

Zhewei Yao | Curriculum Vitae

Soda 465, Berkeley, CA 94704

✉ zhewei@berkeley.edu • [yaozhewei.github.io](https://github.com/yaozhewei) • [in zhewei-yao](https://www.linkedin.com/in/zhewei-yao)
[yaozhewei](https://github.com/yaozhewei)

I am a Ph.D. student in the [RISELab \(former AMPLab\)](#), [BDD](#) and [Math Department](#) at University of California at Berkeley. I am advised by [Michael Mahoney](#). My research interest lies in computing statistics, optimization and machine learning. Currently, I am interested in leveraging tools from randomized linear algebra to provide efficient and scalable solutions for large-scale optimization and learning problems. I am also working on the theory and application of deep learning.

Education

- **University of California at Berkeley** **CA, USA**
Ph.D. in Applied Mathematics, Department of Mathematics *Sep. 2016–Present*
- **Shanghai Jiao Tong University** **Shanghai China**
B.S. in Applied Mathematics, Zhiyuan Honor College *Sep. 2012–Jun. 2016*

Publications

- **Q-BERT: Hessian Based Ultra Low Precision Quantization of BERT**
S. Shen, Z. Dong, J. Ye, L. Ma, Z. Yao, A. Gholami, M. W. Mahoney, K. Keutzer
arxiv preprint [1909.05840](#)
- **ANODEV2: A Coupled Neural ODE Evolution Framework**
T. Zhang, Z. Yao*, A. Gholami*, K. Keutzer, J. Gonzalez, G. Biro, and M. W. Mahoney*
arxiv preprint [1906.04596](#)
Proc. NeurIPS 2019
- **Residual Networks as Nonlinear Systems: Stability Analysis using Linearization**
K. Rothauge, Z. Yao, Z. Hu, and M. W. Mahoney
arxiv preprint [1905.13386](#)
- **HAWQ: Hessian AWARE Quantization of Neural Networks with Mixed-Precision**
Z. Dong, Z. Yao*, A. Gholami*, M. W. Mahoney, K. Keutzer*
arxiv preprint [1905.03696](#)
Proc. ICCV 2019
- **Inefficiency of K-FAC for Large Batch Size Training**
L. Ma, G. Montague, J. Ye, Z. Yao, A. Gholami, K. Keutzer, M. W. Mahoney
arxiv preprint [1903.06237](#)
- **Shallow Learning for Fluid Flow Reconstruction with Limited Sensors and Limited Data**
N. B. Erichson, L. Mathelin, Z. Yao, S. L. Brunton, M. W. Mahoney, J. N. Kutz
arxiv preprint [1902.07358](#)

- **JumpReLU: A Retrofit Defense Strategy for Adversarial Attacks**
N. B. Erichson, Z. Yao*, M. W. Mahoney*
 arxiv preprint [1904.03750](#)

- **Trust Region Based Adversarial Attack on Neural Networks**
Z. Yao, A. Gholami, P. Xu, K. Keutzer, M. W. Mahoney
 arxiv preprint [1812.06371](#)
 Proc. CVPR 2019

- **Parameter Re-Initialization through Cyclical Batch Scheduling**
N. Mu, Z. Yao*, A. Gholami, K. Keutzer, M. W. Mahoney*
 arxiv preprint [1812.01216](#)
 Proc. MLSYS Workshop at NeurIPS 2018

- **On the Computational Inefficiency of Large Batch Sizes for Stochastic Gradient Descent**
N. Golmant, N. Vemuri, Z. Yao, V. Feinberg, A. Gholami, K. Rothauge, M. W. Mahoney, J. Gonzalez
 arxiv preprint [1811.12941](#)

- **Large batch size training of neural networks with adversarial training and second-order information**
Z. Yao, A. Gholami*, K. Keutzer, M. W. Mahoney*
 arxiv preprint [1810.01021](#)

- **Hessian-based Analysis of Large Batch Training and Robustness to Adversaries**
Z. Yao, A. Gholami*, Q. Lei K. Keutzer, M. W. Mahoney*
 arxiv preprint [1802.08241](#)
 Proc. NeurIPS 2018

- **Inexact non-convex Newton-type methods**
Z. Yao, P. Xu, F. Roosta-Khorasani, M. W. Mahoney
 arxiv preprint [1802.06925](#)

- **A hybrid adaptive MCMC algorithm in function spaces**
Q. Zhou, Z. Hu, Z. Yao, J. Li
 arxiv preprint [1607.01458](#)
 SIAM/ASA Journal on Uncertainty Quantification 5 (1), 621-639

- **On an adaptive preconditioned Crank–Nicolson MCMC algorithm for infinite dimensional Bayesian inference**
Z. Hu, Z. Yao*, J. Li*
 arxiv preprint [1511.05838](#)
 Journal of Computational Physics 332, 492-503

- **A TV-Gaussian prior for infinite-dimensional Bayesian inverse problems and its numerical implementation**
Z. Yao, Z. Hu*, J. Li*
 arxiv preprint [1510.05239](#)
 Inverse Problems 32 (7), 075006 (*Highlight Paper*)

Research Experiences

- **University of California at Berkeley** **CA, USA**
Ph.D. Researcher at RiseLab and BDD *Sep. 2016–Present*
 - Develop trust region based adversarial attack and propose statistical based defense method to adversarial attack
 - Use ODE method to explain the behavior of residual neural network
 - Used Hessian information to (i) analyze large batch training and robustness of neural networks (ii) train neural networks for large batch training (iii) determine mixed-precision and fine-tuning order for quantizing neural network
 - Investigated the scaling behavior of stochastic gradient descent and K-FAC with large batch sizes for neural networks
 - Proposed stochastic variants of 2nd-order methods for non-convex optimization problem and establish theories
 - Applied deep learning to other fields, e.g. scientific datasets and fluid dynamics
- **Amazon AWS AI** **CA, USA**
Applied Scientist *May. 2019–Aug. 2019*
 - Applied machine learning algorithm to explore very large scale configurations problems
 - Investigated transfer learning and exploration of TVM computation configuration generation with different batch sizes and GPUs
 - Investigated reinforce learning to explore fast database query answering, particularly on the Materialized View Update and Vacuum frequency.
- **Alibaba** **Beijing, China**
Researcher intern at Alimama *Dec. 2018–Jan. 2019*
 - Investigated over-fitting of recommendation system
 - Investigated large batch training of recommendation system
- **Lawrence Berkeley National Laboratory** **CA, USA**
Researcher intern at NERSC *May. 2018–Aug. 2018*
 - Implemented CPU Parallelization of PyTorch to train large climate dataset (over 400 Gb)
 - Tested robustness on models trained with scientific datasets
- **Shanghai Jiao Tong University** **Shanghai, China**
Undergraduate Researcher *Sep. 2014–Jun. 2016*
 - Considered MCMC algorithm in infinite-dimensional space
 - Designed a TG-prior with better edge-preserving property and two new adaptive algorithms

Others

- **Programming Languages:** C++, Matlab, Python, Pytorch, Tensorflow
- **Conference Reviewer:** NeurIPS 2018, ICLR 2019
- **Teaching:**
 - **Stat 89A: Linear Algebra for Data Science** **UC Berkeley**
Graduate Student Instructor *Spring 2018*

