

# Hao-Jun Michael Shi

✉ [hjmshi@meta.com](mailto:hjmshi@meta.com) • [hjmshi.github.io](https://github.com/hjmshi) • ☎ (516) 343-4934

## Education

### Northwestern University

*PhD in Industrial Engineering and Management Sciences*

Advisor: Prof. Jorge Nocedal

**Evanston, IL**

2016–2021

### Kellogg School of Management at Northwestern University

*Management for Scientists and Engineers Certificate*

**Evanston, IL**

2020

### Northwestern University

*MS in Industrial Engineering and Management Sciences*

**Evanston, IL**

2016–2017

### University of California, Los Angeles

*BS in Applied Mathematics*

College and Departmental Honors

**Los Angeles, CA**

2012–2016

## Research Interests

- deep learning training algorithms and systems
- stochastic optimization
- noisy optimization
- derivative-free optimization

## Awards

- 2019 International Conference on Machine Learning Top 5% Reviewer
- 2017, 2018 IEMS Departmental Service Award
- 2016, 2017 NSF Graduate Research Fellowship: Honorable Mention
- 2016 Walter P. Murphy Fellowship

## Industry Experience

### Meta Platforms

*Research Scientist*

AI and Systems Co-Design

**Menlo Park, CA**

2021–Current

### Facebook

*Research Intern*

Advisor: Dheevatsa Mudigere

**Menlo Park, CA**

2019

## Journal Publications

1. H.-J.M. Shi, T.-H. Lee, S. Iwasaki, J. Gallego-Posada, Z. Li, K. Rangadurai, D. Mudigere, and M. Rabbat. "A Distributed Data-Parallel PyTorch Implementation of the Distributed Shampoo Optimizer for Training Neural Networks At-Scale". Forthcoming.
2. H.-J.M. Shi, Y. Xie, M.Q. Xuan, and J. Nocedal. "Adaptive Finite-Difference Interval Estimation for Noisy Derivative-Free Optimization". *SIAM Journal of Scientific Computing*, 44(4), A2302-A2321, 2022. [ArXiv].
3. H.-J.M. Shi, M.Q. Xuan, F. Oztoprak, and J. Nocedal. "On the Numerical Performance of Derivative-Free Optimization Methods Based on Finite-Difference Approximations". *Optimization Methods and Software*, 1-23, 2022. [ArXiv].

4. H.-J.M. Shi, Y. Xie, R. Byrd, and J. Nocedal. "A Noise-Tolerant Quasi-Newton Algorithm for Unconstrained Optimization". *SIAM Journal of Optimization*, 32(1), 29-55, 2022. [ArXiv].
5. J. Luo, K. Shapiro, H.-J.M. Shi, Q. Yang, and K. Zhu. "Practical Algorithms for Learning Near-Isometric Linear Embeddings". *SIAM Undergraduate Research Online*, vol. 9, 2016. [SIURO].

## Conference Proceedings

1. H.-J.M. Shi, D. Mudigere, M. Naumov, and J. Yang. "Compositional Embeddings Using Complementary Partitions for Memory-Efficient Recommendation Systems". *KDD, Virtual Conference*, August 2020. [KDD].
2. R. Bollapragada, D. Mudigere, J. Nocedal, H.-J.M. Shi, and P.T.P. Tang. "A Progressive Batching L-BFGS Method for Machine Learning". *International Conference on Machine Learning (ICML)*, Stockholm, Sweden, July 2018. [ICML]
3. H.-J.M. Shi, M. Case, X. Gu, S. Tu, and D. Needell. "Methods for Quantized Compressed Sensing". *Proc. Information Theory and Applications (ITA)*, La Jolla, CA, Jan. 2016. [ITA].

## Technical Reports

1. M. Naumov, D. Mudigere, H.-J.M. Shi, J. Huang, N. Sundaraman, J. Park, X. Wang, U. Gupta, C.-J. Wu, A.G. Azzolini, D. Dzulgakov, A. Mallevich, I. Cherniavskii, Y. Lu, R. Krishnamoorthi, A. Yu, V. Kondratenko, S. Pereira, X. Chen, W. Chen, V. Rao, B. Jia, L. Xiong, M. Smelyanskiy. "Deep Learning Recommendation Model for Personalization and Recommendation Systems". Preprint. [ArXiv].
2. H.-J.M. Shi, S. Tu, Y. Xu, and W. Yin. "A Primer on Coordinate Descent Algorithms". Preprint. [ArXiv].
3. X. Gu, S. Tu, H.-J.M. Shi, M. Case, D. Needell, and Y. Plan. "Optimizing Quantization for Lasso Recovery". *IEEE Signal Processing Letters*, vol. 25, issue 1, Jan. 2018. [IEEE].
4. C. Abrahamson, H.-J.M. Shi, and B. Yang. "Ground Motion Prediction Equations for Arias Intensity Consistent with the NGA-West2 Ground Motion Models". *Pacific Earthquake Engineering Research (PEER) Report*, July 2016. [PEER].

## Presentations

1. "Distributed Data-Parallel Training of Neural Networks At-Scale Using Distributed Shampoo". *SIAM Conference on Optimization 2023*, Seattle, Washington, June 2023.
2. "Distributed Data-Parallel Training of Neural Networks At-Scale Using Distributed Shampoo". *FAIR Montreal*, Montreal, Canada, April 2023.
3. "Adaptive Finite-Difference Methods for Noisy Derivative-Free Optimization". *INFORMS 2021*, Remote, Oct. 2021.
4. "Stochastic Line Searches (@FB?)". Facebook, Menlo Park, California, Sept. 2021.
5. "Finite-Difference Methods for Noisy Derivative-Free Optimization". *Math 290J*, UCLA, April 2021.
6. "Recent Advancements in Stochastic, Noisy, and Derivative-Free Optimization Methods for Machine Learning". *William and Mary*, Feb. 2021.
7. "A Noise-Tolerant Quasi-Newton Method for Unconstrained Optimization". *INFORMS 2020*, Remote, Nov. 2020.
8. "Compositional Embeddings Using Complementary Partitions for Memory-Efficient Recommendation Systems". *ACM SIGKDD*, Remote, Aug. 2020.
9. "Compositional Embeddings Using Complementary Partitions for Memory-Efficient Recommendation Systems". *PerSoNAI tutorial, ISCA 2020*, Remote, May 2020.
10. "Towards Understanding Embeddings and Optimization in Deep Neural Recommendation Systems". *INFORMS 2019*, Seattle, Washington, Oct. 2019.

11. "Recent Advancements in Stochastic Quasi-Newton Methods". Facebook, Menlo Park, California, Feb. 2019.
12. "A Progressive Batching L-BFGS Method for Machine Learning". Chicago Area SIAM Student Conference 2018, Chicago, Illinois, Apr. 2018.
13. "A Progressive Batching L-BFGS Method for Machine Learning". Midwest Machine Learning Symposium 2018, Chicago, Illinois, June 2018.
14. "Learning Near-Isometric Linear Embeddings". Joint Mathematics Meetings 2015, San Antonio, Texas, Jan. 2015.

## Teaching Experience

### Northwestern University

- IEMS 351: Optimization Methods in Data Science, Instructor (Spring 2020)
- IEMS 455: Machine Learning, Teaching Assistant (Spring 2018)
- IEMS 1st Year Boot Camp: Analysis, Instructor (Summer 2017)

## Advising

### B.S. and M.S. Students Advised

- Anna Cai, Carnegie Mellon University, Summer 2023.
- Manish Kumar, MS in Analytics, Northwestern University, Winter 2020.  
Current Position: Data Scientist at Microsoft.

## Professional Activities

- Referee for:
  - *SIAM Journal on Optimization*
  - *Mathematical Programming*
  - *Mathematical Programming Computation*
  - *Computational Optimization and Applications*
  - *Journal of Optimization Theory and Applications*
  - *International Conference on Machine Learning*
  - *Neural Information Processing Systems*
  - *International Conference on Learning Representations*
  - *The International Conference for High Performance Computing, Networking, Storage, and Analysis*
- Student Member of SIAM and INFORMS.
- Northwestern INFORMS Student Chapter Board: Webmaster (2017-18).