

Homepage: <http://people.csail.mit.edu/zeyuan>

Email: zeyuan@csail.mit.edu

EDUCATION & WORK EXPERIENCE

Researcher	MSR REDMOND (managers: Sebastien Bubeck, Ofer Dekel)	2017-2021
Postdoc	PRINCETON U + IAS (hosts: Elad Hazan, Avi Wigderson)	2015-2017
Ph.D. in Computer Science	MIT (advisors: Jonathan Kelner, Silvio Micali)	2012–2015
S.M. in Computer Science	MIT (thesis award)	2010–2012
B.S. in Math and Physics	TSINGHUA UNIVERSITY (summa cum laude)	2006–2010

COMPETITION AWARDS

2009 World's 2nd Place	<i>ACM/ICPC World Final</i>
2008 World's 2nd Place	<i>Google Code Jam</i>
2006 World Champion	<i>USA Computing Olympiad</i>
2006 Gold Medal	<i>International Olympiad in Informatics</i>
2005 Gold Medal	<i>International Olympiad in Informatics</i>

RESEARCH INTERESTS

Deep Learning, Machine Learning, Optimization, Algorithms

In recent years, I work on the mathematical foundations of Deep Learning, where the ultimate goal is to turn black magic into scientific theorems, and then design more principled algorithms for a better, safer, more economical world of artificial intelligence. In my past life, I also worked on the theory of machine learning and optimization, and apply them to theoretical computer science, operations research, and statistics.

As part of my research, I maintain my own code base for training ResNet, DenseNet, Transformers, BERT, GPT-2, Vision Transformers, and apply them to vision tasks, Q&A tasks such as SQUAD, language modeling tasks such as WikiText103, text generation tasks such as E2E, WikiSQL, etc.

RESEARCH AWARDS

2021 Microsoft Special Stock Award	<i>Microsoft Research</i>
2018 Microsoft Special Stock + Cash Award	<i>Microsoft Research</i>
2016 Microsoft Azure Research Grant	<i>Microsoft Research</i>
2013 Simons Award for TCS Graduate Students	<i>Simons Foundation</i>
2009 Best Student Paper Runner-Up	<i>IEEE International Conference on Data Mining</i>
2009 Microsoft Young Fellow	<i>Microsoft Research Asia</i>

DRAFTS + PUBLICATIONS

Due to space limitations, many names are abbreviated but can be found in full in my homepage.

🌟 indicators non-first-author papers of mine ; ★ indicates I am an equal-contribution first author

1. **"Towards Understanding Ensemble, Knowledge Distillation and Self-Distillation in Deep Learning"**
Z. Allen-Zhu★, J. Li. *(arxiv draft, July 2021)*
2. **"LoRA: Low-Rank Adaptation of Large Language Models"**
E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu🌟, Y. Li, S. Wang, W. Chen. *(arxiv draft, June 2021)*

Our LoRA algorithm is shipped to some Microsoft products, see e.g. [this blog post](#) on August 23, 2021.

3. **"Forward Super-Resolution: How Can GANs Learn Hierarchical Generative Models for Real-World Distributions"**
Z. Allen-Zhu★, J. Li. *(arxiv draft, June 2021)*
4. **"Backward Feature Correction: How Deep Learning Performs Deep Learning"**
Z. Allen-Zhu★, J. Li. *(arxiv draft, March 2021)*
5. **"Feature Purification: How Adversarial Training Performs Robust Deep Learning"**
Z. Allen-Zhu★, J. Li. *Symposium on Foundations of Computer Science (FOCS 2021)*
6. **"Byzantine-Resilient Non-Convex Stochastic Gradient Descent"**
Z. Allen-Zhu★, F. Ebrahimi, J. Li, D. Alistarh. *Conference on Learning Representations (ICLR 2021)*
7. **"Near-Optimal Discrete Optimization for Experimental Design: A Regret Minimization Approach"**
Z. Allen-Zhu★, Y. Li, A. Singh, Y. Wang. *Mathematical Programming (MAPR 2020)*
8. **"What Can ResNet Learn Efficiently, Going Beyond Kernels?"**
Z. Allen-Zhu★, Y. Li. *Neural Information Processing Systems (NeurIPS 2019)*
9. **"Learning and Generalization in Overparameterized Neural Networks, Going Beyond Two Layers"**
Z. Allen-Zhu★, Y. Li, Y. Liang. *Neural Information Processing Systems (NeurIPS 2019)*
10. **"Can SGD Learn Recurrent Neural Networks with Provable Generalization?"**
Z. Allen-Zhu★, Y. Li. *Neural Information Processing Systems (NeurIPS 2019)*
11. **"On the Convergence Rate of Training Recurrent Neural Networks"**
Z. Allen-Zhu★, Y. Li, Z. Song. *Neural Information Processing Systems (NeurIPS 2019)*
12. **"A Convergence Theory for Deep Learning via Over-Parameterization"**
Z. Allen-Zhu★, Y. Li, Z. Song. *International Conference on Machine Learning (ICML 2019)*
13. **"Natasha 2: Faster Non-Convex Optimization Than SGD"**
Z. Allen-Zhu★. *Neural Information Processing Systems (NeurIPS 2018)*
14. **"How To Make the Gradients Small Stochastically: Even Faster Convex and Nonconvex SGD"**
Z. Allen-Zhu★. *Neural Information Processing Systems (NeurIPS 2018)*
15. **"Is Q-Learning Provably Efficient?"**
C. Jin, Z. Allen-Zhu★, S. Bubeck, M. Jordan. *Neural Information Processing Systems (NeurIPS 2018)*
16. **"The Lingering of Gradients: How to Reuse Gradients Over Time"**
Z. Allen-Zhu★, D. Simchi-Levi, X. Wang. *Neural Information Processing Systems (NeurIPS 2018)*
17. **"Byzantine Stochastic Gradient Descent"**
D. Alistarh, Z. Allen-Zhu★, J. Li. *Neural Information Processing Systems (NeurIPS 2018)*
18. **"Neon2: Finding Local Minima via First-Order Oracles"**
Z. Allen-Zhu★, Y. Li. *Neural Information Processing Systems (NeurIPS 2018)*
19. **"Katyusha X: Practical Momentum Method for Stochastic Sum-of-Nonconvex Optimization"**
Z. Allen-Zhu★. *International Conference on Machine Learning (ICML 2018)*
20. **"Make the Minority Great Again: First-Order Regret Bound for Contextual Bandits"**
Z. Allen-Zhu★, S. Bubeck, Y. Li. *International Conference on Machine Learning (ICML 2018)*
21. **"Operator Scaling via Geodesically Convex Optimization, Invariant Theory and Polynomial Identity Testing"**
Z. Allen-Zhu★, A. Garg, Y. Li, R. Oliveira, A. Wigderson. *Symposium on Theory of Computing (STOC 2018)*

22. **"Linear Convergence of a Frank-Wolfe Type Algorithm over Trace-Norm Balls"**
Z. Allen-Zhu★, E. Hazan, W. Hu, Y. Li. *Neural Information Processing Systems (NeurIPS 2017)*
23. **"First Efficient Convergence for Streaming k-PCA: a Global, Gap-Free, and Near-Optimal Rate"**
Z. Allen-Zhu★, Y. Li. *Symposium on Foundations of Computer Science (FOCS 2017)*
24. **"Much Faster Algorithms for Matrix Scaling"**
Z. Allen-Zhu★, Y. Li, R. Oliveira, A. Wigderson. *Symposium on Foundations of Computer Science (FOCS 2017)*
25. **"Follow the Compressed Leader: Faster Algorithm for Matrix Multiplicative Weight Updates"**
Z. Allen-Zhu★, Y. Li. *International Conference on Machine Learning (ICML 2017)*
26. **"Natasha: Faster Stochastic Non-Convex Optimization via Strongly Non-Convex Parameter"**
Z. Allen-Zhu★. *International Conference on Machine Learning (ICML 2017)*
27. **"Near-Optimal Design of Experiments via Regret Minimization"**
Z. Allen-Zhu★, Y. Li, A. Singh, Y. Wang. *International Conference on Machine Learning (ICML 2017)*
Mathematical Programming (MAPR 2020)
28. **"Faster Principal Component Regression and Stable Matrix Chebyshev Approximation"**
Z. Allen-Zhu★, Y. Li. *International Conference on Machine Learning (ICML 2017)*
29. **"Doubly Accelerated Methods for Faster CCA and Generalized Eigendecomposition"**
Z. Allen-Zhu★, Y. Li. *International Conference on Machine Learning (ICML 2017)*
30. **"Katyusha: The First Direct Acceleration of Stochastic Gradient Methods"**
Z. Allen-Zhu★. *Symposium on Theory of Computing (STOC 2017)*
Journal of Machine Learning Research (JMLR 2018)
31. **"Finding Approximate Local Minima for Nonconvex Optimization in Linear Time"**
N. Agarwal, Z. Allen-Zhu★, B. Bullins, E. Hazan, T. Ma. *Symposium on Theory of Computing (STOC 2017)*
32. **"Linear Coupling: An Ultimate Unification of Gradient and Mirror Descent"**
Z. Allen-Zhu★, L. Orecchia. *Innovations in Theoretical Computer Science (ITCS 2017)*
33. **"Even Faster SVD Decomposition Yet Without Agonizing Pain"**
Z. Allen-Zhu★, Y. Li. *Neural Information Processing Systems (NIPS 2016)*
34. **"Optimal Black-Box Reductions Between Optimization Objectives"**
Z. Allen-Zhu★, E. Hazan. *Neural Information Processing Systems (NIPS 2016)*
35. **"Exploiting the Structure: Stochastic Gradient Methods Using Raw Clusters"**
Z. Allen-Zhu★, Y. Yuan, K. Sridharan. *Neural Information Processing Systems (NIPS 2016)*
36. **"Variance Reduction for Faster Non-Convex Optimization"**
Z. Allen-Zhu★, E. Hazan. *International Conference on Machine Learning (ICML 2016)*
37. **"Even Faster Accelerated Coordinate Descent Using Non-Uniform Sampling"**
Z. Allen-Zhu★, P. Richtárik, Z. Qu, Y. Yuan. *International Conference on Machine Learning (ICML 2016)*
38. **"Improved SVRG for Non-Strongly-Convex or Sum-of-Non-Convex Objectives"**
Z. Allen-Zhu★, Y. Yuan. *International Conference on Machine Learning (ICML 2016)*
39. **"Optimization Algorithms for Faster Computational Geometry"**
Z. Allen-Zhu★, Z. Liao, Y. Yuan. *International Colloquium on Automata, Languages, and Programming (ICALP 2016)*

40. **"The Restricted Isometry Property for the General p -Norms"**
Z. Allen-Zhu★, R. Gelashvili and I. Razenshteyn. *IEEE Transactions on Information Theory (IEEE-IT 2016)*
International Symposium on Computational Geometry (SoCG 2015)
41. **"Using Optimization to Obtain a Width-Independent, Parallel, Simpler, and Faster Positive SDP Solver"**
Z. Allen-Zhu★, Y. T. Lee, and L. Orecchia. *Symposium on Discrete Algorithms (SODA 2016)*
42. **"Expanders via Local Edge Flips"**
Z. Allen-Zhu★, A. Bhaskara, S. Lattanzi, V. Mirrokni, and L. Orecchia.
Symposium on Discrete Algorithms (SODA 2016)
43. **"Nearly-Linear Time Positive LP Solver with Faster Convergence Rate"**
Z. Allen-Zhu★, L. Orecchia. *Symposium on Theory of Computing (STOC 2015)*
Mathematical Programming (MAPR 2018)
44. **"Spectral Sparsification and Regret Minimization Beyond Multiplicative Updates"**
Z. Allen-Zhu★, Z. Liao, L. Orecchia. *Symposium on Theory of Computing (STOC 2015)*
45. **"Using Optimization to Solve Positive LPs Faster in Parallel"**
Z. Allen-Zhu★, L. Orecchia. *Symposium on Discrete Algorithms (SODA 2015)*
46. **"Johnson-Lindenstrauss Compression with Neuroscience-Based Constraints"**
Z. Allen-Zhu★, R. Gelashvili, S. Micali, N. Shavit.
Proceedings of the National Academy of Sciences (PNAS 2014)
- ✧ family name legally changed from Zeyuan Allen Zhu to Zeyuan Allen-Zhu in 2014.
47. **"Reconstructing Markov Processes from Succinct and Independent Experiment"**
S. Micali, Z. A. Zhu★. *Discrete Applied Mathematics, vol. 200 (DAM 2015)*
48. **"Shorter Arithmetization of Nondeterministic Computations"**
A. Chiesa, Z. A. Zhu★. *Theoretical Computer Science, vol. 600 (TCS 2015)*
49. **"Knightian Robustness of the Vickrey Mechanism"**
A. Chiesa, S. Micali and Z. A. Zhu★. *Econometrica, vol. 83, No. 5 (Econometrica 2015)*
50. **"Knightian Self Uncertainty in the VCG Mechanism for Unrestricted Combinatorial Auctions"**
A. Chiesa, S. Micali and Z. A. Zhu★. *ACM Conference on Economics and Computation (EC 2014)*
51. **"Flow-Based Algorithms for Local Graph Clustering"**
L. Orecchia and Z. A. Zhu★. *Symposium on Discrete Algorithms (SODA 2014)*
52. **"A Local Algorithm for Finding Well-Connected Clusters"**
Z. A. Zhu★, S. Lattanzi and V. Mirrokni. *International Conference on Machine Learning (ICML 2013)*
53. **"A Simple, Combinatorial Algorithm for Solving SDD Systems in Nearly-Linear Time"**
J. Kelner, L. Orecchia, A. Sidford and Z. A. Zhu★. *Symposium on Theory of Computing (STOC 2013)*
54. **"Mechanism Design with Approximate Valuations"**
A. Chiesa, S. Micali and Z. A. Zhu★. *Innovations in Theoretical Computer Science (ITCS 2012)*
55. **"Randomized Accuracy-Aware Program Transformations for Efficient Approximate Computations"**
Z. A. Zhu★, S. Misailovic, J. Kelner and M. Rinard.
Symposium on Principles of Programming Languages (POPL 2012)
56. **"Optimal Pricing in Social Networks with Incomplete Information"**
W. Chen, P. Lu, X. Sun, B. Tang, Y. Wang and Z. A. Zhu★.
Workshop on Internet & Network Economics (WINE 2011)

57. **"Asymptotically Optimal Strategy-Proof Mechanisms for Two-Facility Games"**
P. Lu, X. Sun, Y. Wang and Z. A. Zhu★. *ACM Conference on Economics and Computation (EC 2010)*
58. **"A Novel Click Model and Its Applications to Online Advertising"**
Z. A. Zhu★, W. Chen, T. Minka, C. Zhu, Z. Chen.
International Conference on Web Search and Data Mining (WSDM 2010)
59. **"To Divide and Conquer Search Ranking by Learning Query Difficulty"**
Z. A. Zhu★, W. Chen, T. Wan, C. Zhu, G. Wang, Z. Chen.
Conference on Information and Knowledge Management (CIKM 2009)
60. **"A General Magnitude-Preserving Boosting Algorithm for Search Ranking"**
C. Zhu, W. Chen, Zeyuan A. Zhu🌟, G. Wang, D. Wang, Z. Chen.
Conference on Information and Knowledge Management (CIKM 2009)
61. **"Inverse Time Dependency in Convex Regularized Learning"** (best student paper runner-up)
Z. A. Zhu★, W. Chen, C. Zhu, G. Wang, H. Wang, Z. Chen.
International Conference on Data Mining (ICDM 2009)
62. **"P-packSVM: Parallel Primal gradient desCent Kernel SVM"**
Z. A. Zhu★, W. Chen, G. Wang, C. Zhu, Z. Chen. *International Conference on Data Mining (ICDM 2009)*