

EDUCATION

Ph.D. in Statistics, Stanford University, Stanford, CA 2014 - 2019 (Expected)
Specialization in Statistical Learning Theory and Non-convex Optimization.
Advisor: Prof. John Duchi

B.S. in Mathematics, Peking University, Beijing, China Sep 2010 - July 2014
GPA: 3.86 / 4, *Honored Graduate*

RESEARCH INTERESTS

Deep generative models; Principled approaches for deep learning; Non-convex optimization.

PUBLICATIONS

- (1) **Wasserstein Regularization Methods for Quantized Neural Networks.**
Yu Bai, Edo Liberty, Yu-Xiang Wang, 2018. *Working paper*.
- (2) **Complete Dictionary Learning via Nonsmooth Optimization over the Sphere.**
Yu Bai, Qijia Jiang, Ju Sun, 2018. *Working paper*.
- (3) **Approximability of Discriminators Implies Diversity in GANs.**
Yu Bai, Tengyu Ma, Andrej Risteski, 2018. *Submitted*. arXiv preprint arXiv:1806.10586.
- (4) **On the Connection Between Sequential Quadratic Programming and Riemannian Gradient Methods.**
Yu Bai, Song Mei, 2018. arXiv preprint arXiv:1805.08756.
- (5) **Penalty Methods for Low-rank Semidefinite Optimizaion: Local Geometry and Linear Convergence.**
Yu Bai, Song Mei, John Duchi, 2018+. *Working paper*.
- (6) **Model Fidelity, Randomization, and Adaptivity in Linear Experimental Design.**
Yu Bai, John Duchi, 2018+. *Working paper*.
- (7) **TAPAS: Two-pass Approximate Adaptive Sampling for Softmax.**
Yu Bai, Sally Goldman, Li Zhang, 2017. arXiv preprint arXiv:1707.03073.
- (8) **The Landscape of Empirical Risk for Non-convex Losses.**
Song Mei, Yu Bai, Andrea Montanari, 2016. To appear in the *Annals of Statistics*. arXiv preprint arXiv:1607.06534.

INTERNSHIPS

Research Intern, Amazon AI Palo Alto, CA
Host: Edo Liberty June 2018 - Sep 2018

Proposed prox-gradient methods with Wasserstein regularizers for training quantized neural networks. The regularization method is flexible and easy to tune, and matches state-of-the-art numbers on Image Classification with binary ResNet and Language Modeling tasks with 2-bit LSTM. Results suggest a principled understanding of when and why quantized training succeeds.

Research Intern, Google Research Mountain View, CA
Host: Li Zhang June 2016 - Sep 2016

Proposed adaptive sampling strategies for softmax in feedforward neural networks for extreme classification. The adaptive sampling works better than non-adaptive strategies on simulated datasets and achieves new state-of-the-art accuracy on a large-scale Youtube benchmark dataset. The algorithm was made available in Tensorflow (`tf.contrib.nn.rank_sampled_softmax_loss`).

SELECTED COURSEWORK

Convolutional Neural Networks for Visual Recognition (CS231N).
Theories of Deep Learning (Stats 385).
Numerical Linear Algebra (CME302).
Inference, Estimation, and Information Processing (EE378B).
Machine Learning Theory (CS229T).
Convex Optimization (EE364B).
Information Theory and Statistics (Stats 311/EE377).
Theory of Statistics (Stats 300A/B/C).
Theory of Probability (Stats 310A/B/C).

TEACHING EXPERIENCES

Selected courses TA'ed:

Theory of Probability (Stats 310A/B/C).
Theory of Statistics (Stats 300A/B).
Statistical Inference (Stats 200).
Introduction to Stochastic Processes (Stats 217).

COMPUTER SKILLS

Machine Learning Libraries: Tensorflow, PyTorch.
Languages & Software: Python, Julia, C, C++, Matlab, R, Linux.