Jason D. Lee

Curriculum Vitae

Link to Google Scholar Profile.
email: jasonlee@princeton.edu
https://jasondlee88.github.io/

Research Interests

Machine Learning Theory and Methods: provable and practical methods for deep learning, reinforcement learning, domain adaptation, representation learning, uncertainty quantification, and the understanding/improvement of machine learning heuristics.

Optimization and Algorithm Design: non-convex optimization, multi-agent learning, and distributed optimization.

Algorithmic Statistics: Selective Inference, High-dimensional estimation and inference

Employment

2019–Present **Associate Professor**, *Electrical Engineering and Computer Science (Secondary)*, Princeton University

2019–2020 **Member**, *School of Mathematics*, Institute for Advanced Study

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.
– 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

2022 **NSF Career Award**

2021 **ONR Young Investigator Award for Mathematical Data Science**

2020 **Princeton Commendation for Excellent Teaching for ECE538B**

2019 **Sloan Research Fellowship in Computer Science**

2019 **Finalist for Best Paper Prize for Young Researchers in Continuous Optimization**

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

Representative Publications

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

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

[3] 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.

[4] Alekh Agarwal, Sham M Kakade, Jason D Lee, and Gaurav Mahajan. On the Theory of Policy Gradient Methods. *JMLR (short version at COLT)*, 2020.

[5] Simon S Du, Sham M Kakade, Jason D Lee, Shachar Lovett, Gaurav Mahajan, Wen Sun, and Ruosong Wang. Bilinear Classes: A Structural Framework for Provable Generalization in RL. *International Conference on Machine Learning (ICML)*, 2021.

[6] Alex Damian, Tengyu Ma, and Jason D. Lee. Label Noise SGD Provably Prefers Flat Global Minimizers. *Neural Information Processing Systems (NeurIPS)*, 2021.

[7] Jason D Lee, Qi Lei, Nikunj Saunshi, and Jiacheng Zhuo. Predicting what you already know helps: Provable self-supervised learning. *Neural Information Processing Systems (NeurIPS)*, 2021.

[8] Alexandru Damian, Jason Lee, and Mahdi Soltanolkotabi. "Neural networks can learn representations with gradient descent". *Conference on Learning Theory*. PMLR. 2022, pp. 5413–5452.

Refereed Conference Publications

[1] Christopher De Sa, Satyen Kale, Jason D Lee, Ayush Sekhari, and Karthik Sridharan. From Gradient Flow on Population Loss to Learning with Stochastic Gradient Descent. *Neural Information Processing Systems (NeurIPS)*, 2022.

[2] Eshaan Nichani, Yu Bai, and Jason D Lee. Identifying good directions to escape the NTK regime and efficiently learn low-degree plus sparse polynomials. *Neural Information Processing Systems (NeurIPS)*, 2022.

[3] Zhiyuan Li, Tianhao Wang, JasonD Lee, and Sanjeev Arora. Implicit Bias of Gradient Descent on Reparametrized Models: On Equivalence to Mirror Descent. *Neural Information Processing Systems (NeurIPS)*, 2022.

[4] Masatoshi Uehara, Ayush Sekhari, Jason D Lee, Nathan Kallus, and Wen Sun. Provably efficient reinforcement learning in partially observable dynamical systems. *Neural Information Processing Systems (NeurIPS)*, 2022.

[5] Alexandru Damian, Jason Lee, and Mahdi Soltanolkotabi. "Neural networks can learn representations with gradient descent". *Conference on Learning Theory*. PMLR. 2022, pp. 5413–5452.

[6] Itay Safran and Jason Lee. "Optimization-based separations for neural networks". *Conference on Learning Theory*. PMLR. 2022, pp. 3–64.

[7] Itay Safran, Gal Vardi, and Jason D Lee. On the Effective Number of Linear Regions in Shallow Univariate ReLU Networks: Convergence Guarantees and Implicit Bias. *Neural Information Processing Systems (NeurIPS)*, 2022.

[8] Jiaqi Yang, Qi Lei, Jason D Lee, and Simon S Du. Nearly minimax algorithms for linear bandits with shared representation. *arXiv preprint arXiv:2203.15664*, 2022.

[9] Wenhao Zhan, Baihe Huang, Audrey Huang, Nan Jiang, and Jason Lee. "Offline reinforcement learning with realizability and single-policy concentrability". *Conference on Learning Theory*. PMLR. 2022, pp. 2730–2775.

[10] Baihe Huang, Jason D Lee, Zhaoran Wang, and Zhuoran Yang. Towards General Function Approximation in Zero-Sum Markov Games. *ICLR*, 2022.

[11] Yulai Zhao, Yuandong Tian, Jason D Lee, and Simon S Du. Provably efficient policy gradient methods for two-player zero-sum Markov games. *AISTATS*, 2022.

[12] DiJia Su, Jason D Lee, John M Mulvey, and H Vincent Poor. MUSBO: Model-based Uncertainty Regularized and Sample Efficient Batch Optimization for Deployment Constrained Reinforcement Learning. *ICASSP*, 2022.

[13] Kurtland Chua, Qi Lei, and Jason D Lee. How fine-tuning allows for effective meta-learning. *Neural Information Processing Systems (NeurIPS)*, 2021.

[14] Alex Damian, Tengyu Ma, and Jason D. Lee. Label Noise SGD Provably Prefers Flat Global Minimizers. *Neural Information Processing Systems (NeurIPS)*, 2021.

[15] Jason D Lee, Qi Lei, Nikunj Saunshi, and Jiacheng Zhuo. Predicting what you already know helps: Provable self-supervised learning. *Neural Information Processing Systems (NeurIPS)*, 2021.

[16] Baihe Huang, Kaixuan Huang, Sham M Kakade, Jason D Lee, Qi Lei, Runzhe Wang, and Jiaqi Yang. Going Beyond Linear RL: Sample Efficient Neural Function Approximation. *Neural Information Processing Systems (NeurIPS)*, 2021.

[17] Baihe Huang, Kaixuan Huang, Sham M Kakade, Jason D Lee, Qi Lei, Runzhe Wang, and Jiaqi Yang. Optimal Gradient-based Algorithms for Non-concave Bandit Optimization. *Neural Information Processing Systems (NeurIPS)*, 2021.

[18] Yu Bai, Minshuo Chen, Pan Zhou, Tuo Zhao, Jason Lee, Sham Kakade, Huan Wang, and Caiming Xiong. How Important is the Train-Validation Split in Meta-Learning? *International Conference on Machine Learning (ICML)*, 2021.

[19] Qi Lei, Wei Hu, and Jason Lee. Near-Optimal Linear Regression under Distribution Shift. *International Conference on Machine Learning (ICML)*, 2021.

[20] Tianle Cai, Ruiqi Gao, Jason D Lee, and Qi Lei. A Theory of Label Propagation for Subpopulation Shift. *International Conference on Machine Learning (ICML)*, 2021.

[21] Simon S Du, Sham M Kakade, Jason D Lee, Shachar Lovett, Gaurav Mahajan, Wen Sun, and Ruosong Wang. Bilinear Classes: A Structural Framework for Provable Generalization in RL. *International Conference on Machine Learning (ICML)*, 2021.

[22] Jeff Z. HaoChen, Colin Wei, Jason D. Lee, and Tengyu Ma. Shape Matters: Understanding the Implicit Bias of the Noise Covariance. *Conference on Learning Theory (COLT 2021)*, 2021.

[23] Cong Fang, Jason D Lee, Pengkun Yang, and Tong Zhang. Modeling from Features: a Mean-field Framework for Over-parameterized Deep Neural Networks. *Conference on Learning Theory (COLT)*, 2021.

[24] Simon S Du, Wei Hu, Sham M Kakade, Jason D Lee, and Qi Lei. Few-shot learning via learning the representation, provably. *International Conference on Learning Representations (ICLR)*, 2021.

[25] Jiaqi Yang, Wei Hu, Jason D Lee, and Simon S Du. Provable Benefits of Representation Learning in Linear Bandits. *International Conference on Learning Representations (ICLR)*, 2021.

[26] Simon S Du, Jason D Lee, Gaurav Mahajan, and Ruosong Wang. Agnostic Q-learning with function approximation in deterministic systems: Tight bounds on approximation error and sample complexity. *Neural Information Processing Systems (NeurIPS)*, 2020.

[27] Edward Moroshko, Suriya Gunasekar, Blake Woodworth, Jason D Lee, Nathan Srebro, and Daniel Soudry. Implicit Bias in Deep Linear Classification: Initialization Scale vs Training Accuracy. *Neural Information Processing Systems (NeurIPS)*, 2020.

[28] Minshuo Chen, Yu Bai, Jason D Lee, Tuo Zhao, Huan Wang, Caiming Xiong, and Richard Socher. Towards Understanding Hierarchical Learning: Benefits of Neural Representations. *Neural Information Processing Systems (NeurIPS)*, 2020.

[29] Kaiyi Ji, Jason D Lee, Yingbin Liang, and H Vincent Poor. Convergence of Meta-Learning with Task-Specific Adaptation over Partial Parameters. *Neural Information Processing Systems (NeurIPS)*, 2020.

[30] Jingtong Su, Yihang Chen, Tianle Cai, Tianhao Wu, Ruiqi Gao, Liwei Wang, and Jason D Lee. Sanity-Checking Pruning Methods: Random Tickets can Win the Jackpot. *Neural Information Processing Systems (NeurIPS)*, 2020.

[31] Xiang Wang, Chenwei Wu, Jason D Lee, Tengyu Ma, and Rong Ge. Beyond Lazy Training for Over-parameterized Tensor Decomposition. *Neural Information Processing Systems (NeurIPS)*, 2020.

[32] Jason D Lee, Ruoqi Shen, Zhao Song, Mengdi Wang, et al. Generalized Leverage Score Sampling for Neural Networks. *Neural Information Processing Systems (NeurIPS)*, 2020.

[33] Yihong Gu, Weizhong Zhang, Cong Fang, Jason D Lee, and Tong Zhang. How to Characterize The Landscape of Overparameterized Convolutional Neural Networks. *Neural Information Processing Systems (NeurIPS)*, 2020.

[34] Qi Lei, Jason D Lee, Alexandros G Dimakis, and Constantinos Daskalakis. SGD Learns One-Layer Networks in WGANs. *International Conference on Machine Learning (ICML)*, 2020.

[35] Alekh Agarwal, Sham M Kakade, Jason D Lee, and Gaurav Mahajan. Optimality and Approximation with Policy Gradient Methods in Markov Decision Processes. *Conference on Learning Theory (COLT)*, 2020.

[36] Yu Bai and Jason D Lee. Beyond Linearization: On Quadratic and Higher-Order Approximation of Wide Neural Networks. *International Conference on Learning Representations (ICLR)*, 2020.

[37] Blake Woodworth, Suriya Gunasekar, Jason Lee, Daniel Soudry, and Nathan Srebro. Kernel and Deep Regimes in Overparametrized Models. *Conference on Learning Theory (COLT)*, 2020.

[38] Ashok Vardhan Makkuva, Amirhossein Taghvaei, Sewoong Oh, and Jason D Lee. Optimal transport mapping via input convex neural networks. *International Conference on Machine Learning (ICML)*, 2020.

[39] 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.

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

[41] 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.

[42] 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.

[43] Mor Sphigel Nacson, Suriya Gunasekar, Jason D Lee, Nathan Srebro, and Soudry Daniel. Lexicographic and Depth-Sensitive Margins in Homogeneous and Non-Homogeneous Training. *International Conference on Machine Learning (ICML)*, 2019.

[44] 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.

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

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

[47] 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.

[48] 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.

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

[50] 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.

[51] 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.

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

[53] 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.

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

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

[56] 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.

[57] 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.

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

[59] 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.

[60] 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.

[61] 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.

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

[63] 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.

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

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

[66] 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.

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

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

[69] 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.

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

[71] 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.

[72] 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.

Journal Articles

[1] Qi Cai, Zhuoran Yang, Jason D Lee, and Zhaoran Wang. Neural Q-Learning Converges to Global Optima. *Mathematics of Operations Research*, 2022.

[2] Songtao Lu, Jason D. Lee, Meisam Razaviyayn, and Mingyi Hong. Linearized ADMM Converges to Second-Order Stationary Points for Non-Convex Problems. *IEEE Transactions on Signal Processing*, 2021.

[3] Xi Chen, Jason D Lee, He Li, and Yun Yang. Distributed Estimation for Principal Component Analysis: a Gap-free Approach. *Journal of the American Statistical Association*, 2021.

[4] Alekh Agarwal, Sham M Kakade, Jason D Lee, and Gaurav Mahajan. On the Theory of Policy Gradient Methods. *JMLR (short version at COLT)*, 2020.

[5] Adel Javanmard and Jason D Lee. A Flexible Framework for Hypothesis Testing in High-dimensions. *Accepted Journal of the Royal Statistical Society Series B*, 2019.

[6] Xi Chen, Jason D Lee, Xin T Tong, Yichen Zhang, et al. Statistical inference for model parameters in stochastic gradient descent. *The Annals of Statistics* 48.1, 2020.

[7] 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.

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

[9] 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.

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

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

[12] 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.

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

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

[15] 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.

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

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

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

Under Review

[1] Alex Damian, Eshaan Nichani, and Jason D Lee. Self-Stabilization: The Implicit Bias of Gradient Descent at the Edge of Stability. *arXiv preprint arXiv:2209.15594*, 2022.

[2] Wenhao Zhan, Jason D Lee, and Zhuoran Yang. Decentralized Optimistic Hyperpolicy Mirror Descent: Provably No-Regret Learning in Markov Games. *arXiv preprint arXiv:2206.01588*, 2022.

[3] Kurtland Chua, Qi Lei, and Jason D Lee. Provable Hierarchy-Based Meta-Reinforcement Learning. *arXiv preprint arXiv:2110.09507*, 2021.

[4] Wenhao Zhan, Masatoshi Uehara, Wen Sun, and Jason D Lee. Pac reinforcement learning for predictive state representations. *arXiv preprint arXiv:2207.05738*, 2022.

[5] Masatoshi Uehara, Ayush Sekhari, Jason D Lee, Nathan Kallus, and Wen Sun. Computationally Efficient PAC RL in POMDPs with Latent Determinism and Conditional Embeddings. *arXiv preprint arXiv:2206.12081*, 2022.

[6] Xinyi Chen, Edgar Minasyan, Jason D Lee, and Elad Hazan. Provable Regret Bounds for Deep Online Learning and Control. *arXiv preprint arXiv:2110.07807*, 2022.

[7] Wenhao Zhan, Shicong Cen, Baihe Huang, Yuxin Chen, Jason D. Lee, and Yuejie Chi. *Policy Mirror Descent for Regularized Reinforcement Learning: A Generalized Framework with Linear Convergence*. 2021. arXiv: `2105.11066 [cs.LG]`.

[8] Kaixuan Huang, Sham M Kakade, Jason D Lee, and Qi Lei. A Short Note on the Relationship of Information Gain and Eluder Dimension. *arXiv preprint arXiv:2107.02377*, 2021.

[9] Lemeng Wu, Mao Ye, Qi Lei, Jason D Lee, and Qiang Liu. Steepest Descent Neural Architecture Optimization: Escaping Local Optimum with Signed Neural Splitting. *arXiv preprint arXiv:2003.10392*, 2020.

[10] 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*.

[11] 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.

[12] 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. 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.

Selected Talks

2019-2021 **Delivered invited tutorials on Foundation of Deep Learning at MIT, Simons Institute at UC Berkeley, Machine Learning Summer School, and JHU MINDS Institute.**

2021 **Algorithmic Regularization of Stochastic Gradient**, *Google Research, Workshop on Conceptual Understanding of Deep Learning*

2020-2021 **Provable Representation Learning in Deep Learning**, *Cambridge, Baidu Research, SImons Institute, Institute of Advanced Studies, Northwestern University, TTI-Chicago, ITA 2020, University of British Columbia*

2020-2021 **Beyond Linearization in Deep Learning**, *Army Research Office, Tsinghua University, Beijing Academy of Artificial Intelligence, Rensselaer Polytechnic Institute, Simons Institute, Institute of Advanced Studies*

2018-2019 **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 Non-convex 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,*

External Service

Area Chair/Senior PC for ICML, COLT, and NeurIPS conferences.

Delivered invited tutorials on Foundation of Deep Learning at MIT, Simons Institute at UC Berkeley, Machine Learning Summer School, and JHU MINDS Institute.

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

PhD Supervision

Alex Damian (PACM), 2020–current

Kurtland Chua (CS), 2019–current

Tianle Cai (ECE), 2021–current

Eshaan Nichani (ECE), co-advised with Yuxin Chen, 2021 – current

Wenhao Zhan (ECE), co-advised with Yuxin Chen, 2021 – current

Postdoctoral Supervision

Qi Lei, 2019 – 2022, first employment: Assistant Professor at NYU

Qian Yu, 2021 – current

Itay Safran, 2020 – current

Cong Fang, 2019 – 2020, first employment: Assistant Professor at Peking University

Maziar Sanjabi, 2017 – 2019, first employment: Researcher at Meta AI Research