

ZHIYUAN LI

zhiyuanli.ttic.edu

zhiyuanli@ttic.edu

+1 609-356-4247

RESEARCH INTEREST

Machine learning theory, including adaptive optimization methods in deep learning, reasoning capabilities of language models, modern paradigm of generalization (overparametrization, out-of-domain generalization) and its connection to the implicit bias of optimization algorithms.

PROFESSIONAL EXPERIENCE

Toyota Technological Institute at Chicago

September 2023 - Now

Tenure-track Assistant Professor

Stanford University

September 2022 - August 2023

Postdoctoral Researcher in Computer Science

Advisor: Prof. Tengyu Ma

EDUCATION

Princeton University

September 2017 - August 2022

Ph.D. in Computer Science

Advisor: Prof. Sanjeev Arora

Tsinghua University

August 2013 - June 2017

B.E. in Computer Science & Engineering

Special CS Pilot Class (Yao Class)

PUBLICATIONS

(α - β) denotes alphabetical ordering. (*) denotes equal contribution.

- [1] Nirmal Joshi, Gal Vardi, Adam Block, Surbhi Goel, **Zhiyuan Li**, Theodor Misiakiewicz, Nathan Srebro.
A Theory of Learning with Autoregressive Chain of Thought.
Proceedings of the 38th Conference on Learning Theory (COLT). 2025.
- [2] Shuo Xie, Tianhao Wang, Sashank Reddi, Sanjiv Kumar, **Zhiyuan Li**.
Structured Preconditioners in Adaptive Optimization: A Unified Analysis.
Proceedings of the 42nd International Conference on Machine Learning (ICML). 2025.
- [3] Chenxiao Yang, Nathan Srebro, David McAllester, **Zhiyuan Li**.
PENCIL: Long Thoughts with Short Memory.
Proceedings of the 42nd International Conference on Machine Learning (ICML). 2025.
- [4] Marko Medvedev, Kaifeng Lyu, Dingli Yu, Sanjeev Arora, **Zhiyuan Li**, Nathan Srebro.
Weak-to-Strong Generalization Even in Random Feature Networks, Provably.
Proceedings of the 42nd International Conference on Machine Learning (ICML). 2025.
- [5] Thomas Chen, Tengyu Ma, **Zhiyuan Li**.
Non-Asymptotic Length Generalization.
Proceedings of the 42nd International Conference on Machine Learning (ICML). 2025.

- [6] Chenxiao Yang, **Zhiyuan Li**, David Wipf.
Chain-of-Thought Provably Enables Learning the (Otherwise) Unlearnable.
The Thirteenth International Conference on Learning Representations (ICLR). 2025.
- [7] Yida Yin, Zhiqiu Xu, **Zhiyuan Li**, Trevor Darrell, Zhuang Liu.
A Coefficient Makes SVRG Effective.
The Thirteenth International Conference on Learning Representations (ICLR). 2025.
- [8] Nikunj Saunshi, Nishanth Dikkala, **Zhiyuan Li**, Sanjiv Kumar, Sashank J Reddi.
Reasoning with Latent Thoughts: On the Power of Looped Transformers.
The Thirteenth International Conference on Learning Representations (ICLR). 2025.
- [9] Kaiyue Wen, **Zhiyuan Li**, Jason Wang, David Hall, Percy Liang, Tengyu Ma.
Understanding Warmup-Stable-Decay Learning Rates: A River Valley Loss Landscape Perspective.
The Thirteenth International Conference on Learning Representations (ICLR). 2025.
- [10] Shuo Xie, Mohamad Amin Mohamadi, **Zhiyuan Li**.
Adam Exploits ℓ_∞ -geometry of Loss Landscape via Coordinate-wise Adaptivity.
The Thirteenth International Conference on Learning Representations (ICLR). 2025. (Spotlight)
- [11] Shuo Xie, **Zhiyuan Li**.
Implicit Bias of AdamW: ℓ_∞ -Norm Constrained Optimization.
Proceedings of the 41st International Conference on Machine Learning (ICML). PMLR. 2024.
- [12] Mohamad Amin Mohamadi, **Zhiyuan Li**, Lei Wu, Danica J Sutherland.
Why Do You Grok? A Theoretical Analysis of Grokking Modular Addition.
Proceedings of the 41st International Conference on Machine Learning (ICML). PMLR. 2024.
- [13] Khashayar Gatmiry, **Zhiyuan Li**, Sashank J Reddi, Stefanie Jegelka.
Simplicity Bias via Global Convergence of Sharpness Minimization.
International Conference on Machine Learning (ICML). PMLR. 2024.
- [14] **Zhiyuan Li**, Hong Liu, Denny Zhou, Tengyu Ma.
Chain of Thought Empowers Transformers to Solve Inherently Serial Problems.
The Twelfth International Conference on Learning Representations (ICLR). 2024.
- [15] **Zhiyuan Li**, Yi Wang, Zhiren Wang. (α - β)
Fast Equilibrium of SGD in Generic Situations.
The Twelfth International Conference on Learning Representations (ICLR). 2024.
- [16] Kaifeng Lyu, Jikai Jin, **Zhiyuan Li**, Simon S Du, Jason D Lee, Wei Hu.
Dichotomy of Early and Late Phase Implicit Biases Can Provably Induce Grokking.
The Twelfth International Conference on Learning Representations (ICLR). 2024.
- [17] Hong Liu, **Zhiyuan Li**, David Hall, Percy Liang, Tengyu Ma.
Sophia: A Scalable Stochastic Second-Order Optimizer for Language Model Pre-Training.
The Twelfth International Conference on Learning Representations (ICLR). 2024.
- [18] Runzhe Wang, Sadhika Malladi, Tianhao Wang, Kaifeng Lyu, **Zhiyuan Li**.
The Marginal Value of Momentum for Small Learning Rate SGD.
The Twelfth International Conference on Learning Representations (ICLR). 2024.
- [19] Khashayar Gatmiry, **Zhiyuan Li**, Tengyu Ma, Sashank Reddi, Stefanie Jegelka, Ching-Yao Chuang.
What is the Inductive Bias of Flatness Regularization? A Study of Deep Matrix Factorization Models.
Advances in Neural Information Processing Systems 36 (NeurIPS). 2023.
- [20] Kaiyue Wen, **Zhiyuan Li**, Tengyu Ma.

Sharpness Minimization Algorithms Do Not Only Minimize Sharpness to Achieve Better Generalization.

Advances in Neural Information Processing Systems 36 (NeurIPS). 2023. (Oral)

- [21] Jikai Jin, **Zhiyuan Li**, Kaifeng Lyu, Simon Shaolei Du, Jason D Lee.
Understanding Incremental Learning of Gradient Descent: A Fine-Grained Analysis of Matrix Sensing.
Proceedings of the 40th International Conference on Machine Learning (ICML). PMLR. 2023.
- [22] Hong Liu, Sang Michael Xie, **Zhiyuan Li**, Tengyu Ma.
Same Pre-Training Loss, Better Downstream: Implicit Bias Matters for Language Models.
Proceedings of the 40th International Conference on Machine Learning (ICML). PMLR. 2023.
- [23] Kaiyue Wen, Tengyu Ma, **Zhiyuan Li**.
How Does Sharpness-Aware Minimization Minimize Sharpness?
The Eleventh International Conference on Learning Representations (ICLR). 2023.
- [24] **Zhiyuan Li**, Tianhao Wang, Dingli Yu.
Fast Mixing of Stochastic Gradient Descent with Normalization and Weight Decay.
Advances in Neural Information Processing Systems 35 (NeurIPS). 2022.
- [25] Zhiyuan Li*, Tianhao Wang*, Jason D Lee, Sanjeev Arora.
Implicit Bias of Gradient Descent on Reparametrized Models: On Equivalence to Mirror Descent.
Advances in Neural Information Processing Systems 35 (NeurIPS). 2022.
- [26] Kaifeng Lyu, **Zhiyuan Li**, Sanjeev Arora.
Understanding the Generalization Benefit of Normalization Layers: Sharpness Reduction.
Advances in Neural Information Processing Systems 35 (NeurIPS). 2022.
- [27] **Zhiyuan Li**. “Bridging Theory and Practice in Deep Learning: Optimization and Generalization”. PhD thesis. Princeton University, 2022.
- [28] Sanjeev Arora, **Zhiyuan Li**, Abhishek Panigrahi. (α - β)
Understanding Gradient Descent on the Edge of Stability in Deep Learning.
Proceedings of the 39th International Conference on Machine Learning (ICML). PMLR. 2022.
- [29] **Zhiyuan Li**, Srinadh Bhojanapalli, Manzil Zaheer, Sashank Reddi, Sanjiv Kumar.
Robust Training of Neural Networks Using Scale Invariant Architectures.
Proceedings of the 39th International Conference on Machine Learning (ICML). PMLR. 2022. (Long Presentation)
- [30] Kaifeng Lyu*, Zhiyuan Li*, Runzhe Wang*, Sanjeev Arora.
Gradient Descent on Two-Layer Nets: Margin Maximization and Simplicity Bias.
Advances in Neural Information Processing Systems 34 (NeurIPS). 2021.
- [31] **Zhiyuan Li**, Tianhao Wang, Sanjeev Arora.
What Happens after SGD Reaches Zero Loss?—A Mathematical Framework.
The Tenth International Conference on Learning Representations (ICLR). 2022. (Spotlight)
- [32] Simon S Du, Wei Hu, **Zhiyuan Li**, Ruoqi Shen, Zhao Song, Jiajun Wu.
When Is Particle Filtering Efficient For Planning In Partially Observed Linear Dynamical Systems?
Proceedings of the 37th Conference on Uncertainty in Artificial Intelligence (UAI). PMLR. 2021.
- [33] Yaqi Duan, Chi Jin, **Zhiyuan Li**. (α - β)
Risk Bounds And Rademacher Complexity In Batch Reinforcement Learning.
Proceedings of the 38th International Conference on Machine Learning (ICML). PMLR. 2021.
- [34] **Zhiyuan Li**, Sadhika Malladi, Sanjeev Arora.

- On The Validity Of Modeling Sgd With Stochastic Differential Equations (SDEs).**
Advances in Neural Information Processing Systems 34 (NeurIPS). 2021.
- [35] **Zhiyuan Li**, Yuping Luo, Kaifeng Lyu. (α - β)
Towards Resolving The Implicit Bias Of Gradient Descent For Matrix Factorization: Greedy Low-Rank Learning.
The Ninth International Conference on Learning Representations (ICLR). 2021.
- [36] **Zhiyuan Li**, Yi Zhang, Sanjeev Arora.
Why Are Convolutional Nets More Sample-Efficient Than Fully-Connected Nets?
The Ninth International Conference on Learning Representations (ICLR). 2021. **(Oral)**
- [37] Xiaoxia Wu, Edgar Dobriban, Tongzheng Ren, Shanshan Wu, **Zhiyuan Li**, Suriya Gunasekar, Rachel Ward, Qiang Liu.
Implicit Regularization And Convergence For Weight Normalization.
Advances in Neural Information Processing Systems 33 (NeurIPS). 2020.
- [38] Zhiyuan Li*, Kaifeng Lyu*, Sanjeev Arora.
Reconciling Modern Deep Learning With Traditional Optimization Analyses: The Intrinsic Learning Rate.
Advances in Neural Information Processing Systems 33 (NeurIPS). 2020.
- [39] Wei Hu, **Zhiyuan Li**, Dingli Yu.
Simple and Effective Regularization Methods for Training on Noisily Labeled Data with Generalization Guarantee.
The Eighth International Conference on Learning Representations (ICLR). 2020.
- [40] Sanjeev Arora, Simon S Du, **Zhiyuan Li**, Ruslan Salakhutdinov, Ruosong Wang, Dingli Yu. (α - β)
Harnessing the power of infinitely wide deep nets on small-data tasks.
The Eighth International Conference on Learning Representations (ICLR). 2020. **(Spotlight)**
- [41] **Zhiyuan Li**, Sanjeev Arora.
An Exponential Learning Rate Schedule for Deep Learning.
The Eighth International Conference on Learning Representations (ICLR). 2020. test. **(Spotlight)**
- [42] Zhiyuan Li*, Ruosong Wang*, Dingli Yu*, Simon S Du, Wei Hu, Ruslan Salakhutdinov, Sanjeev Arora.
Enhanced Convolutional Neural Tangent Kernels.
arXiv preprint arXiv:1911.00809 (arXiv). 2019.
- [43] Rohith Kuditipudi, Xiang Wang, Holden Lee, Yi Zhang, **Zhiyuan Li**, Wei Hu, Rong Ge, Sanjeev Arora.
Explaining Landscape Connectivity of Low-Cost Solutions for Multilayer Nets.
Advances in Neural Information Processing Systems 32 (NeurIPS). 2019.
- [44] Sanjeev Arora, Simon S Du, Wei Hu, **Zhiyuan Li**, Russ R Salakhutdinov, Ruosong Wang. (α - β)
On Exact Computation with an Infinitely Wide Neural Net.
Advances in Neural Information Processing Systems 32 (NeurIPS). 2019. **(Spotlight)**
- [45] Sanjeev Arora, Simon Du, Wei Hu, **Zhiyuan Li**, Ruosong Wang. (α - β)
Fine-Grained Analysis of Optimization and Generalization for Overparameterized Two-Layer Neural Networks.
Proceedings of the 36th International Conference on Machine Learning (ICML). PMLR. 2019.
- [46] Sanjeev Arora, **Zhiyuan Li**, Kaifeng Lyu. (α - β)
Theoretical Analysis of Auto Rate-Tuning by Batch Normalization.
The Seventh International Conference on Learning Representations (ICLR). 2019.
- [47] Behnam Neyshabur, **Zhiyuan Li**, Srinadh Bhojanapalli, Yann LeCun, Nathan Srebro.

Towards Understanding the Role of Over-Parametrization in Generalization of Neural Networks.

The Seventh International Conference on Learning Representations (ICLR). 2019.

- [48] Elad Hazan, Wei Hu, Yuanzhi Li, **Zhiyuan Li**. (α - β)
Online Improper Learning with an Approximation Oracle.
Advances in Neural Information Processing Systems 31 (NeurIPS). 2018.
- [49] **Zhiyuan Li**, Yicheng Liu, Pingzhong Tang, Tingting Xu, Wei Zhan. (α - β)
Stability of Generalized Two-Sided Markets with Transaction Thresholds.
Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS). 2017. **(Best Paper Nomination)**
- [50] Dylan J Foster, **Zhiyuan Li**, Thodoris Lykouris, Karthik Sridharan, Eva Tardos. (α - β)
Learning in Games: Robustness of Fast Convergence.
Advances in Neural Information Processing Systems 29 (NeurIPS). 2016.
- [51] Yexiang Xue, **Zhiyuan Li**, Stefano Ermon, Carla P Gomes, Bart Selman.
Solving Marginal MAP Problems with NP Oracles and Parity Constraints.
Advances in Neural Information Processing Systems 29 (NeurIPS). 2016.

TALKS

PENCIL: LONG THOUGHTS WITH SHORT MEMORY

- Research at TTIC, February, 2025
- MoDL Workshop, Simons Institute, February, 2025
- Annual Workshop, IDEAL Institute, June, 2025

Weak-to-Strong Generalization Even in Random Feature Networks, Provably

- EnCORE Institute, March, 2025
- PLI Seminar, Princeton University, March, 2025

New Frontiers of Deep Learning Theory in the Era of Language Models

- Research at TTIC, April, 2024

Chain of thought empowers transformers to solve inherently serial problems

- SPARK Seminar, Google, Nov, 2024
- Seminar at Special Year on Large Language Models and Transformers, Simons Institute, Nov, 2024
- Mathematics Colloquia and Seminars, UC Davis, Dec, 2024
- Formal Languages and Neural Networks Seminar (FLANN), August, 2024

Sharpness Minimization Algorithms Do Not Only Minimize Sharpness to Achieve Better Generalization

- Oral talk, NeurIPS, December, 2023

SOPHIA: A Scalable Stochastic Second-Order Optimizer for Language Model Pre-Training

- Workshop on Theory and Practice of Foundation Models, Institute for Foundations of Data Science, Yale University, October, 2023

Generalization Benefit of Flatness Regularization

- FAI Seminar, September, 2023

How Does Sharpness-Aware Minimization Minimizes Sharpness?

- Contributed talk, NeurIPS, December, 2022
- INFORMS, 2023
- Workshop on Machine Learning Theory and Foundations, MSRA, March, 2023

Robust Training of Neural Networks Using Scale Invariant Architectures

- ICML long presentation, July, 2022

Understanding Gradient Descent on Edge of Stability in Deep Learning

- RWTH Aachen University, November, 2022
- Workshop on Representational Learning Theory, TTIC, August, 2022

What Happens after SGD Reaches Zero Loss? –A Mathematical Framework

- Poggio Lab Meeting, MIT, March, 2022
- Foundations Of Data Science, IIIS seminar, Tsinghua University, March, 2022

Reconciling Modern Deep Learning with Traditional Optimization Analyses: The Intrinsic Learning Rate

- UCLA Big Data and ML seminar, March, 2021
- Mathematical Machine Learning Seminar MPI, Feb, 2021

Why Are Convolutional Nets More Sample-Efficient than Fully-Connected Nets?

- ICLR, May, 2021
- International Joint Conference on Theoretical Computer Science, August, 2020

On Exact Computation with an Infinitely Wide Neural Net

- Asilomar Conference on Signals, Systems, and Computers, Asilomar, November, 2019

An Exponential Learning Rate Schedule For Deep Learning

- ICLR, April, 2020
- IIIS-Haihua Frontier Seminar Series, Tsinghua University, December, 2019
- Computer Science Department, Princeton University, November, 2019
- Workshop on Theory of Deep Learning: Where next? Institute for Advanced Study, October, 2019

Theoretical Analysis of Auto Rate-Tuning by Batch Normalization

- Computer Science Department, Princeton University, July, 2019

Fine-Grained Analysis of Optimization and Generalization for Overparameterized Two-Layer Neural Networks

- ICML, Long Beach, June, 2019

Online Improper Learning with an Approximation Oracle

- Institute for Advanced Study, April, 2018

AWARDS

APR. 2025	OpenAI Superalignment Fast Grants
JUNE. 2020-2022	Microsoft Research PhD Fellowship
SEPT. 2017	The William G. Bowen Merit Fellowship, Princeton University
MAY. 2017	AAMAS Best Paper Nomination
JUNE. 2016	Excellent Graduate Award, Tsinghua University (top 10%)
NOV. 2016	Outstanding Student Leader, Tsinghua University
SEPT. 2016	Bronze Medal, Yao Award
OCT. 2015	CASC Scholarship
OCT. 2014	Kwang-Hua Scholarship
JAN. 2013	Second Prize in Chinese Mathematics Olympiad (CMO)

PROFESSIONAL SERVICES

Organization Committee on :

- 1st and 2nd Workshop on Mathematics of Modern Machine Learning, NeurIPS 2023,2024
- Midwest Machine Learning Symposium, 2025
- Understanding the Mechanisms of Deep Learning and Generative Modeling, IDEAL Institute, 2025

Conference Area Chair for:

- NeurIPS 2023,2024,2025
- ICML 2025

Conference Reviewer for:

- NeurIPS(NIPS) 2016,2019,2020,2021
- ALT 2019
- COLT 2019,2020,2021,2022
- AISTATS 2020
- ICML 2020,2021,2023
- ICLR 2023,2024,2025
- STOC 2022,2025

Journal Reviewer for:

- Journal of Machine Learning Research (JMLR)
- Machine Learning
- IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)

TEACHING

Instructor-of-Record, TTIC & University of Chicago

- TTIC 31120 Statistical and Computational Learning Theory, Spring 2025
- STAT 31015 Mathematical Computation IIA: Convex Optimization, Winter 2024

Teaching Assistant, Princeton University

- COS511: Theoretical Machine Learning, Fall 2018
- ELE524: Foundations of Reinforcement Learning, Spring 2020

Guest Lecturer, Princeton University

- COS597G: Theoretical Foundations of Deep Learning, Fall 2018
- COS598B: Mathematical Understanding of Deep Learning, Spring 2021