Shashanka Ubaru

PO Box 218 1101 Kitchawan Rd
Yorktown Heights, NY , USA 10598
✉ *Shashanka.Ubaru@ibm.com*
shashankaubaru.github.io

Research Interests

Machine learning, numerical linear and tensor algebra, coding theory applications, approximation theory and algorithms, quantum topological data analysis, mutlilabel classification, and graph neural networks.

Education

2012-18 **University of Minnesota**, *Minneapolis*, MN, USA.
Ph.D. Computer Science May 2018
Advisor: Yousef Saad
Thesis: *Algorithmic advances in learning from large dimensional matrices and scientific data*

M.S. Computer Science October 2015
M.S. Electrical Engineering November 2014
Advisors: Yousef Saad and Arya Mazumdar
Thesis: *Randomized techniques for matrix decomposition and estimating the approximate rank of a matrix*

2008-12 **M.S. Ramaiah Institute of Technology**, *Bangalore*, India.
B.Eng. Electronics and Communication

Employment

2018 - present **IBM T.J. Watson Research Center**, *1101 Kitchawan Rd,Yorktown Heights*, NY, USA.
Research Staff Member Jan, 2020 - present
Goldstine Postdoctoral Fellow Aug, 2018 - Dec, 2019

2013-2018 **University of Minnesota**, *200 Union St SE, Minneapolis*, MN, USA.
Research Assistant, Department of Computer Science. Sep, 2013 - May, 2018

2016 & 2017 **Lawrence Berkeley National Laboratory**, *1 Cyclotron Rd, Berkeley*, CA, USA.
Research Intern, Neural Systems and Data Science Lab June - Aug, 2017
Research Intern, Scientific Data Management group May - Aug, 2016

2013 **Seagate Technology**, *1280 Disc Dr, Shakopee*, MN, USA.
Signal Processing Intern May - Aug, 2013

Publications

Conference proceedings

2021 *Efficient Scaling of Dynamic Graph Neural Networks*.
V. Chakaravarthy, S. Pandian, S. Raje, Y. Sabharwal, T. Suzumura, and **S. Ubaru**.
Supercomputing (SC21)

2021 *Analysis of stochastic Lanczos quadrature for spectrum approximation*.
T. Chen, T. Trogdon, and **S. Ubaru**.
International Conference on Machine Learning (ICML). *Long presentation*

2021 *Projection techniques to update the truncated SVD of evolving matrices*.
V. Kalantzis, G. Kollias, **S. Ubaru**, A. Nikolakopoulos, L. Horesh, and K.L. Clarkson.
International Conference on Machine Learning (ICML).

2021 *Sparse graph based sketching for fast numerical linear algebra*.
D. Hu, **S. Ubaru**, A. Gittens, K. Clarkson, L. Horesh, and V. Kalantzis.
International Conference on Acoustics, Speech, and Signal Processing (ICASSP).

2021 *Dynamic Graph Convolutional Networks Using the Tensor M-Product*.
O. Malik, **S. Ubaru**, L. Horesh, M. Kilmer, and H. Avron.
SIAM International Conference on Data Mining (SDM).

2020   *Multilabel Classification by Hierarchical Partitioning and Data-dependent Grouping*.
       S. Ubaru, S. Dash, O. Gunluk, and A. Mazumdar.
       Advances in Neural Information Processing Systems (NeurIPS)

2019   *Spectrum-Adapted Polynomial Approximation for Matrix Functions*.
       L. Fan, D. Shuman, **S. Ubaru**, and Y. Saad
       International Conference on Acoustics, Speech, and Signal Processing (ICASSP).

2019   *Find the dimension that counts: Fast dimension estimation and Krylov PCA*.
       S. Ubaru, A.-K. Seghouane, and Y. Saad
       SIAM International Conference on Data Mining (SDM).

2018   *Spectral Sums Beyond Fast Matrix Multiplication: Algorithms and Hardness*.
       C. Musco, P. Netrapalli, A. Sifford, **S. Ubaru**, and D. P. Woodruff
       Innovations in Theoretical Computer Science (ITCS).

2017 UoI-$NMF_{cluster}$: *A Robust Nonnegative Matrix Factorization Algorithm for Improved Parts-Based
       Decomposition and Reconstruction of Noisy Data*.
       S. Ubaru, K. Wu, and K. E. Bouchard
       IEEE International Conference on Machine Learning and Applications (ICMLA).
       Best Paper Award.

2017   *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, and S. Bhattacharyya
       Advances in Neural Information Processing Systems (NeurIPS).

2017   *Multilabel Classification with Group Testing and Codes*.
       S. Ubaru and A. Mazumdar
       International Conference on Machine Learning (ICML).

2016   *Fast methods for estimating the Numerical rank of large matrices*.
       S. Ubaru and Y. Saad
       International Conference on Machine Learning (ICML).

2016   *Group testing schemes from low-weight codewords of BCH codes*.
       S. Ubaru, A. Mazumdar, and A. Barg
       IEEE International Symposium on Information Theory (ISIT).

2015   *Low rank approximation using error correcting coding matrices*.
       S. Ubaru, A. Mazumdar, and Y. Saad
       International Conference on Machine Learning (ICML).

Journal articles

2020   *Spectrum-Adapted Polynomial Approximation for Matrix Functions with Applications in Graph Signal
       Processing*.
       L. Fan, D. Shuman, **S. Ubaru**, and Y. Saad.
       Algorithms, Special Issue : Efficient Graph Algorithms in Machine Learning, 13(11), 295.

2019   *Sampling and multilevel coarsening algorithms for fast matrix approximations*.
       S. Ubaru and Y. Saad
       Numerical Linear Algebra with Applications 26.3 : e2234.

2017   *Fast estimation of* $\mathtt{tr}(f(A))$ *via Stochastic Lanczos Quadrature*.
       S. Ubaru, J. Chen, and Y. Saad
       SIAM Journal on Matrix Analysis and Applications (SIMAX), 38(4), 1075–1099.

2017   *Low rank approximation and decomposition of large matrices using error correcting codes*.
       S. Ubaru, A. Mazumdar, and Y. Saad
       IEEE Transactions on Information Theory, 63(9), 5544–5558.

2017   *Formation enthalpies for transition metal alloys using machine learning*.
       S. Ubaru, A. Miedlar, Y. Saad, and J R. Chelikowsky
       Physical Review B, (Vol.95, No.21).

2017   *Fast estimation of approximate matrix ranks using spectral densities*.
       S. Ubaru, Y. Saad, and A.-K. Seghouane
       Neural Computation, 29(5):1317–1351.

2017   *Improving the Incoherence of a Learned Dictionary via Rank Shrinkage*.
       S. Ubaru, A.-K. Seghouane, and Y. Saad
       Neural Computation, 29(1):263–285.

2012 *Displaying gray scales by cross pairing select and data voltages in multi-line addressed LCD*.
S. Ubaru and T.N. Ruckmongathan
IEEE Journal of Display Technology, 8(11), 669–677.

Book chapter

2018 *Applications of trace estimation techniques*.
S. Ubaru and Y. Saad
High Performance Computing in Science and Engineering, LNCS book series, vol. 11087, ch. 2, pp 19–33.

Workshop papers

2019 *Multilabel prediction in log time and data-dependent grouping*.
S. Ubaru, S. Dash, O. Gunluk, and A. Mazumdar
NeurIPS - Workshop on Information Theory and Machine Learning.

2019 *Tensor graph neural networks for prediction on time varying graphs*.
O. Malik, **S. Ubaru**, L. Horesh, M. Kilmer, and H. Avron
NeurIPS - Workshop on Graph Representation Learning.

2019 *Union of Intersections (UoI) for interpretable data driven discovery and prediction in neuroscience*.
K. E. Bouchard, P. Sachdeva, S. Bhattacharyya, M. Balasubramanian, and **S. Ubaru**
Computational and Systems Neuroscience (Cosyne).

Submissions

2021 *Quantum Topological Data Analysis with Linear Depth and Exponential Speedup*.
S. Ubaru, I. Akhalwaya, M. Squillante, K. Clarkson, and L. Horesh.

2021 *Dynamic graph based epidemiological model for COVID-19 contact tracing data analysis and optimal testing prescription*.
S. Ubaru, G. Cohen, and L. Horesh.
Journal of Bioinformatics, Special Issue. In Revision

2020 *Unsupervised Hierarchical Graph Representation Learning with Variational Bayes*.
S. Ubaru and J. Chen.

2019 *Provably convergent acceleration in factored gradient descent with applications in matrix sensing*.
T Ajayi, D Mildebrath, A. Krylidis, **S. Ubaru**, G. Kollias, and K. E. Bouchard.

Patents

2021 *US20210174242A1 - Automatic multilabel classification using machine learning* .
S. Ubaru, S. Dash, O. Gunluk, A. Mazumdar and L. Horesh.

2021 *US20210090182A1 - Tensor-based predictions from analysis of time-varying graphs*.
O. Malik, **S. Ubaru**, L. Horesh, M. Kilmer, and H. Avron.

2019 *P201809232 - Short-depth quantum amplitude estimation without eigenstate collapse*.
I. Akhalwaya, K. Clarkson, L. Horesh, M. Squillante, **S. Ubaru**, and V. Kalantzis. (Filled)

2020 *P201907231 - Streaming Algorithm Based Matrix Sketching Using Analog Crossbar Architectures*.
L. Horesh, M. Onen, H. Avron, T. Gokmen, **S. Ubaru**, and V. Kalantzis. (Filled)

2020 *P202002430 - Accelerated Quasi-Newton methods on analog crossbar hardware*.
V. Kalantzis, **S. Ubaru**, L. Horesh, H. Avron, and M. Onen. (Filled)

2020 *P202006188 - Fast PCA of evolving data using analog crossbar array*.
S. Ubaru, V. Kalantzis, L. Horesh, M. Squillante, and H. Avron. (Filled)

2021 *P202006530 - Determining triangles in graph data structures using crosspoint array* .
V. Kalantzis, **S. Ubaru**, L. Horesh, and H. Avron. (Filled)

2021 *P202008482 - Projection-based techniques for updating SVD in evolving datasets*.
V. Kalantzis, G. Kollias, **S. Ubaru**, L. Horesh, and K. Clarkson. (Filled)

Presentations and Visits

2021 *Into another dimension - a new tensor algebra for learning on dynamic graphs*
- Seminar at University of Albany, NY.

Algorithms for estimating spectral sums for large matrices
- Oden Institute Seminar Series, University of Texas, Austin.

Novel Tensor Framework for Data Representation and Compression
- SIAM Conference on Computational Science and Engineering (CSE21).

2020 *Multilabel Classification by Hierarchical Partitioning and Data-dependent Grouping*
- Neural Information Processing Systems (NeurIPS) Virtually held.

Novel tensor framework for model reduction and neural networks
- ICERM: Workshop on Algorithms for Dimension and Complexity Reduction, Brown University, RI

2019 *Tensor graph neural networks for prediction on time varying graphs*
- NeurIPS: Workshop on Graph Representation Learning, Vancouver, Canada

Multilabel prediction in log time and data-dependent grouping
- NeurIPS: Workshop on Information Theory and Machine Learning, Vancouver, Canada

Spectrum approximation by Lanczos Quadrature and Preconditioned SVRG
- 20th International Conference On Preconditioning, Minneapolis, MN, USA.

Find the dimension that counts: Fast dimension estimation and Krylov PCA
- SIAM International Conference on Data Mining (SDM), Calgary, Canada.

Error Correcting Codes for Machine Learning.
- CSA Seminar, Indian Institute of Science (IISc), Bengaluru, India.

2018 *Error Correcting Codes for Machine Learning.*
- Theory Seminar, University of Massachusetts, Amherst, MA.
- IP Seminar, IBM T.J. Watson Research Center, NY.

Lawrence Berkeley National Laboratory, CA (Two weeks visit in June).

2017 $UoI\text{-}NMF_{cluster}$: *A Robust Nonnegative Matrix Factorization Algorithm for Noisy Data*
- International Conference on Machine Learning and Applications (ICMLA), Cancun, Mexico.

Multilabel Classification with Group Testing and Codes
- Neural Systems and Engineering Labs, Lawrence Berkeley National Laboratory, CA.
- International Conference on Machine Learning (ICML), Sydney, Australia.

Error Correcting Codes for Machine Learning.
- The University of Melbourne, Melbourne, Australia (Two weeks visit in August).

UoI-NMFcluster and UoI-CUR: Union of Intersections methods for matrix approximations
- Neural Systems and Engineering Labs, Lawrence Berkeley National Laboratory,CA.

Microsoft Research, Bangalore, India (Two days visit in May).

2016 *Error correcting codes for low rank approximation and group testing*
- BLISS Seminar, University of California, Berkeley, CA.

Fast methods for estimating the Numerical rank of large matrices
- International Conference on Machine Learning (ICML), New York, NY.

2015 *Low rank approximation using error correcting coding matrices.*
- International Conference on Machine Learning (ICML), Lille, France.

Awards

2019 *UMN Best Dissertation Award*, Computer Science Department nominee.

2018 **Herman Goldstine Fellowship**, IBM Research.

2017 **Best Paper Award**, International Conference on Machine Learning and Applications (ICMLA).

2015,16,17 *ICML Travel Scholarship.*

2015 *CS Department Travel Award.*

2011   *Visiting Research Student Program*, Raman Research Institute.

Teaching

IBM Research

Summer 2021   Mathematics of Big Data: Sketching and (Multi-) Linear Algebra at MSRI Berkeley, *Instructor*.

University of Minnesota

Spring 2018   CSci 2033, Elementary Computational Linear Algebra, *Recitation Instructor*.

Fall 2017   CSci 5304, Computational Aspects of Matrix Theory, *Teaching Assistant*.

Spring 2017   CSci 8314, Sparse Matrix Computations, *Temporary Instructor*.

Mentoring

IBM Research - Interns

2021   Paz Fink Sustin. (PhD Student - Tel Aviv University)

2021   Lucky Yerimah. (PhD Student - RPI)

2021   Emily Thompson. (REU Student via. DIMACS)

2020   Dong Hu. (PhD Student - RPI)

2019   Osman Malik. (PhD Student - UC, Boulder)

Service

Reviewer: *Conferences:*

Neural Information Processing Systems (NeurIPS) 2016 - 21 ; International Conference on Machine Learning (ICML) 2018 - 21 ; International Conference on Learning Representations (ICLR) 2021; AAAI conference on Artificial Intelligence 2019 - 21 ; Conference on Uncertainty in Artificial Intelligence (UAI) 2019 - 20; IEEE International Conference On Machine Learning And Applications 2018 - 20; IEEE International Symposium on Information Theory (ISIT) 2017.

Journals:

PLOS One, SIAM Journal on Matrix Analysis and Applications, Acta Materialia, IEEE Transactions on Signal Processing, IEEE Transactions on Image Porcessing, Signal Processing Letters, IEEE Transactions on Information Theory, Electronic Transactions on Numerical Analysis, IEEE Transactions on Signal and Information Processing over Networks, Journal of Mathematics and Applications, Mathematical Modelling and Analysis.

Organizer: ICMLA Challenge 2018 - Parts based decomposition of noisy data.

MSRIT ROBONXG-2012, a week long robotics festival.

Programming skills

C, C++, Matlab, Python, LATEX, Basic web programming.

References

Yousef Saad
CSE Distinguished Professor,
Computer Science and Engineering,
University of Minnesota- Twin Cities, MN, USA.
saad@umn.edu.

Arya Mazumdar
Associate Professor,
The Halicioglu Data Science Institute,
University of California, San Diego, CA, USA.
 arya@ucsd.edu.

Lior Horesh
Senior Manager
Mathematics of AI
IBM T.J. Watson Research Center, NY, USA
lhoresh@us.ibm.com

Sanjeeb Dash
Manager
Foundation of Optimization
IBM T.J. Watson Research Center, NY, USA
sanjeebd@us.ibm.com