahoney, *SIAM News* 49:1 January/February 2016.
65. "Faster Parallel Solver for Positive Linear Programs via Dynamically-Bucketed Selective Coordinate Descent," D. Wang, M. W. Mahoney, N. Mohan, and S. Rao, *Technical Report, Preprint*: arXiv:1511.06468 (2015).
66. "A Local Perspective on Community Structure in Multilayer Networks," L. G. S. Jeub, M. W. Mahoney, P. J. Mucha, and M. A. Porter, *Technical Report, Preprint*: arXiv:1510.05185 (2015); *Network Science*, 5(2): 144-163 (2017).
67. "Optimal Subsampling Approaches for Large Sample Linear Regression," R. Zhu, P. Ma, M. W. Mahoney, and B. Yu, *Technical Report, Preprint*: arXiv:1509.05111 (2015).
68. "Unified Acceleration Method for Packing and Covering Problems via Diameter Reduction," D. Wang, S. Rao, and M. W. Mahoney, *Technical Report, Preprint*: arXiv:1508.02439 (2015); *Proc. of the 43rd ICALP Conference*, 50:1-50:13 (2016).
69. "Using local spectral methods to robustify graph-based learning algorithms," D. F. Gleich and M. W. Mahoney, *Proc. of the 21st Annual SIGKDD*, 359-368 (2015).
70. "Structured Block Basis Factorization for Scalable Kernel Matrix Evaluation," R. Wang, Y. Li, M. W. Mahoney, and E. Darve, *Technical Report, Preprint*: arXiv:1502.03571 (2015); Journal version submitted for publication.
71. "Identifying important ions and positions in mass spectrometry imaging data using CUR matrix decompositions," J. Yang, O. Rubel, Prabhat, M. W. Mahoney, and B. P. Bowen, *Analytical Chemistry*, 87 (9), 4658-4666 (2015).
72. "Spectral Gap Error Bounds for Improving CUR Matrix Decomposition and the Nystrom Method," D. G. Anderson, S. S. Du, M. W. Mahoney, C. Melgaard, K. Wu, and M. Gu, *Proc. of the 18th International Conference on AISTATS*, 38: 19-27 (2015).
73. "Weighted SGD for Lp Regression with Randomized Preconditioning," J. Yang, Y.-L. Chow, C. Re, and M. W. Mahoney, *Technical Report, Preprint*: arXiv:1502.03571 (2015); *Proc. of the 27-th Annual SODA*, 558-569 (2016); *J. Machine Learning Research*, 18(211): 1-43 (2018).
74. "Implementing Randomized Matrix Algorithms in Parallel and Distributed Environments," J. Yang, X. Meng, and M. W. Mahoney, *Technical Report*,

- Preprint: arXiv:1502.03032 (2015); Proceedings of the IEEE 104(1): 58-92 (2016).*
75. "Tree decompositions and social graphs," A. B. Adcock, B. D. Sullivan, and M. W. Mahoney, *Technical Report, Preprint: arXiv:1411.1546 (2014); Internet Mathematics*, 12(5), 315-361 (2016).
 76. "Fast Randomized Kernel Methods With Statistical Guarantees," A. El Alaoui and M. W. Mahoney, *Technical Report, Preprint: arXiv:1411.0306 (2014); Proc. of the 2015 NIPS Conference*, 775-783 (2015).
 77. "A Statistical Perspective on Randomized Sketching for Ordinary Least-Squares," G. Raskutti and M. W. Mahoney, *Technical Report, Preprint: arXiv:1406.5986 (2014); Proc. of the 32nd ICML Conference*, 617-625 (2015); *J. Machine Learning Research*, 17(214): 1-31 (2016).
 78. "Random Laplace Feature Maps for Semigroup Kernels on Histograms," J. Yang, V. Sindhwani, Q. Fan, H. Avron, and M. W. Mahoney, *Proc. of the 27th CVPR Conference*, 971-978 (2014).
 79. "Anti-differentiating Approximation Algorithms: A case study with Min-cuts, Spectral, and Flow," D. F. Gleich and M. W. Mahoney, *Proc. of the 31st ICML Conference, JMLR W&CP 32 (1): 1018-1025 (2014).*
 80. "Quasi-Monte Carlo Feature Maps for Shift-Invariant Kernels," J. Yang, V. Sindhwani, H. Avron, and M. W. Mahoney, *Proc. of the 31st ICML Conference, JMLR W&CP 32 (1): 485-493 (2014); Technical Report, Preprint: arXiv:1412.8293 (2014); J. Machine Learning Research*, 17(120): 1-38 (2016).
 81. "Think Locally, Act Locally: The Detection of Small, Medium-Sized, and Large Communities in Large Networks," L. G. S. Jeub, P. Balachandran, M. A. Porter, P. J. Mucha, and M. W. Mahoney, *Technical Report, Preprint: arXiv:1403.3795 (2014); Physical Review E*, 91, 012821 (2015).
 82. "A new spin on an old algorithm: technical perspective on Communication costs of Strassen's matrix multiplication," M. W. Mahoney, *Communications of the ACM*, 57(2): 106 (2014).
 83. "Tree-like Structure in Large Social and Information Networks," A. B. Adcock, B. D. Sullivan, and M. W. Mahoney, *Proc. of the 2013 IEEE ICDM*, 1-10 (2013).
 84. "Objective Identification of Informative Wavelength Regions in Galaxy Spectra," C.-W. Yip, M. W. Mahoney, A. S. Szalay, I. Csabai, T. Budavari, R. F. G. Wyse, and L. Dobos, *Technical Report, Preprint: arXiv:1312.0637 (2013); Astronomical Journal*, 147, 110 (2014).

85. "Evaluating OpenMP Tasking at Scale for the Computation of Graph Hyperbolicity," A. B. Adcock, B. D. Sullivan, O. R. Hernandez, and M. W. Mahoney, *Proc. of the 9-th IWOMP*, 71-83 (2013).
86. "Frontiers in Massive Data Analysis," Committee on the Analysis of Massive Data, et al. (M. I. Jordan, et al.), The National Academies Press (2013).
87. "A Statistical Perspective on Algorithmic Leveraging," P. Ma, M. W. Mahoney, and B. Yu, *Technical Report, Preprint: arXiv:1306.5362* (2013); *Proc. of the 31st ICML Conference, JMLR W&CP* 32 (1): 91-99 (2014); *J. Machine Learning Research*, 16, 861-911 (2015).
88. "Robust Regression on MapReduce," X. Meng, and M. W. Mahoney, *Proc. of the 30th ICML Conference, JMLR W&CP* 28(3): 888-896 (2013).
89. "Quantile Regression for Large-scale Applications," J. Yang, X. Meng, and M. W. Mahoney, *Technical Report, Preprint: arXiv:1305.0087* (2013); *Proc. of the 30th ICML Conference, JMLR W&CP* 28(3): 881-887 (2013); *SIAM J. Scientific Computing*, 36(5), S78-S110 (2014).
90. "Revisiting the Nystrom Method for Improved Large-Scale Machine Learning," A. Gittens and M. W. Mahoney, *Technical Report, Preprint: arXiv:1303.1849* (2013); *Proc. of the 30th ICML Conference, JMLR W&CP* 28(3): 567-575 (2013); *J. Machine Learning Research*, 17(117): 1-65 (2016).
91. "Semi-supervised Eigenvectors for Large-scale Locally-biased Learning," T. J. Hansen and M. W. Mahoney, *Proc. of the 2012 NIPS Conference*, 2528-2536 (2012); *Technical Report, Preprint: arXiv:1304.7528* (2013); *J. Machine Learning Research*, 15, 3691-3734 (2014).
92. "Low-distortion Subspace Embeddings in Input-sparsity Time and Applications to Robust Linear Regression," X. Meng and M. W. Mahoney, *Technical Report, Preprint: arXiv:1210.3135* (2012); *Proc. of the 45-th STOC*, 91-100 (2013).
93. "The Fast Cauchy Transform: with Applications to Basis Construction, Regression, and Subspace Approximation in L1," K. L. Clarkson, P. Drineas, M. Magdon-Ismail, M. W. Mahoney, X. Meng, D. P. Woodruff, *Technical Report, Preprint: arXiv:1207.4684* (2012); *Proc. of the 24-th Annual SODA*, 466-477 (2013); *SIAM J. Computing*, 45, 763-810 (2016).
94. "rCUR: an R package for CUR matrix decomposition," A. Bodor, I. Csabai, M. W. Mahoney, and N. Solymosi, *BMC Bioinformatics*, 13:103, 2012.
95. "Approximate Computation and Implicit Regularization for Very Large-scale Data Analysis," M. W. Mahoney, *Technical Report, Preprint: arXiv:1203.0786* (2012), *Proc. of the 2012 ACM Symposium on PODS*, 143-154, 2012.

96. “On the Hyperbolicity of Small-World Networks and Tree-Like Graphs,” W. Chen, W. Fang, G. Hu, M. W. Mahoney, *Technical Report, Preprint*: arXiv:1201.1717 (2012); *Proc. of the 23rd ISAAC* 278-288 (2012); *Internet Mathematics*, 9(4), 434-491 (2013).
97. “Randomized Dimensionality Reduction for K-means Clustering,” C. Boutsidis, A. Zouzias, M. W. Mahoney, and P. Drineas, *Technical Report, Preprint*: arXiv:1110.2897 (2011); *IEEE Transactions on Information Theory*, 61(2), 1045-1062 (2015).
98. “Regularized Laplacian Estimation and Fast Eigenvector Approximation,” P. O. Perry and M. W. Mahoney, *Technical Report, Preprint*: arXiv:1110.1757 (2011); *Proceedings of the 2011 NIPS Conference*, 2420-2428 (2011).
99. “LSRN: A Parallel Iterative Solver for Strongly Over- or Under-Determined Systems,” X. Meng, M. A. Saunders, and M. W. Mahoney, *Technical Report, Preprint*: arXiv:1109.5981 (2011); *SIAM J. Scientific Computing*, 36(2), C95-C118 (2014).
100. “Fast approximation of matrix coherence and statistical leverage,” P. Drineas, M. Magdon-Ismail, M. W. Mahoney, and D. P. Woodruff, *Technical Report, Preprint*: arXiv:1109.3843 (2011); *Proceedings of the 29th ICML Conference*, 1051-1058 (2012); *J. Machine Learning Research*, 13, 3475-3506 (2012).
101. “Localization on low-order eigenvectors of data matrices,” M. Cucuringu and M. W. Mahoney, *Technical Report, Preprint*: arXiv:1109.1355 (2011).
102. “Efficient Genomewide Selection of PCA-Correlated tSNPs for Genotype Imputation,” A. Javed, P. Drineas, M. W. Mahoney, and P. Paschou, *Annals of Human Genetics*, 75, 707-722 (2011).
103. “Randomized Algorithms for Matrices and Data,” M. W. Mahoney, *Technical Report, Preprint*: arXiv:1104.5557 (2011); abridged version as a chapter in: *Advances in Machine Learning and Data Mining for Astronomy*, M. J. Way, et al., editors, CRC Press, Taylor and Francis Group, pp. 647-672, 2012; full version as a monograph in: *Foundations and Trends in Machine Learning*, NOW Publishers, Volume 3, Issue 2, 2011.
104. “Computation in Large-Scale Scientific and Internet Data Applications is a Focus of MMDS 2010,” M. W. Mahoney, *Technical Report, Preprint*: arXiv:1012.4231 (2010) (arXiv), appeared in *SIGKDD Explorations*, *SIGACT News*, *ASA-SCGN Newsletter*, and *IMS Bulletin*.

105. "CUR from a Sparse Optimization Viewpoint," J. Bien, Y. Xu, and M. W. Mahoney, *Technical Report, Preprint: arXiv:1011.0413* (2010); *Proceedings of the 2010 NIPS Conference*, 217-225 (2010).
106. "Algorithmic and Statistical Perspectives on Large-Scale Data Analysis," M. W. Mahoney, *Technical Report, Preprint: arXiv:1010.1609* (2010); In: *Combinatorial Scientific Computing*, pp. 427-469, edited by U. Naumann and O. Schenk, CRC Press, Taylor and Francis Group (2012).
107. "Implementing Regularization Implicitly via Approximate Eigenvector Computation," M. W. Mahoney and L. Orecchia, *Technical Report, Preprint: arXiv:1010.0703* (2010); *Proceedings of the 28th ICML Conference*, 121-128 (2011).
108. "Approximating Higher-Order Distances Using Random Projections," P. Li, M. W. Mahoney, and Y. She, *Proc. of the 26th UAI Conference*, 312-321 (2010); *Technical Report, Preprint: arXiv:1203.3492* (2012).
109. "Effective Resistances, Statistical Leverage, and Applications to Linear Equation Solving," P. Drineas and M. W. Mahoney, *Technical Report, Preprint: arXiv:1005.3097* (2010).
110. "Empirical Comparison of Algorithms for Network Community Detection," J. Leskovec, K. J. Lang, and M. W. Mahoney, *Technical Report, Preprint: arXiv:1004.3539* (2010); *Proc. 19-th International WWW*, 631-640 (2010).
111. "A Local Spectral Method for Graphs: with Applications to Improving Graph Partitions and Exploring Data Graphs Locally," M. W. Mahoney, L. Orecchia, and N. K. Vishnoi, *Technical Report, Preprint: arXiv: 0912.0681* (2009); *J. Machine Learning Research*, 13, 2339-2365 (2012).
112. "Unsupervised Feature Selection for the k-means Clustering Problem," C. Boutsidis, M. W. Mahoney, and P. Drineas, *Proceedings of the 2009 NIPS Conference*, 153-161 (2009).
113. "Learning with Spectral Kernels and Heavy-Tailed Data," M. W. Mahoney and H. Narayanan, *Technical Report, Preprint: arXiv:0906.4539* (2009).
114. "Empirical Evaluation of Graph Partitioning using Spectral Embeddings and Flow," K. J. Lang, M. W. Mahoney, and L. Orecchia, *Proc. 8th International SEA*, 197-208 (2009).
115. "CUR Matrix Decompositions for Improved Data Analysis," M. W. Mahoney and P. Drineas, *Proc. Natl. Acad. Sci. USA*, 106, 697-702 (2009).

116. “An Improved Approximation Algorithm for the Column Subset Selection Problem,” C. Boutsidis, M. W. Mahoney, and P. Drineas, *Technical Report, Preprint: arXiv: 0812.4293* (2008); *Proc. 20th Annual SODA*, 968-977 (2009).
117. “Algorithmic and Statistical Challenges in Modern Large-Scale Data Analysis are the Focus of MMDS 2008,” M. W. Mahoney, L.-H. Lim, and G. E. Carlsson, *Technical Report, Preprint: arXiv: 0812.3702* (2008); Appeared in *SIGKDD Explorations, SIAM News*, and *ASA-SCGN Newsletter*.
118. “Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters,” J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney, *Technical Report, Preprint: arXiv: 0810.1355* (2008); *Internet Mathematics*, 6(1), 29-123 (2009).
119. “Unsupervised Feature Selection for Principal Components Analysis,” C. Boutsidis, M. W. Mahoney, and P. Drineas, *Proc. 14th Annual SIGKDD*, 61-69 (2008).
120. “Statistical Properties of Community Structure in Large Social and Information Networks,” J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney, *Proc. 17th International WWW*, 695-704 (2008).
121. “Faster Least Squares Approximation,” P. Drineas, M. W. Mahoney, S. Muthukrishnan, and T. Sarlos, *Technical Report, Preprint: arXiv:0710.1435* (2007); *Numerische Mathematik*, 117, 219-249 (2011).
122. “PCA-Correlated SNPs for Structure Identification in Worldwide Human Populations,” P. Paschou, E. Ziv, E. G. Burchard, S. Choudhry, W. Rodriguez-Cintron, M. W. Mahoney, and P. Drineas, *PLOS Genetics*, 3, 1672-1686 (2007).
123. “Relative-Error CUR Matrix Decompositions,” P. Drineas, M. W. Mahoney, and S. Muthukrishnan, *Technical Report, Preprint: arXiv:0708.3696* (2007); *SIAM J. Matrix Analysis and Applications*, 30, 844-881 (2008).
124. “Feature Selection Methods for Text Classification,” A. Dasgupta, P. Drineas, B. Harb, V. Josifovski, and M. W. Mahoney, *Proc. 13th Annual SIGKDD*, 230-239 (2007).
125. “Sampling Algorithms and Coresets for Lp Regression,” A. Dasgupta, P. Drineas, B. Harb, R. Kumar, and M. W. Mahoney, *Technical Report, Preprint: arXiv:0707.1714* (2007); *Proc. 19th Annual SODA*, 932-941 (2008); *SIAM J. Computing*, 38, 2060-2078 (2009).

126. "Web Information Retrieval and Linear Algebra Algorithms," A. Frommer, M. W. Mahoney, and D. B. Szyld (Eds.), *Proc. of Dagstuhl Seminar 07071*, (2007).
127. "Intra- and Interpopulation Genotype Reconstruction from Tagging SNPs," P. Paschou, M. W. Mahoney, A. Javed, J. R. Kidd, A. J. Pakstis, S. Gu, K. K. Kidd, and P. Drineas, *Genome Research*, 17(1), 96-107 (2007).
128. "Bridging the Gap Between Numerical Linear Algebra, Theoretical Computer Science, and Data Applications," G. H. Golub, M. W. Mahoney, P. Drineas, and L.-H. Lim, *SIAM News* 39:8 October 2006.
129. "Randomized Algorithms for Matrices and Massive Data Sets," P. Drineas and M. W. Mahoney, *Proc. 32nd Annual VLDB*, 1269 (2006).
130. "Subspace Sampling and Relative-Error Matrix Approximation: Column-Row-Based Methods," P. Drineas, M. W. Mahoney, and S. Muthukrishnan, *Proc. 14th Annual ESA*, 304-314 (2006).
131. "Subspace Sampling and Relative-Error Matrix Approximation: Column-Based Methods," P. Drineas, M. W. Mahoney, and S. Muthukrishnan, *Proc. 10th Annual RANDOM*, 316-326 (2006).
132. "Tensor-CUR Decompositions For Tensor-Based Data," M. W. Mahoney, M. Maggioni, and P. Drineas, *Proc. 12th Annual SIGKDD*, 327-336 (2006); *SIAM J. Matrix Analysis and Applications*, 30, 957-987 (2008).
133. "Polynomial Time Algorithm for Column-Row-Based Relative-Error Low-Rank Matrix Approximation," P. Drineas, M. W. Mahoney, and S. Muthukrishnan, *Technical Report*, DIMACS TR 2006-04 March 2006.
134. "Sampling Algorithms for L2 Regression and Applications," P. Drineas, M. W. Mahoney, and S. Muthukrishnan, *Proc. 17th Annual SODA*, 1127-1136 (2006).
135. "A Randomized Algorithm for a Tensor-Based Generalization of the Singular Value Decomposition," P. Drineas and M. W. Mahoney, *Technical Report*, YALEU/DCS/TR-1327, June 2005; *Linear Algebra and its Applications*, 420, 553-571 (2007).
136. "On the Nystrom Method for Approximating a Gram Matrix for Improved Kernel-Based Learning," P. Drineas and M. W. Mahoney, *Technical Report*, YALEU/DCS/TR-1319, April 2005; *Proc. 18th Annual COLT*, 323-337 (2005); *J. Machine Learning Research*, 6, 2153-2175 (2005).

137. "Sampling Sub-problems of Heterogeneous Max-Cut Problems and Approximation Algorithms," P. Drineas, R. Kannan, and M. W. Mahoney, *Technical Report*, YALEU/DCS/TR-1283, April 2004; *Proc. 22nd Annual STACS*, 57-68 (2005); *Random Structures and Algorithms*, 32:3, 307-333 (2008).
138. "Fast Monte Carlo Algorithms for Matrix Operations III: Computing an Efficient Approximate Decomposition of a Matrix," P. Drineas, R. Kannan, and M. W. Mahoney, *Technical Report*, YALEU/DCS/TR-1271, February 2004; *SIAM J. Computing*, 36, 184-206 (2006).
139. "Fast Monte Carlo Algorithms for Matrix Operations II: Computing Low-Rank Approximations to a Matrix," P. Drineas, R. Kannan, and M. W. Mahoney, *Technical Report*, YALEU/DCS/TR-1270, February 2004; *SIAM J. Computing*, 36, 158-183 (2006).
140. "Fast Monte Carlo Algorithms for Matrix Operations I: Approximating Matrix Multiplication," P. Drineas, R. Kannan, and M. W. Mahoney, *Technical Report*, YALEU/DCS/TR-1269, February 2004; *SIAM J. Computing*, 36, 132-157 (2006).
141. "Rapid Mixing of Several Markov Chains for a Hard-Core Model," R. Kannan, M. W. Mahoney, and R. Montenegro, *Proc. 14th Annual I.S.A.A.C.*, 663-675 (2003).
142. "Quantum, Intramolecular Flexibility, and Polarizability Effects on the Reproduction of the Density Anomaly of Liquid Water by Simple Potential Functions," M. W. Mahoney and W. L. Jorgensen, *J. Chem. Phys.* 115, 10758-10768 (2001).
143. "Rapid Estimation of Electronic Degrees of Freedom in Monte Carlo Calculations for Polarizable Models of Liquid Water," M. W. Mahoney and W. L. Jorgensen, *J. Chem. Phys.*, 114, 9337-9349 (2001).
144. "Diffusion Constant of the TIP5P Model of Liquid Water," M. W. Mahoney and W. L. Jorgensen, *J. Chem. Phys.*, 114, 363-366 (2001).
145. "A Five-Site Model for Liquid Water and the Reproduction of the Density Anomaly by Rigid, Nonpolarizable Potential Functions," M. W. Mahoney and W. L. Jorgensen, *J. Chem. Phys.*, 112, 8910-8922 (2000).
146. "Repression and Activation of Promoter-Bound RNA Polymerase Activity by Gal Repressor," H. E. Choy, R. R. Hanger, T. Aki, M. Mahoney, K. Murakami, A. Ishihama, and S. Adhya, *J. Mol. Biol.* 272: 293-300 1997.
147. "Discrete Representations of the Protein C-alpha Chain," X. F. de la Cruz, M. W. Mahoney, and B. K. Lee, *Fold. & Des.* 2: 223-234, 1997.

Classes Taught

- UCB Stat89a: “Linear Algebra for Data Science,” Spring 2018-19 (originally a 2-unit DS8 Connector on “Matrices and Graphs in Data Science,” Spring 2016-17), (Designed and taught; sophomore-level jupyter-based textbook on the applied mathematics and linear algebra of data in preparation.)
- PCMI Graduate Summer School on “Mathematics of Data” (with A. Gilbert and J. Duchi), July 2016. (Edited volume of lectures in press.)
- UCB Stat260/CS294: “Topics in Spectral Graph Methods,” Spring 2015. (Designed and taught; lecture notes available as a technical report.)
- UCB Stat260/CS294: “Randomized Algorithms for Matrices and Data,” Fall 2013. (Designed and taught; lecture notes available as a technical report; advanced undergraduate / graduate textbook in preparation.)
- Numerous graduate and undergraduate classes at Yale and Stanford in computer science, statistics, and applied mathematics.
- “Algorithms for Massive Data Set Analysis” Stanford, Fall 2009.
- Designed and co-taught (with Donald Brown) “Markov Chain Methods in General Equilibrium Theory” at Yale in Summer 2003.

Academic

Service to the Community

- Editor: M. W. Mahoney, J. C. Duchi, and A. C. Gilbert, *The Mathematics of Data: IAS / Park City Mathematics Institute (PCMI) Graduate Summer School Lecture Series, Volume 25*, 2018.
- Associate Editor, SIAM Journal on Mathematics of Data Science, 2018-present.
- Associate Editor, SIAM Book Series on Data Science, 2018-present.
- Associate Editor, Applied and Computational Harmonic Analysis, 2017-present.
- Associate Editor, Journal of Machine Learning Research, 2015-present.
- Guest Editor: A. Grama, T. Courtade, M. W. Mahoney, and T. Weissman, *Proceedings of the IEEE: Special Issue on the “Science of Information,”* 2017.
- Guest Editor: G. B. Giannakis, F. Bach, R. Cendrillon, M. W. Mahoney, and J. Neville, “Signal Processing for Big Data,” *IEEE Signal Processing Magazine*, vol. 31, pp. 15-16, (Editorial for Special Issue), September 2014.
- National Advisory Committee, Statistical and Applied Mathematical Sciences Institute (SAMSI), 2014-2020.
- Scientific Advisory Board, Helsinki Institute for Information Technology (HIIT), 2014-2017.
- Committee on the Analysis of Massive Data, The National Research Council of the National Academies, 2010-2011

- Science Foundation Ireland, Strategic Research Cluster External Review Panel, June 2011
- Regular member of review panel for NSF and related organizations
- Regular PC / Senior PC Member for conferences, reviewer for journals, etc.
- MMDS: Chair, founder, and lead organizer of MMDS Foundation; started (in 2006 with G. Golub) and ran (in 2008, etc., with P. Drineas, A. Shkolnik, G. Carlsson, and others) the biennial Workshop on Algorithms for Modern Massive Data Sets (the MMDS Meetings)

Grants:

- ONR Research Grant (with A. Shrivastava and R. Baraniuk), “Randomized Numerical Linear Algebra for Large-scale Learning and Inference,” 2018-2022, **\$400K**.
- NSF Research Grant (with B. Yu, F. Perez, R. Karp, and M. Jordan), “Berkeley Institute on Foundations of Data Analysis,” 2017-2020, **\$1.5M**.
- NSF Research Grant (with K. Ramchandran and S. Avestimehr), “Foundations of Coding for Modern Distributed Computing,” 2017-2021, **\$350K**.
- DOE Research Grant, “Scalable Inference for Adversarial Network Data,” 2016-2018, **\$90K**.
- DARPA Research Grant, D3M program, “Robust, Efficient, and Local Machine Learning Primitives,” 2017-2021, **\$1.35M**.
- Academic Research Gift: Adobe, Inc. ARO Research Grant, “Terabyte-scale Regression Diagnostic Methods for Interactive and Exploratory Analytics,” 2016-2018, **\$50K**.
- ARO Research Grant, “Local Algorithms for Large Informatics Graphs,” 2016-2019, **\$375K**.
- Cray Research Grant, “Implementing and Evaluating Matrix Algorithms in Spark on High Performance Computing Platforms for Science Applications,” 2015-2018, **\$1.0M**.
- UCB Internal Research Grant, via BDD, “Improving the scaling of deep learning networks by characterizing and exploiting soft convexity,” 2016-2018, **\$150K**.
- NSF Research Grant, via Purdue CSoI, (with D. Gleich) “Quantifying the information content of a graph via information in graph diffusions,” 2015-2018, **\$225K**.
- NSF Travel Grant, “Streaming Algorithms for Fundamental Computations in Numerical Linear Algebra,” (with J. Demmel, O. Schwartz, and S. Toldeo) 2015-2019, **\$40K**.
- DARPA Research Grant, GRAPHS program, “Machine learning methods and large informatics graphs,” 2014-2016, **\$450K**.
- NSF Research Grant, “Characterizing and exploiting tree-like structure in large social and information networks,” 2014-2017, **\$500K**.

- NSF Research Grant, “Randomized Numerical Linear Algebra (RandNLA) for multi-linear and non-linear data,” (with P. Drineas) 2014-2017, **\$500K**.
- DARPA Research Grant, XDATA program, “Implementing Randomized Linear Algebra in BDAS: Sub-award to: The Berkeley Data Analysis System (BDAS): An Open Source Platform for Big Data,” 2014-2017, **\$885K**.
- DOE Research Grant, “Scalable Statistics and Machine Learning for Data-Centric Science,” 2013-2016, **\$300K**.
- ARO Research Grant, “Scalable Matrix Algorithms for Interactive Analytics of Very Large Informatics Graphs,” 2013-2016, **\$300K**.
- NSF Research Grant, “Leverage Subsampling for Regression and Dimension Reduction” (with P. Ma and B. Yu), 2012-2014, **\$225K**.
- DARPA Research Grant, GRAPHS program, “Sparse Projections Achieving Randomization in Tree-like Networks,” (with B. Sullivan) 2012-2013, **\$300K** covers partial salary and one graduate student.
- NSF Research Grant, “Randomized Algorithms in Linear Algebra and Numerical Evaluations on Massive Datasets,” (with P. Drineas and M. Saunders) 2010-2012, **\$500K** covers two graduate students.
- NSF Research Grant, “Geometric Network Analysis Tools: Algorithmic Methods for Identifying Structure in Large Informatics Graphs,” (with S. Rao) 2010-2013, **\$1.2M** covers two graduate students and one postdoctoral fellow.
- AFOSR/ONR grant for MMDS 2010 (with G. Carlsson) **\$10K**.
- NSF grant for MMDS 2010 (with G. Carlsson) **\$20K**.
- AFOSR Research Grant, “Geometric Network Analysis,” (with G. Carlsson) 2009-2012, **\$500K** covers MWM salary and one graduate student.
- NSF grant for MMDS 2006, (with G. Golub) **\$15K**.

Other Dissertation Committees

- Vipul Gupta, UC Berkeley, 2019
- Danqing Zhang, UC Berkeley, 2017
- Di Wang, UC Berkeley, 2017
- Peter Jin, UC Berkeley, 2017
- Francois Belletti, UC Berkeley, 2016
- Vu Pham, UC Berkeley, 2015
- Rajendra Shinde, Stanford, 2012
- Christos Boutsidis, Rensselaer Polytechnic Institute, 2011
- Pradipta Mitra, Yale University, June 2008
- Lek-Heng Lim, Stanford University, April 2007

Patents Applied for and Issued

- “Systems and Methods for Identifying Potentially Interesting Events in Extended Recordings,” (with Y. Galant, M. Boliek, and E. Steinberg) 2015.
- “Suggesting Contacts for Social Networks,” (with M. Slaney, S. Ravikumar, A. Anagnostopoulos, M. Mahdian, A. Ghosh, A. Jaffe, and M. Rothenberg) 2008.

- “System and method of feature selection for text classification using subspace sampling,” (with V. Josifovski, A. Dasgupta, P. Drineas, B. Harb) 2007.

Miscellaneous

- Leadership Council, International Computer Science Institute, 2016-present.
- Industry Alliance Program Committee, UC Berkeley Statistics Department, 2015-present.
- Member, Center for Science of Information (CSoI), NSF STC, Purdue University, 2014-2019.
- Regular consultant to industry.
- Kavli Frontiers Fellow, Kavli Foundation, 2012.
- J. Tinsley Oden Visiting Faculty Fellowship, UT Austin, Fall 2005.
- Visiting Fellow, Program in Multiscale Geometry and Analysis in High Dimensions, IPAM, UCLA, Fall 2004.

Meetings, Workshops, and Events Organized or Co-organized

XLDB Extremely Large Databases 2019 Meeting (with J. Becla, et al.), SLAC, Stanford, CA, April 2019

UC Berkeley Simons Institute Program on “Foundations of Data Science,” (with D. Woodruff, K. Clarkson, R. Kannan, A. Montanari, S. Vempala, and R. Ward, Berkeley, CA, Fall 2018

Recent Advances in Randomized Numerical Linear Algebra (with R. Kannan and D. Woodruff), NII Shonan Meeting, Shonan Village Center, Japan, July 2016.

Park City Mathematics Institute (PCMI) Graduate Summer School on “Mathematics of Data” (with A. Gilbert and J. Duchi), Park City, UT, July 2016.

MMDS 2016: Workshop on Algorithms for Modern Massive Data Sets (with A. Shkolnik and P. Drineas), Berkeley, CA, June 2016

MMDS 2014: Workshop on Algorithms for Modern Massive Data Sets (with A. Shkolnik, P. Drineas, R. Zadeh, and F. Perez), Berkeley, CA, June 2014

UC Berkeley Simons Institute Program on “Theoretical Foundations of Big Data Analysis,” (with M. Jordan, S. Muthukrishnan, P. Buehlmann, R. Kannan, and S. Boyd), Berkeley, CA, Fall 2013

2012-2013 Massive Datasets Program Opening Tutorials and Workshop, SAMSI, (with N. Altman, J. Sun, and D. Ushizima) Research Triangle Park, NC, September 2012

SAMSI Program on Statistical and Computational Methodology for Massive Datasets, (with M. Jordan, K. Kafadar, S. Sain, J. Sun, and A. Szalay) Research Triangle Park, NC, 2012-2013

MMDS 2012: Workshop on Algorithms for Modern Massive Data Sets (with G. Carlsson, P. Drineas, and A. Shkolnik), Stanford, CA, July 2012

Sparse Representation and Low-rank Approximation, workshop in conjunction with NIPS 2011 (with A. Talwalkar, L. Mackey, M. Mohri, F. Bach, M. Davies, R. Gribonval, and G. Obozinski), Granada, Spain, December 2011

Large Graphs: Modeling, Algorithms and Applications, workshop at IMA, (with E. Kolaczyk and M. Maggioni) Minneapolis, MN, October 2011

Low-rank Methods for Large-scale Machine Learning, workshop in conjunction with NIPS 2010 (with A. Talwalkar and M. Mohri), Whistler, Canada, December 2010

2010-2011 Program on Complex Networks Opening Tutorials and Workshop, SAMSI, Research Park Triangle, NC, August 2010

MMDS 2010: Workshop on Algorithms for Modern Massive Data Sets (with G. Carlsson, P. Drineas, L.-H. Lim, and A. Shkolnik), Stanford, CA, June 2010

Randomized Algorithms for Scalable Dense Linear Algebra, session in conjunction with 14th SIAM Conference on Parallel Processing and Scientific Computing (with S. A. Toledo), Seattle, WA, February 2010

SIAM 2009 Applied Linear Algebra Conference Organizing Committee, Monterey, CA, October 2009

European Workshop on Challenges in Modern Massive Data Sets (with M. Morup, L.-H. Lim, L. K. Hansen, G. Carlsson), Lyngby, Denmark, July 2009

MMDS 2008: Workshop on Algorithms for Modern Massive Data (with G. Carlsson, P. Drineas, and L.-H. Lim), Stanford, CA, June 2008

Data Mining for Biomedical Informatics, workshop in conjunction with SIAM SDM 2008, (with P. Drineas and R. Kuang), Atlanta, GA, April 2008

Novel Matrix Methods for Internet Data Mining, session in conjunction with ICIAM 2007 (with L.-H. Lim, G. Golub, and A. Dasgupta) Zurich, Switzerland, July 2007

Data Mining for Biomedical Informatics, workshop in conjunction with SIAM SDM 2007, (with P. Drineas and V. Kumar), Minneapolis, MN, April 2007

Web Information Retrieval and Linear Algebra Algorithms, (with A. Frommer and D. B. Szyld), Schloss Dagstuhl, Dagstuhl, Germany, February 2007

MMDS 2006: Workshop on Algorithms for Modern Massive Data Sets, (with G.H. Golub, P. Drineas, and L.-H. Lim), Stanford, CA, June 2006

References

- Available upon request.