

Satyen Kale

satyen@satyenkale.com
www.satyenkale.com

Curriculum Vitae

RESEARCH INTERESTS

Design of efficient and practical algorithms for fundamental problems in **machine learning**, specifically:

- decision-making under uncertainty
- optimization
- statistical learning theory
- privacy

EDUCATION

Princeton University, M.A. (2004) and Ph.D. (2007), Computer Science

Thesis: Efficient Algorithms Using the Multiplicative Weights Updates Method. Advised by Prof. Sanjeev Arora.

Indian Institute of Technology Bombay, B.Tech. (2002), Computer Science and Engineering

Thesis: Spectral Algorithms for Data Representation and Manipulation. Advised by Prof. Abhiram Ranade.

Indian Institute of Statistics, Calcutta, summers 1999-2002

Participant in Nurture Program in Mathematics, equivalent of a Master's course in mathematics.

WORK EXPERIENCE

Google Research, New York, NY, Staff Research Scientist, 2016-present

Yahoo! Research, New York, NY, Senior Research Scientist, 2013-2016

Columbia University, New York, NY, Adjunct Professor, Spring 2016 and Fall 2019

IBM Research, Yorktown Heights, NY, Research Staff Member, 2011-2013

Yahoo! Research, Santa Clara, CA, Postdoctoral Researcher, 2009-2011

Microsoft Research Redmond and New England, Postdoctoral Researcher, 2007-2009

Microsoft Research Silicon Valley Center, Mountain View, CA, Summer Intern, 2006

IBM Almaden Research Center, San Jose, CA, Summer Intern, 2005

ETH Zürich, Switzerland, Summer Intern, 2001

PUBLICATIONS

PEER-REVIEWED CONFERENCE AND JOURNAL PAPERS

[1] **Differentially Private and Lazy Online Convex Optimization**

N. Agarwal, S. Kale, K. Singh, and A. Thakurta. *Conference on Learning Theory (COLT)*, 2023

- [2] **Beyond uniform lipschitz condition in differentially private optimization**
R. Das, S. Kale, Z. Xu, T. Zhang, and S. Sanghavi. *International Conference on Machine Learning (ICML)*, 2023
- [3] **Efficient Training of Language Models using Few-Shot Learning**
S. J. Reddi, S. Miryoosefi, S. Karp, S. Krishnan, S. Kale, S. Kim, and S. Kumar. *International Conference on Machine Learning (ICML)*, 2023
- [4] **On the Convergence of Federated Averaging with Cyclic Client Participation**
Y. J. Cho, P. Sharma, G. Joshi, Z. Xu, S. Kale, and T. Zhang. *International Conference on Machine Learning (ICML)*, 2023
- [5] **From Gradient Flow on Population Loss to Learning with Stochastic Gradient Descent**
C. M. De Sa, S. Kale, J. D. Lee, A. Sekhari, and K. Sridharan. *Neural Information Processing Systems (NeurIPS)*, 2022
- [6] **Reproducibility in optimization: Theoretical framework and limits**
K. Ahn, P. Jain, Z. Ji, S. Kale, P. Netrapalli, and G. I. Shamir. *Neural Information Processing Systems (NeurIPS)*, 2022
- [7] **Private Matrix Approximation and Geometry of Unitary Orbits**
O. Mangoubi, Y. Wu, S. Kale, A. Thakurta, and N. K. Vishnoi. *Conference on Learning Theory (COLT)*, 2022
- [8] **Self-Consistency of the Fokker Planck Equation**
Z. Shen, Z. Wang, S. Kale, A. Ribeiro, A. Karbasi, and H. Hassani. *Conference on Learning Theory (COLT)*, 2022
- [9] **Pushing the efficiency-regret Pareto frontier for online learning of portfolios and quantum states**
J. Zimmert, N. Agarwal, and S. Kale. *Conference on Learning Theory (COLT)*, 2022
- [10] **Agnostic learnability of halfspaces via logistic loss**
Z. Ji, K. Ahn, P. Awasthi, S. Kale, and S. Karp. *International Conference on Machine Learning (ICML)*, 2022
- [11] **Federated functional gradient boosting**
Z. Shen, H. Hassani, S. Kale, and A. Karbasi. *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2022
- [12] **Efficient Methods for Online Multiclass Logistic Regression**
N. Agarwal, S. Kale, and J. Zimmert. *International Conference on Algorithmic Learning Theory (ALT)*, 2022
- [13] **Breaking the centralized barrier for cross-device federated learning**
S. P. Karimireddy, M. Jaggi, S. Kale, M. Mohri, S. Reddi, S. U. Stich, and A. T. Suresh. *Neural Information Processing Systems (NeurIPS)*, 2021
- [14] **SGD: The role of implicit regularization, batch-size and multiple-epochs**
A. Sekhari, K. Sridharan, and S. Kale. *Neural Information Processing Systems (NeurIPS)*, 2021
- [15] **Learning with user-level privacy**
D. Levy, Z. Sun, K. Amin, S. Kale, A. Kulesza, M. Mohri, and A. T. Suresh. *Neural Information Processing Systems (NeurIPS)*, 2021

- [16] **A Deep Conditioning Treatment of Neural Networks**
N. Agarwal, P. Awasthi, S. Kale. *International Conference on Algorithmic Learning Theory (ALT)*, 2021.
- [17] **PAC-Bayes Learning Bounds for Sample-Dependent Priors**
P. Awasthi, S. Kale, S. Karp, M Mohri. *Neural Information Processing Systems (NeurIPS)*, 2020.
- [18] **Estimating Training Data Influence by Tracing Gradient Descent**
G. Pruthi, F. Liu, S. Kale, M. Sundararajan. *Neural Information Processing Systems (NeurIPS)*, 2020.
- [19] **SCAFFOLD: Stochastic Controlled Averaging for Federated Learning**
S. P. Karimireddy, S. Kale, M. Mohri, S. J. Reddi, S. U. Stich, A. T. Suresh. *International Conference on Machine Learning (ICML)*, 2020.
- [20] **Breaking the Glass Ceiling for Embedding-Based Classifiers for Large Output Spaces**
C. Guo, A. Mousavi, X. Wu, D. N. Holtmann-Rice, S. Kale, S. J. Reddi, S. Kumar. *Neural Information Processing Systems (NeurIPS)*, 2019.
- [21] **Hypothesis Set Stability and Generalization**
D. J. Foster, S. Greenberg, S. Kale, H. Luo, M. Mohri, K. Sridharan. *Neural Information Processing Systems (NeurIPS)*, 2019.
- [22] **Escaping Saddle Points with Adaptive Gradient Methods**
M. Staib, S. J. Reddi, S. Kale, S. Kumar, S. Sra. *International Conference on Machine Learning (ICML)*, 2019.
- [23] **Stochastic Negative Mining for Learning with Large Output Spaces**
S. J. Reddi, S. Kale, F. X. Yu, D. Holtmann-Rice, J. Chen, S. Kumar. *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2019.
- [24] **Adaptive Methods for Nonconvex Optimization**
M. Zaheer, S. J. Reddi, D. S. Sachan, S. Kale, S. Kumar. *Neural Information Processing Systems (NeurIPS)*, 2018.
- [25] **Online Learning of Quantum States**
S. Aaronson, X. Chen, E. Hazan, S. Kale, and A. Nayak. *Neural Information Processing Systems (NeurIPS)*, 2018.
- [26] **On the convergence of Adam and Beyond**
S. J. Reddi, S. Kale, S. Kumar. *International Conference on Learning Representations (ICLR)*, 2018. **Best paper award.**
- [27] **Loss Decomposition for Fast Learning in Large Output Spaces**
I. E. H. Yen, S. Kale, F. X. Yu, D. Holtmann-Rice, S. Kumar and P. Ravikumar. *International Conference on Machine Learning (ICML)*, 2018.
- [28] **Logistic Regression: The Importance of Being Improper**
D. J. Foster, S. Kale, H. Luo, M. Mohri and Karthik Sridharan. *Conference on Learning Theory (COLT)*, 2018. **Best student paper award.**

- [29] **Parameter-Free Online Learning via Model Selection**
D. J. Foster, S. Kale, M. Mohri and K. Sridharan. *Neural Information Processing Systems* (NeurIPS), 2017.
- [30] **Adaptive Feature Selection: Computationally Efficient Online Sparse Linear Regression under RIP**
S. Kale, Z. Karnin, T. Liang and D. Pál. *International Conference on Machine Learning* (ICML), 2017.
- [31] **Online Semidefinite Programming**
N. Elad, S. Kale and J. Naor. *International Colloquium on Automata, Languages and Programming* (ICALP), 2016.
- [32] **Online Sparse Linear Regression**
D. Foster, S. Kale, and H. Karloff. *Conference on Learning Theory* (COLT), 2016.
- [33] **Online Gradient Boosting**
A. Beygelzimer, E. Hazan, S. Kale, and H. Luo. *Neural Information Processing Systems* (NeurIPS), 2015.
- [34] **Optimal and Adaptive Algorithms for Online Boosting**
A. Beygelzimer, S. Kale, and H. Luo. *International Conference on Machine Learning* (ICML), 2015. **Best paper award.**
- [35] **Budgeted Prediction With Expert Advice**
K. Amin, S. Kale, G. Tesauro, and D. Turaga. *Conference of Association for the Advancement of Artificial Intelligence* (AAAI), 2015.
- [36] **Beyond the Regret Minimization Barrier: an Optimal Algorithm for Stochastic Strongly-Convex Optimization**
E. Hazan and S. Kale. *Journal of Machine Learning Research* (JMLR), 2014. Preliminary version in *Conference on Learning Theory* (COLT), 2011.
- [37] **Taming the Monster: A Fast and Simple Algorithm for Contextual Bandits**
A. Agarwal, D. Hsu, J. Langford, S. Kale, L. Li, and R. E. Schapire. *International Conference on Machine Learning* (ICML), 2014.
- [38] **Multiarmed Bandits With Limited Expert Advice**
S. Kale. *Conference on Learning Theory* (COLT), 2014.
- [39] **Adaptive Market Making via Online Learning**
J. Abernethy and S. Kale. *Neural Information Processing Systems* (NeurIPS), 2013.
- [40] **Bargaining for Revenue Shares on Tree Trading Networks**
A. Ghosh, S. Kale, K. Lang and B. Moseley. *International Joint Conference on Artificial Intelligence* (ICJAI), 2013.
- [41] **The Approximability of the Binary Paintshop Problem**
A. Gupta, S. Kale, V. Nagarajan, R. Saket and B. Schieber. *International Workshop on Approximation Algorithms for Combinatorial Optimization Problems* (APPROX), 2013.
- [42] **Noise Tolerance of Expanders and Sublinear Expander Reconstruction**
S. Kale, Y. Peres, and C. Seshadhri. *SIAM Journal on Computing* (SICOMP), 2013. Preliminary version in *Foundations of Computer Science* (FOCS), 2008.

- [43] **Near-Optimal Algorithms for Online Matrix Prediction**
E. Hazan, S. Kale and S. Shalev-Shwartz. *Conference on Learning Theory (COLT)*, 2012.
- [44] **Projection-free Online Learning**
E. Hazan and S. Kale. *International Conference on Machine Learning (ICML)*, 2012.
- [45] **Efficient and Practical Stochastic Subgradient Descent for Nuclear Norm Regularization**
H. Avron, S. Kale, S. P. Kasiviswanathan, and V. Sindhvani. *International Conference on Machine Learning (ICML)*, 2012.
- [46] **Contextual Bandit Learning with Predictable Rewards**
A. Agarwal, M. Dudik, S. Kale, J. Langford and R. E. Schapire. *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2012.
- [47] **The Multiplicative Weights Update method: a Meta-Algorithm and some Applications**
S. Arora, E. Hazan, and S. Kale. *Theory of Computing Journal*, 2012.
- [48] **An Online Portfolio Selection Algorithm with Regret Logarithmic in Price Variation**
E. Hazan and S. Kale. *Mathematical Finance*, 2012. Preliminary version appeared as **On Stochastic and Worst-case Models for Investing** in *Neural Information Processing Systems (NeurIPS)*, 2009.
- [49] **Newtron: an Efficient Bandit algorithm for Online Multiclass Prediction**
E. Hazan and S. Kale. *Neural Information Processing Systems (NeurIPS)*, 2011.
- [50] **Efficient Optimal Learning for Contextual Bandits**
M. Dudik, D. Hsu, S. Kale, N. Karampatziakis, J. Langford, L. Reyzin, and T. Zhang. *Conference on Uncertainty in Artificial Intelligence (UAI)*, 2011.
- [51] **Who Moderates the Moderators?: Crowdsourcing Abuse Detection in User-Generated Content**
A. Ghosh, S. Kale, and R. P. McAfee. *Conference on Electronic Commerce (EC)*, 2011.
- [52] **Cross-Validation and Mean-Square Stability**
S. Kale, R. Kumar and S. Vassilvitskii. *Innovations in Computer Science (ICS)*, 2011.
- [53] **Combinatorial Approximation Algorithms for MaxCut Using Random Walks**
S. Kale and C. Seshadhri. *Innovations in Computer Science (ICS)*, 2011.
- [54] **An Expansion Tester for Bounded Degree Graphs**
S. Kale and C. Seshadhri. *SIAM Journal on Computing (SICOMP)*, 2011. Preliminary version in *International Colloquium on Automata, Languages and Programming (ICALP)*, 2008.
- [55] **Better Algorithms for Benign Bandits**
E. Hazan and S. Kale. *Journal of Machine Learning Research (JMLR)*, 2011. Preliminary version in *Symposium on Discrete Algorithms (SODA)*, 2009.
- [56] **Non-Stochastic Bandit Slate Problems**
S. Kale, L. Reyzin and R. E. Schapire. *Neural Information Processing Systems (NeurIPS)*, 2010.
- [57] **Learning Rotations with Little Regret**
E. Hazan, S. Kale and M. K. Warmuth. *Conference on Learning Theory (COLT)*, 2010.

- [58] **Extracting Certainty from Uncertainty: Regret Bounded by Variation in Costs**
E. Hazan and S. Kale. *Machine Learning Journal* (MLJ), 2010, special issue for COLT 2008. Preliminary version in *Conference on Learning Theory* (COLT), 2008.
- [59] $O(\sqrt{\log n})$ **Approximation to SPARSEST CUT in $\tilde{O}(n^2)$ Time**
S. Arora, E. Hazan, and S. Kale. Accepted to *SIAM Journal on Computing* (SICOMP), 2010. Preliminary version in *Foundations of Computer Science* (FOCS), 2004. conference.
- [60] **Online Submodular Minimization**
E. Hazan and S. Kale. *Neural Information Processing Systems* (NeurIPS), 2009.
- [61] **The Uniform Hardcore Lemma via Approximate Bregman Projections**
B. Barak, M. Hardt, and S. Kale. *Symposium on Discrete Algorithms* (SODA), 2009.
- [62] **Computational Equivalence of Fixed Points and No Regret Algorithms, and Convergence to Equilibria**
E. Hazan and S. Kale. *Conference on Neural Information Processing Systems* (NeurIPS), 2007.
- [63] **A Combinatorial, Primal-Dual approach to Semidefinite Programs**
S. Arora and S. Kale. *Symposium of Theory of Computing* (STOC), 2007.
- [64] **Privacy, Accuracy, and Consistency Too: A Holistic Solution to Contingency Table Release**
B. Barak, K. Chaudhuri, C. Dwork, S. Kale, F. McSherry, and K. Talwar. *Symposium on Principles of Database Systems* (PODS), 2007.
- [65] **Efficient Aggregation Algorithms for Probabilistic Data**
T. S. Jayram, S. Kale, and E. Vee. *Symposium on Discrete Algorithms* (SODA), 2007.
- [66] **A Variation on SVD Based Image Compression**
A. Ranade, S. S. Mahabalarao, and S. Kale. *Image and Vision Computing*, 2007. Preliminary version in *Workshop on Computer Vision, Graphics, and Image Processing* (WCVGIP), 2006.
- [67] **Logarithmic Regret Algorithms for Online Convex Optimization**
E. Hazan, A. Agarwal, and S. Kale. *Machine Learning Journal* (MLJ), 2007, special issue for COLT 2006. Preliminary version with A. Kalai in *Conference on Learning Theory* (COLT), 2006.
- [68] **Algorithms for Portfolio Management Based on the Newton Method**
A. Agarwal, E. Hazan, S. Kale, and R. E. Schapire. *International Conference on Machine Learning* (ICML), 2006.
- [69] **A Fast Random Sampling Algorithm for Sparsifying Matrices**
S. Arora, E. Hazan, and S. Kale. *Workshop on Randomization and Computation* (RANDOM), 2006.
- [70] **Fast Algorithms for Approximate Semidefinite Programming Using the Multiplicative Weights Update method**
S. Arora, E. Hazan, and S. Kale. *Foundations of Computer Science* (FOCS), 2005. conference.
- [71] **Analysis and Algorithms for Content-based Event Matching**
S. Kale, E. Hazan, F. Cao and J. P. Singh. *Workshop on Distributed Event-Based Systems* (DEBS), 2005.

- [1] **On the Unreasonable Effectiveness of Federated Averaging with Heterogeneous Data**
J. Wang, R. Das, G. Joshi, S. Kale, Z. Xu, and T. Zhang. *arXiv preprint arXiv:2206.04723*, 2022
- [2] **Mixed Federated Learning: Joint Decentralized and Centralized Learning**
S. Augenstein, A. Hard, L. Ning, K. Singhal, S. Kale, K. Partridge, and R. Mathews. *arXiv preprint arXiv:2205.13655*, 2022
- [3] **A Multiclass Boosting Framework for Achieving Fast and Provable Adversarial Robustness**
J. Abernethy, P. Awasthi, and S. Kale. *arXiv preprint arXiv:2103.01276*, 2021
- [4] **A Field Guide to Federated Optimization**
J. Wang, Z. Charles, Z. Xu, G. Joshi, H. B. McMahan, M. Al-Shedivat, G. Andrew, S. Avestimehr, K. Daly, D. Data, S. Diggavi, H. Eichner, A. Gadhikar, Z. Garrett, A. M. Girgis, F. Hanzely, A. Hard, C. He, S. Horvath, Z. Huo, A. Ingerman, M. Jaggi, T. Javidi, P. Kairouz, S. Kale, S. P. Karimireddy, J. Konecny, S. Koyejo, T. Li, L. Liu, M. Mohri, H. Qi, S. J Reddi, P. Richtarik, K. Singhal, V. Smith, M. Soltanolkotabi, W. Song, A. T. Suresh, S. U. Stich, A. Talwalkar, H. Wang, B. Woodworth, S. Wu, F. X. Yu, H. Yuan, M. Zaheer, M. Zhang, T. Zhang, C. Zheng, C. Zhu, and W. Zhu. *arXiv preprint arXiv:2107.06917*, 2021
- [5] **Efficient Algorithms Using the Multiplicative Weights Update Method**
S. Kale. Ph.D. Thesis. Princeton Technical Report TR-804-07, 2007.
- [6] **Approximating Quadratic Programs with Positive Semidefiniteness Constraints**
E. Hazan and S. Kale. Princeton Technical Report TR-746-06, 2004.

PATENTS

Adaptive Optimization with Improved Convergence

S. J. Reddi, S. Kumar, S. C. Kale. U.S. Patent 11,586,904.

Controlled adaptive optimization

S. J. Reddi, S. Kumar, M. Zaheer, S. C. Kale. U.S. Patent 10,769,529.

Systems and methods for evaluating a loss function or a gradient of a loss function via dual decomposition

S. C. Kale, D. Holtmann-Rice, S. Kumar, E. Yan, X. Yu. U.S. Patent 10,510,021.

Message classification

D. Pal, S. C. Kale, Y. Xi, I. Kanniah, Y. Peduel, Z. S. Karnin, J. S. Shue. U.S. Patent 10,594,640.

Methods and Systems for Ad Placement Planning

E. Vee, D. Chen, P. Chen, S. Kale, S. Mandalapu, and C. Nagarajan. U.S. Patent 9,135,632

Consistent Contingency Table Release

B. Barak, K. Chaudhuri, C. Dwork, S. Kale, F. McSherry, and K. Talwar. U.S. Patent Application 11,972,618.

TEACHING EXPERIENCE

Lecturer for COMS 4995-004: *Optimization for Machine Learning*, Fall 2019
Computer Science Department, Columbia University, New York, NY.

Lecturer for COMS 4771: *Introduction to Machine Learning*, Spring 2016
Computer Science Department, Columbia University, New York, NY.

Guest lecturer for COS 521: *Advanced Algorithm Design*, Fall 2005, 2006
Computer Science Department, Princeton University, Princeton, NJ.

Teaching Assistant for COS 226: *Algorithms and Data Structures*, Fall 2004
Computer Science Department, Princeton University, Princeton, NJ.

Teaching Assistant for COS 341: *Discrete Mathematics*, Fall 2003
Computer Science Department, Princeton University, Princeton, NJ.

HONORS

Best student paper award at Conference on Learning Theory, 2018
“Logistic Regression: The Importance of Being Improper” awarded best student paper at COLT 2018.

Best paper award at International Conference on Learning Representations, 2018
“On the convergence of Adam and Beyond” awarded best paper at ICLR 2018.

Best paper award at International Conference on Machine Learning, 2015
“Optimal and Adaptive Algorithms for Online Boosting” awarded best paper at ICML 2015.

Bronze medal in 37th International Mathematics Olympiad, 1997
Represented India at the IMO in Mar Del Plata, Argentina.

Scholarship in National Talent Search Examination, 1996
One of 750 students are awarded scholarships every year all over India.

PROFESSIONAL SERVICE

Ph.D. committee member: Giulia DeSalvo (New York University, 2017), Scott Yang (New York University, 2017), Haipeng Luo (Princeton University, 2016), Qiang Ma (Rutgers University, 2016)

Program committee member: International Conference on Machine Learning (ICML) 2009, 2012, (area chair), 2013 (area chair), 2016 (area chair), 2019 (area chair), 2020 (area chair) and 2021 (area chair); Neural Information Processing Systems (NeurIPS) 2018 (senior area chair), 2019 (senior area chair) and 2021 (senior area chair), 2022 (senior area chair), and 2023 (area chair); Conference on Learning Theory (COLT) 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020 (area chair), 2021 (area chair), 2022 (area chair), and 2023 (area chair); Algorithmic Learning Theory (ALT) 2020, 2021, 2022, 2023, and 2024; Symposium on Discrete Algorithms (SODA) 2011, and Conference on Knowledge Discovery and Mining (SIGKDD) 2010.

Conference organization: Served as ALT 2019 program chair, and COLT 2017 program chair. Served as COLT sponsorship chair in 2018, 2019, and 2020. Served as COLT Steering Committee Member (2019-2022) and ALT Steering Committee Member (2019-2023). Served as COLT 2016

local chair and COLT 2015 publications chair. Organized DIMACS/TRIPODS workshop on optimization in Machine Learning at Lehigh University in 2018. Organized sessions on *Online Learning* at INFORMS 2009 and *Semidefinite Programming and Applications* at INFORMS 2008.

Editorship and reviewer for journals and conferences: Associate Editor of Mathematics of Operations Research (MOR) journal. Reviewer for Journal of the ACM (JACM), SIAM Journal on Computing (SICOMP), Journal of Machine Learning Research (JMLR), Machine Learning Journal (MLJ), Algorithmica, Theoretical Computer Science journal (TCS), Transactions on Algorithms (TALG), and Statistics and Computing journal (STCO), Foundations of Computer Science (FOCS), Symposium on Theory of Computation (STOC), Symposium on Discrete Algorithms (SODA), Conference on Learning Theory (COLT), Innovations in Theoretical Computer Science (ITCS), International Conference on Machine Learning (ICML), Neural Information Processing Systems (NeurIPS), ACM Conference on Knowledge Discovery and Mining (SIGKDD), Conference on Uncertainty in Artificial Intelligence (UAI), Conference on Integer Programming and Combinatorial Optimization (IPCO), International Conference on Very Large Data Bases (VLDB), Symposium on Theoretical Aspects of Computer Science (STACS), International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX), International Workshop on Randomization and Computation (RANDOM), and International Symposium on Information Theory (ISIT).