Jason D. Lee

Curriculum Vitae

✉ *email: jasonlee@marshall.usc.edu*
http://www-bcf.usc.edu/~lee715/

Research Interests

Machine Learning: Provable methods for deep learning and non-convex problems, Distributed machine learning algorithms

Large-scale Optimization: Distributed optimization, Non-convex optimization, Stochastic optimization

Statistics: Post-model selection inference, High-dimensional estimation and inference

Employment

2019–Present **Assistant Professor**, *Electrical Engineering and Associated Faculty in Computer Science*, Princeton University.

2016–2019 **Assistant Professor**, *Data Science and Operations Department*, University of Southern California.

2015–2016 **Postdoctoral Scholar**, *Computer Science*, UC Berkeley.
– Adviser: Michael I. Jordan

Education

2010–2015 **Stanford University, Computational and Mathematical Engineering**.
Ph.D. Candidate
– Advisers: Trevor Hastie and Jonathan Taylor

2006–2010 **Duke University, Department of Mathematics**.
Bachelor of Science in Mathematics
Graduation with High Distinction, Phi Beta Kappa, Magna Cum Laude
– Adviser: Mauro Maggioni

Awards and Honors

2019 **Sloan Research Fellowship in Computer Science**.

2018 **ICML Workshop on Nonconvex Optimization Best Paper Award for "Algorithmic Regularization in Learning Deep Homogeneous Models"**.

2016 **NIPS Best Student Paper Award for "Matrix Completion has no Spurious Local Minima"**.

2010 **National Defense Science and Engineering Graduate Fellowship, Stanford Graduate Fellowship, and NSF Graduate Fellowship**.

2010 **Julia Dale Prize in Mathematics**, *Duke University*.
Awarded to top graduate of Duke Mathematics Department

Refereed Conference Publications

[1] Ruiqi Gao, Tianle Cai, Haochuan Li, Liwei Wang, Cho-Jui Hsieh, and Jason D Lee. Convergence of Adversarial Training in Overparametrized Networks. *Neural Information Processing Systems (NeurIPS)*, 2019.

[2] Qi Cai, Zhuoran Yang, Jason D Lee, and Zhaoran Wang. Neural Temporal-Difference Learning Converges to Global Optima. *Neural Information Processing Systems (NeurIPS)*, 2019.

[3] Colin Wei, Jason D Lee, Qiang Liu, and Tengyu Ma. Regularization Matters: Generalization and Optimization of Neural Nets v.s. their Induced Kernel. *Neural Information Processing Systems (NeurIPS)*, 2019.

[4] Maher Nouiehed, Maziar Sanjabi, Tianjian Huang, Jason D Lee, and Meisam Razaviyayn. Solving a class of non-convex min-max games using iterative first order methods. *Neural Information Processing Systems (NeurIPS)*, 2019.

[5] Simon S Du, Jason D Lee, Haochuan Li, Liwei Wang, and Xiyu Zhai. Gradient Descent Finds Global Minima of Deep Neural Networks. *International Conference on Machine Learning (ICML)*, 2019.

[6] Mor Shpigel Nacson, Jason D. Lee, Suriya Gunasekar, Nathan Srebro, and Daniel Soudry. Convergence of Gradient Descent on Separable Data. *Artificial Intelligence and Statistics (AISTATS)*, 2019.

[7] Sham Kakade and Jason D Lee. Provably Correct Automatic Subdifferentiation for Qualified Programs. *Neural Information Processing Systems (NIPS)*, 2018.

[8] Maziar Sanjabi, Jimmy Ba, Meisam Razaviyayn, and Jason D Lee. Solving Approximate Wasserstein GANs to Stationarity. *Neural Information Processing Systems (NIPS)*, 2018.

[9] Shiyu Liang, Ruoyu Sun, Jason D Lee, and R Srikant. Adding One Neuron Can Eliminate All Bad Local Minima. *Neural Information Processing Systems (NIPS)*, 2018.

[10] Simon S Du, Wei Hu, and Jason D Lee. Algorithmic Regularization in Learning Deep Homogeneous Models: Layers are Automatically Balanced. *Neural Information Processing Systems (NIPS)*, 2018.

[11] Suriya Gunasekar, Jason Lee, Daniel Soudry, and Nathan Srebro. Implicit Bias of Gradient Descent on Linear Convolutional Networks. *Neural Information Processing Systems (NIPS)*, 2018.

[12] Simon S Du, Jason D Lee, Yuandong Tian, Barnabas Poczos, and Aarti Singh. Gradient Descent Learns One-hidden-layer CNN: Don't be Afraid of Spurious Local Minima. *International Conference on Machine Learning (ICML)*, 2018.

[13] Simon S Du and Jason D Lee. On the Power of Over-parametrization in Neural Networks with Quadratic Activation. *International Conference on Machine Learning (ICML)*, 2018.

[14] Suriya Gunasekar, Jason Lee, Daniel Soudry, and Nathan Srebro. Characterizing Implicit Bias in Terms of Optimization Geometry. *International Conference on Machine Learning (ICML)*, 2018.

[15] Mingyi Hong, Jason D Lee, and Meisam Razaviyayn. Gradient Primal-Dual Algorithm Converges to Second-Order Stationary Solutions for Nonconvex Distributed Optimization. *International Conference on Machine Learning (ICML)*, 2018.

[16] Rong Ge, Jason D Lee, and Tengyu Ma. Learning One-hidden-layer Neural Networks with Landscape Design. *International Conference on Learning Representations (ICLR)*, 2018.

[17] Simon S Du, Jason D Lee, and Yuandong Tian. When is a Convolutional Filter Easy to Learn? *International Conference on Learning Representations (ICLR)*, 2018.

[18] Simon S Du, Chi Jin, Jason D Lee, Michael I Jordan, Aarti Singh, and Barnabas Poczos. Gradient Descent Can Take Exponential Time to Escape Saddle Points. *Neural Information Processing Systems (NIPS)*, 2017.

[19] Jialei Wang, Jason D Lee, Mehrdad Mahdavi, Mladen Kolar, and Nathan Srebro. Sketching Meets Random Projection in the Dual: A Provable Recovery Algorithm for Big and High-dimensional Data. *Artificial Intelligence and Statistics (AISTATS)*, 2017.

[20] Qiang Liu and Jason D Lee. Black-box importance sampling. *Artificial Intelligence and Statistics (AISTATS)*, 2017.

[21] Yuchen Zhang, Jason D. Lee, Martin J Wainwright, and Michael I Jordan. Learning Halfspaces and Neural Networks with Random Initialization. *Artificial Intelligence and Statistics (AISTATS)*, 2017.

[22] Rong Ge, Jason D. Lee, and Tengyu Ma. Matrix Completion has No Spurious Local Minimum. *Neural Information Processing Systems (NIPS)*, 2016.

[23] Jason D. Lee, Max Simchowitz, Michael I Jordan, and Benjamin Recht. Gradient Descent Converges to Minimizers. *Conference on Learning Theory (COLT)*, 2016.

[24] Yuchen Zhang, Jason D. Lee, and Michael I Jordan. l1-regularized Neural Networks are Improperly Learnable in Polynomial Time. *International Conference on Machine Learning (ICML)*, 2016.

[25] Qiang Liu, Jason D. Lee, and Michael I Jordan. A Kernelized Stein Discrepancy for Goodness-of-fit Tests and Model Evaluation. *International Conference on Machine Learning (ICML)*, 2016.

[26] Jason D. Lee, Yuekai Sun, and Jonathan E. Taylor. Evaluating the Statistical Significance of Biclusters. *Neural Information Processing Systems (NIPS)*, 2015, pp. 1–9.

[27] Austin R Benson, Jason D. Lee, Bartek Rajwa, and David F. Gleich. Scalable Methods for Nonnegative Matrix Factorizations of Near-Separable Tall-and-Skinny Matrices. *Neural Information Processing Systems (NIPS)*, 2014, pp. 1–9.

[28] Jason D. Lee and Jonathan E. Taylor. Exact Post Model Selection Inference for Marginal Screening. *Neural Information Processing Systems (NIPS)*, 2014, pp. 1–9.

[29] Jason D. Lee, Ran Gilad-Bachrach, and Rich Caruana. Using Multiple Samples to Learn Mixture Models. *Neural Information Processing Systems (NIPS)*, 2013, pp. 324–332.

[30] Jason D. Lee, Yuekai Sun, and Jonathan E. Taylor. On Model Selection Consistency of Penalized M-Estimators: a Geometric Theory. *Neural Information Processing Systems (NIPS)*, 2013, pp. 342–350.

[31] Jason D. Lee and Trevor Hastie. Structure Learning of Mixed Graphical Models. *Artificial Intelligence and Statistics (AISTATS)*, 2013, pp. 388–396.

[32] Jason D. Lee, Yuekai Sun, and Michael Saunders. Proximal Newton-type Methods for Convex Optimization. *Neural Information Processing Systems (NIPS)*, 2012, pp. 836–844.

[33] Jason D. Lee, Ben Recht, Nathan Srebro, Joel Tropp, and Ruslan Salakhutdinov. Practical Large-Scale Optimization for Max-Norm Regularization. *Neural Information Processing Systems (NIPS)*, 2010, pp. 1297–1305.

[34] Jason D. Lee and Mauro Maggioni. Multiscale Analysis of Time Series of Graphs. *International Conference on Sampling Theory and Applications (SAMPTA)*, 2011.

[35] Markus Kliegl, Jason D. Lee, Jun Li, Xinchao Zhang, Chuanxiong Guo, and David Rincón. Generalized DCell Structure for Load-Balanced Data Center Networks. *IEEE Conference on Computer Communications (INFOCOM)*, 2010, pp. 1–5.

[36] Anna V. Little, Jason D. Lee, Yoon-Mo Jung, and Mauro Maggioni. Estimation of Intrinsic Dimensionality of Samples from Noisy Low-Dimensional Manifolds in High Dimensions with Multiscale SVD. *IEEE Workshop on Statistical Signal Processing (SSP)*, 2009, pp. 85–88.

Journal Articles

[1] Xi Chen, Jason D Lee, Xin T Tong, and Yichen Zhang. Statistical Inference for Model Parameters in Stochastic Gradient Descent. *Accepted at Annals of Statistics*.

[2] Jason D Lee, Ioannis Panageas, Georgios Piliouras, Max Simchowitz, Michael I Jordan, and Benjamin Recht. First-order Methods Almost Always Avoid Saddle Points. *Accepted at Math Programming*, 2018.

[3] Damek Davis, Dmitriy Drusvyatskiy, Sham Kakade, and Jason D Lee. Stochastic subgradient method converges on tame functions. *Foundations of Computational Mathematics*, 2018.

[4] Mahdi Soltanolkotabi, Adel Javanmard, and Jason D Lee. Theoretical insights into the optimization landscape of over-parameterized shallow neural networks. *Transactions on Information Theory*, 2018.

[5] Michael I Jordan, Jason D. Lee, and Yun Yang. Communication-efficient distributed statistical learning. *Journal of the American Statistics Association*, 2018.

[6] Jason D. Lee, Qiang Liu, Yuekai Sun, and Jonathan E. Taylor. Communication-Efficient Distributed Sparse Regression. *Journal of Machine Learning Research*, 2017.

[7] Jialei Wang, Jason D Lee, Mehrdad Mahdavi, Mladen Kolar, and Nathan Srebro. Sketching Meets Random Projection in the Dual: A Provable Recovery Algorithm for Big and High-dimensional Data. *Electronic Journal of Statistics*, 2017.

[8] Jason D. Lee, Tengyu Ma, Qihang Lin, and Tianbao Yang. Distributed Stochastic Variance Reduced Gradient Methods. *Journal of Machine Learning Research*, 2017.

[9] Jason D. Lee, Dennis L. Sun, Yuekai Sun, and Jonathan E. Taylor. Exact Inference after Model Selection via the Lasso. *Annals of Statistics*, 2016.

[10] Trevor Hastie, Rahul Mazumder, Jason D. Lee, and Reza Zadeh. Matrix Completion and Low-Rank SVD via Fast Alternating Least Squares. *Journal of Machine Learning Research*, 2015.

[11] Jason D. Lee, Yuekai Sun, and Jonathan E. Taylor. On Model Selection Consistency of Regularized M-Estimators. *Electronic Journal of Statistics*, 2015.

[12] Jason D. Lee, Yuekai Sun, and Michael Saunders. Proximal Newton-Type Methods for Minimizing Composite Functions. *SIAM Journal on Optimization*, 2014.

[13] Jason D. Lee and Trevor J Hastie. Learning the Structure of Mixed Graphical Models. *Journal of Computational and Graphical Statistics*, 2014.

Under Review

[1] Qi Lei, Jason D Lee, Alexandros G Dimakis, and Constantinos Daskalakis. SGD Learns One-Layer Networks in WGANs. *arXiv preprint arXiv:1910.07030*, 2019.

[2] Alekh Agarwal, Sham M Kakade, Jason D Lee, and Gaurav Mahajan. Optimality and Approximation with Policy Gradient Methods in Markov Decision Processes. *arXiv preprint arXiv:1908.00261*, 2019.

[3] Yu Bai and Jason D Lee. Beyond Linearization: On Quadratic and Higher-Order Approximation of Wide Neural Networks. *arXiv preprint arXiv:1910.01619*, 2019.

[4] Blake Woodworth, Suriya Gunasekar, Jason Lee, Daniel Soudry, and Nathan Srebro. Kernel and Deep Regimes in Overparametrized Models. *arXiv preprint arXiv:1906.05827*, 2019.

[5] Ashok Vardhan Makkuva, Amirhossein Taghvaei, Sewoong Oh, and Jason D Lee. Optimal transport mapping via input convex neural networks. *arXiv preprint arXiv:1908.10962*, 2019.

[6] Xuanqing Liu, Cho-Jui Hsieh, Jason D Lee, and Yuekai Sun. An inexact subsampled proximal Newton-type method for large-scale machine learning. *Submitted to Journal of Machine Learning Research*.

[7] Adel Javanmard and Jason D Lee. A Flexible Framework for Hypothesis Testing in High-dimensions. *Submitted to Journal of the Royal Statistical Society Series B*.

[8] Maher Nouiehed, Jason D Lee, and Meisam Razaviyayn. Convergence to Second-Order Stationarity for Constrained Non-Convex Optimization. *Submitted to SIAM Journal on Optimization*, 2018.

[9] Xiao Li, Zhihui Zhu, Anthony Man-Cho So, and Jason D. Lee. Incremental (Sub)-Gradient Descent for Weakly Convex Optimization. *Submitted to SIOPT*, 2019.

Technical Reports

[1] Chenwei Wu, Jiajun Luo, and Jason D Lee. No Spurious Local Minima in a Two Node Neural Network. *International Conference on Learning Representations (ICLR) Workshop Track*, 2018.

[2] Jason D. Lee, Yuekai Sun, and Michael A Saunders. Convergence Analysis of Inexact Proximal Newton-Type Methods. *NIPS Workshop on Optimization in Machine Learning*, 2012.

[3] Jason D. Lee. Multiscale Estimation of Intrinsic Dimensionality of Point Cloud Data and Multiscale Analysis of Dynamic Graphs. *Senior Thesis, Duke University*, 2010.

[4] Markus Kliegl, Jason D. Lee, Jun Li, Xinchao Zhang, Chuanxiong Guo, and David Rincón. Generalized DCell Structure for Load-Balanced Data Center Networks. *Microsoft Research Technical Report*, 2009, pp. 1–14. URL: `http://research.microsoft.com/apps/pubs/default.aspx?id=103129`.

[5] Jason D. Lee and John Neuberger. Existence of Asymptotic Solutions to Semi-linear Partial Difference Equations. *Joint Mathematics Meetings*, 2008.

Software

fastALS, *Spark implementation for* Matrix Completion and Low-Rank SVD via Fast Alternating Least Squares , Trevor Hastie, Rahul Mazumder, Jason D. Lee, and Reza Zadeh, `http://git.io/sparkfastals`.

PNOPT, *Package for* Proximal Newton-type methods for non-smooth optimization, Jason D. Lee, Michael Saunders, and Yuekai Sun, `https://github.com/yuekai/PNopt`.

Nonnegative matrix factorizations in MapReduce, *Hadoop implementation of* Scalable methods for nonnegative matrix factorizations of near-separable tall-and-skinny matrices, Austin R. Benson, Jason D. Lee, Bartek Rajwa, and David F. Gleich, `https://github.com/arbenson/mrnmf`.

Gradient and Newton methods for structure learning of mixed graphical models, *Matlab implementation for* structure learning of mixed graphical models , Jason D. Lee and Trevor Hastie, `http://web.stanford.edu/~jdl17/learningmgm.html`.

Teaching Experience

Spring 2018 **Instructor**, *Statistical Machine Learning*, University of Southern California.

Fall 2016, Spring 2018 **Instructor**, *Applied Business Statistics*, University of Southern California.

Winter 2013 **Instructor**, *Introduction to Scientific Computing with MATLAB (CME 192)*, Stanford University.

Selected Talks

2018 **Towards the Understanding of Overparametrization in Deep Learning**, *CIMI Workshop in Toulouse, France, University of Wisconsin Madison Tripods Workshop, ICML 2018 Workshop on the Theory of Deep Learning, Conference on Statistical Learning and Data Science 2018, UT Austin CS seminar, Facebook AI Research, Informs Conference on Optimization, and Information Theory and its Applications 2018.*

2017-2018 **Landscape Design for Deep Learning**, *Army Research Lab, Conference on Nonconvex Statistical Learning, Fudan University, Asilomar 2017, Foundations of Computational Math 2017, Georgia Tech Isye Seminar, Caltech CMS Seminar.*

2016-2017 **Tractable Non-Convexity:Matrix Completion, Saddlepoints, and Gradient Descent**, *Information Theory and its Applications 2017, SIAM Conference on Optimization 2017, Informs 2017, UC Irvine Statistics Seminar, UCLA Statistics Seminar, USC Math seminar, COLT 2016, University of Miami seminar, .*

Reviewing

Reviewer for ICML, AISTATS, and NIPS conferences.

Reviewer for Journal of Machine Learning Research, AOS, Math Programming, and JRSS-B.