

# Shusen Wang

Postdoctoral Scholar, Department of Statistics and RISELab, UC Berkeley

**Phone:** (510)229-0382    **Email:** shusen@berkeley.edu    **Homepage:** wangshusen.github.io

## Research

**Keywords**      Machine learning (ML), randomized linear algebra, convex optimization, big data, distributed computing.

**Summary**      I am most interested in statistical ML and computational methods for ML. The goal of my research is making large-scale ML more accurate and efficient. My thesis work is randomized algorithms for speeding up ML and matrix computation with loss of a limited amount of accuracy. My postdoctoral work at Berkeley is mainly the design of parallel ML and data analysis algorithms and the implementation using Apache Spark.

## Work Experience

**2016-Now**      **Postdoctoral Scholar;** Department of Statistics, UC Berkeley (Berkeley, CA, USA)

*Advisor: Michael W. Mahoney*

*Research: randomized linear algebra and implementation*

**2014-2015**      **Research Intern;** Baidu Big Data Lab (Beijing, China)

*Mentor: Tong Zhang*

*Research: optimization algorithms*

**2012**      **Research Intern;** Google Research (Beijing, China)

*Research: randomized algorithms*

**2011-2012**      **Intern;** Microsoft Research Asia (Beijing, China)

*Mentors: Haixun Wang and Yangqiu Song*

*Project: probabilistic knowledge base*

## Education

**2011-2016**      **Doctor of Engineering;** Zhejiang University (Hangzhou, China)

*Major: Computer Science*

*Advisor: Zhihua Zhang*

*Thesis: Large-Scale Machine Learning: A Randomized Approach and Theoretical Analysis*

**2007-2011**      **Bachelor of Engineering;** Zhejiang University (Hangzhou, China)

## Academic Service

- **Journal Reviewer**
  - Journal of Machine Learning Research 2015 to 2018
  - SIAM Journal on Scientific Computing, 2017
  - ACM Transactions on Mathematical Software, 2017
  - Journal of Econometrics, 2017
  - SIAM Journal on Matrix Analysis and Applications, 2017
- **Conference Reviewer or Committee Member**
  - NIPS 2014, 2015, 2017
  - ICML 2017, 2018
  - IJCAI 2015, 2017, 2018
  - AAAI 2016, 2018

## Major Honors & Awards

- |           |  |
|-----------|--|
| 2014      | <b>Baidu Scholarship</b> , awarded to 8 Chinese students in the world, US\$30,000                    |
| 2013      | <b>Microsoft Research Asia Fellow</b> , awarded to 10 students in Asia Pacific, US\$10,000           |
| 2012      | <b>Scholarship Award for Excellent Doctoral Student Granted by Ministry of Education</b> , US\$5,000 |
| 2012-2014 | <b>National Scholarship for Graduate Students</b> , 3 times, each time US\$5,000                     |

## Journal Papers

- **Scalable Kernel K-Means Clustering with Nystrom Approximation: Relative-Error Bounds.** Shusen Wang, Alex Gittens, and Michael W. Mahoney. **Accepted by** *Journal of Machine Learning Research* **conditioned on minor revisions**; available at *arXiv:1706.02803*, 2017.
- **Sketched Ridge Regression: Optimization Perspective, Statistical Perspective, and Model Averaging.** Shusen Wang, Alex Gittens, and Michael W. Mahoney. **Accepted by** *Journal of Machine Learning Research*; a short version is published in *ICML 2017*.
- **Towards More Efficient SPSD Matrix Approximation and CUR Matrix Decomposition.** Shusen Wang, Zihua Zhang, and Tong Zhang. *Journal of Machine Learning Research*, 17(210):1–49, 2016. (**JMLR** 2016).
- **SPSD Matrix Approximation via Column Selection: Theories, Algorithms, and Extensions.** Shusen Wang, Luo Luo, and Zihua Zhang. *Journal of Machine Learning Research*, 17(49):1–49, 2016. (**JMLR** 2016).
- **Improving CUR Matrix Decomposition and the Nystrom Approximation via Adaptive Sampling.** Shusen Wang and Zihua Zhang. *Journal of Machine Learning Research*, 14: 2729–2769, 2013. (**JMLR** 2013).
- **EP-GIG Priors and Applications in Bayesian Sparse Learning.** Zihua Zhang, Shusen Wang, Dehua Liu, and Michael I. Jordan. *Journal of Machine Learning Research*, 13: 2031–2061, 2012. (**JMLR** 2012).

## Preprints

- **Error Estimation for Randomized Least-Squares Algorithms via the Bootstrap.** Miles E. Lopes, **Shusen Wang**, and Michael W. Mahoney. *arXiv:1803.08021*, 2018.
- **GIANT: Globally Improved Approximate Newton Method for Distributed Optimization.** **Shusen Wang**, Farbod Roosta-Khorasani, Peng Xu, and Michael W. Mahoney. *arXiv:1709.03528*, 2017.
- **A Bootstrap Method for Error Estimation in Randomized Matrix Multiplication.** Miles E. Lopes, **Shusen Wang**, Michael W. Mahoney. Under the review of *Journal of Machine Learning Research*; available at *arXiv:1708.01945*, 2017.
- **Improved Analyses of the Randomized Power Method and Block Lanczos Method.** **Shusen Wang**, Zhihua Zhang, and Tong Zhang. *arXiv:1508.06429*, 2015.

## Conference Papers

- **Sketched Ridge Regression: Optimization Perspective, Statistical Perspective, and Model Averaging.** **Shusen Wang**, Alex Gittens, and Michael W. Mahoney. In *International Conference on Machine Learning*, 2017. (**ICML** 2017).
- **Towards Real-Time Geologic Feature Detection from Seismic Measurements using a Randomized Machine-Learning Algorithm.** Youzuo Lin, **Shusen Wang**, Jayaraman Thiagarajan, George Guthrie, and David Coblentz. In *Proceeding of Society of Exploration Geophysics (SEG)*, 2017.
- **Open Domain Short Text Conceptualization: A Generative + Descriptive Modeling Approach.** Yangqiu Song, **Shusen Wang**, and Haixun Wang. In *International Joint Conference on Artificial Intelligence*, 2015. (**IJCAI** 2015).
- **Improving the Modified Nystrom Method Using Spectral Shifting.** **Shusen Wang**, Chao Zhang, Hui Qian, and Zhihua Zhang. In *the 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2014. (**KDD** 2014).
- **Efficient Algorithms and Error Analysis for the Modified Nystrom Method.** **Shusen Wang** and Zhihua Zhang. In *Proceedings of the 17th International Conference on Artificial Intelligence and Statistics, JMLR W&CP*, 2014. (**AISTATS** 2014).
- **Making Fisher Discriminant Analysis Scalable.** Bojun Tu, Zhihua Zhang, **Shusen Wang**, and Hui Qian. In *the International Conference on Machine Learning*. (**ICML** 2014).
- **Exact Subspace Clustering in Linear Time.** **Shusen Wang**, Bojun Tu, Congfu Xu, and Zhihua Zhang. In *the 28th AAAI Conference on Artificial Intelligence*. (**AAAI** 2014).
- **Using The Matrix Ridge Approximation to Speedup Determinantal Point Processes Sampling Algorithms.** **Shusen Wang**, Chao Zhang, Hui Qian, and Zhihua Zhang. In *the 28th AAAI Conference on Artificial Intelligence*. (**AAAI** 2014).
- **Transfer Understanding from Head Queries to Tail Queries.** Yangqiu Song, Haixun Wang, Weizhu Chen, and **Shusen Wang**. In *the 23rd ACM International Conference on Information and Knowledge Management*. (**CIKM** 2014).
- **Nonconvex Relaxation Approaches to Robust Matrix Recovery.** **Shusen Wang**, Dehua Liu, and Zhihua Zhang. In *International Joint Conference on Artificial Intelligence*. (**IJCAI** 2013).
- **A Scalable CUR Matrix Decomposition Algorithm: Lower Time Complexity and Tighter Bound.** **Shusen Wang** and Zhihua Zhang. In *Advances in Neural Information Processing Systems 25*. (**NIPS** 2012).
- **Colorization by Matrix Completion.** **Shusen Wang** and Zhihua Zhang. In *the 26th AAAI Conference on Artificial Intelligence*. (**AAAI** 2012).

- **Efficient Subspace Segmentation via Quadratic Programming.** Shusen Wang, Xiaotong Yuan, Tiansheng Yao, Shuicheng Yan, and Jialie Shen. In *the 25th AAAI Conference on Artificial Intelligence*. (**AAAI** 2011).