Hadi Daneshmand

1-857-799-0358 ◇ https://scholar.google.com/citations?user=roFM8XsAAAAJ&hl=en ◇ dhadi@virginia.edu

Research Interests

General: Foundations of Machine Learning
Specific: Foundations of Generative AI
Applications: In-context learning with large language models, and generative models

Academic Appointments

University of Virginia, Assistant Professor of Computer Science *USA, Since 2024*
Massachusetts Institute of Technology and Boston University, Postdoctoral Researcher *USA, 2022-24*
Princeton University, Postdoctoral Fellow *Princeton, USA, 2022*
French Institute for Research in Computer Science, Postdoctoral Researcher *Paris, France, 2020-22*

Education

ETH Zurich, PhD in Computer Science *Switzerland, 2014-2020*
Sharif University of Technology, MS in Artificial Intelligence *Iran, 2011-2014*
Sharif University of Technology, BS in Computer Engineering *Iran, 2007-2011*

Research Experience

Massachusetts Institute of Technology and Boston University USA, Since 2022
Postdoctoral researcher, mentors: *Professor Suvrit Sra* and *Professor Francesco Orabona*
Recipient of a FODSI (Foundations Of Data Science Institute) postdoctoral fellowship

(a) Dec 2022- Dec 2023 Postdoctoral associate at MIT and visiting scholar at Boston University

(b) Since Dec 2023: Research affiliate at MIT and Postdoctoral associate at Boston University

Princeton University *USA, 2022*
Postdoctoral fellow
Recipient of early postdoc mobility grant of Swiss National Science Foundation

French Institute for Research in Computer Science and Automation (INRIA) *France, 2020-22*
Postdoctoral researcher, mentor: *Professor Francis Bach*

ETH Zurich Switzerland, 2014-2020
Graduate research assistant, advisor: *Professor Thomas Hofmann*
Thesis: Optimization for Neural Networks: Quest for Theoretical Understandings
Committee: *Professor Francis Bach* and *Professor Andreas Krause*

Vector Institute at the University of Toronto *Canada, 2019*
Research intern, mentor: *Professor Murat A. Erdogdu*
Research on Markov chain theory: Non-asymptotic central limit theorem for discretized diffusion processes

Max Planck Institute for Intelligent Systems *Germany, 2014*
Research intern, mentor: *Professor Bernhard Scholkopf*
Research on sample complexity of graph inference from information cascade

Award

Research

Stanford CPAL Rising Star Award *2025*
Conference on Parsimony and Learning (CPAL)

Spotlight Award of ICML Workshop on In-context Learning *2024*
For paper (20) in publications on large language models

Postdoctoral Fellowship (148K USD) of Foundations of Data Science Institute *2023*
Outputs: papers (17), (18), (19) and (20) in publications

Early Postdoc Mobility Grant (86K USD), Swiss National Science Foundation *2020*
Proposal: bridging the gap between local and global optimization in machine learning
Outputs: papers (15) and (16) in publications

Best Poster Award *2016*
Max Planck–ETH center for learning systems, Deep Learning Workshop

Service

International Conference on Machine Learning, Reviewer Award *Baltimore, USA, 2022*
Neural Information Processing Systems, Reviewer Award *Virtual, 2020*
International Conference on Machine Learning, Reviewer Award *Long Beach, USA, 2019*

Publications

* equal contributions, ◇ top publications, Google Scholar

▷ **Theory of Language Models**

◇(21) Linear Transformers Implicitly Discover Unified Numerical Algorithms
Patrick Lutz, Aditya Gangrade, Hadi Daneshmand, and Venkatesh Saligrama.
Conference on Neural Information Processing Systems 2025

◇(20) Transformers Learn Temporal Difference Methods for In-Context Reinforcement Learning
Jiuqi Wang*, Ethan Blaser*, Hadi Daneshmand, and Shangtong Zhang.
International Conference on Learning Representations 2025
◇*Special recognition:* Spotlight Award of ICML Workshop on In-context Learning (2024)

(19) Towards Training Without Depth Limits: Batch Normalization Without Gradient Explosion
Alexandru Meterez, Amir Joudaki, Francesco Orabona, Alexander Immer, Gunnar Rätsch and Hadi Daneshmand
International Conference on Learning Representations 2024

◇(18) Transformers Learn to Implement Preconditioned Gradient Descent for In-context Learning
Kwangjun Ahn*, Xiang Cheng*, Hadi Daneshmand* and Suvrit Sra
Conference on Neural Information Processing Systems 2023

(17) On the Impact of Activation and Normalization in Obtaining Isometric Embeddings at Initialization
Amir Joudaki, Hadi Daneshmand and Francis Bach
Conference on Neural Information Processing Systems 2023

Publications

▷ **Beyond Theoretical Mean-Field Neural Networks:** *Bridging the gap between theory and practice*

(16) Efficient Displacement Convex Optimization with Particle Gradient Descent
Hadi Daneshmand, Jason D Lee and Chi Jin
International Conference on Machine Learning 2023

(15) On Bridging the Gap between Mean Field and Finite Width in Deep Random Neural Networks with Batch Normalization
Amir Joudaki, Hadi Daneshmand and Francis Bach
International Conference on Machine Learning 2023

◇(14) Batch Normalization Orthogonalizes Representations in Deep Random Networks
Hadi Daneshmand, Amir Joudaki and Francis Bach
Conference on Neural Information Processing Systems 2021
◇*Special recognition:* This work was spotlighted among the top 3% of submissions

▷ **Bridging Optimization and Integration**

(13) Rethinking the Variational Interpretation of Nesterov's Accelerated Method
Peiyuan Zhang*, Antonio Orvieto* and Hadi Daneshmand
Conference on Neural Information Processing Systems 2021

(12) Revisiting the Role of Euler Numerical Integration on Acceleration and Stability in Convex Optimization
Peiyuan Zhang, Antonio Orvieto, Hadi Daneshmand, Thomas Hofmann, Roy S. Smith
International Conference on Artificial Intelligence and Statistics 2021

▷ **Non-convex Optimization for Neural Networks**

(11) Batch Normalization Provably Avoids Rank Collapse for Randomly Initialised Deep Networks
Hadi Daneshmand*, Jonas Kohler*, Francis Bach, Thomas Hofmann and Aurelien Lucchi
Conference on Neural Information Processing Systems 2020

(10) Optimization for Neural Networks: Quest for Theoretical Understandings
Hadi Daneshmand
PhD Thesis, ETH Zurich 2020

(9) Exponential convergence rates for Batch Normalization: The power of length-direction decoupling in non-convex optimization
Jonas Kohler* ,Hadi Daneshmand* Aurelien Lucchi, Ming Zhou , Klaus Neymeyr and Thomas Hofmann
International Conference on Artificial Intelligence and Statistics 2019

◇(8) Local Saddle Point Optimization: A Curvature Exploitation Approach
Leonard Adolphs, Hadi Daneshmand, Aurelien Lucchi and Thomas Hofmann
International Conference on Artificial Intelligence and Statistics 2019

(7) Escaping Saddles with Stochastic Gradients
Hadi Daneshmand*, Jonas Kohler*, Aurelien Lucchi and Thomas Hofmann
International Conference on Machine Learning 2018
◇ *Special recognition:* Elected among the top %8 submissions for a long presentation

Publications

▷ **Efficient Stochastic Optimization for Statistical Learning**

(6) Adaptive Newton method for empirical risk minimization to statistical accuracy
Aryan Mokhtari*, Hadi Daneshmand*, Aurelien Lucchi, Thomas Hofmann and Alejandro Ribeiro
Conference on Neural Information Processing Systems 2016

(5) Starting Small — Learning with Adaptive Sample Sizes
Hadi Daneshmand, Aurelien Lucchi and Thomas Hofmann
International Conference on Machine Learning 2016

▷ **The Inference of Hidden Graphs from Temporal Dynamics**

(4) Inferring causal molecular networks: empirical assessment through a community-based effort
Steven M Hill, Laura M Heiser, . . . , Hadi Daneshmand, . . .
Nature Methods 2016

(3) Estimating Diffusion Network Structure: Recovery Conditions, Sample Complexity, and a Soft-thresholding algorithm
Manuel Gomez Rodriguez, Le Song, Hadi Daneshmand, and Bernhard Scholkopf
Journal of Machine Learning Researches 2016

◇(2) Estimating Diffusion Network Structures: Recovery Conditions, Sample Complexity & Soft-thresholding Algorithm
Hadi Daneshmand, Manuel Gomez Rodriguez, Le Song, and Bernhard Scholkopf
International Conference on Machine Learning 2014
◇ *Special recognition:* Elected among top 18 submissions (out of 1260+) recommended to Journal of Machine Learning Research

(1) A Time-aware Recommender System based on Dependency Network of Items
Hadi Daneshmand, Amin Javari, Seyed Ebrahim Abtahi and Mahdi Jalili
Oxford computer journal 2014

Workshop papers

(1) Data Generation without Function Estimation
Hadi Daneshmand and Ashkan Soleymani
NeurIPS 2025 workshop on Optimization for Machine Learning
◇ *Special recognition:* Invited for a contributed talk (among 6 selected out of +160 submissions)

(2) How Do Transformers Align Tokens?
Hadi Daneshmand
NeurIPS 2025 Workshop on Differentiable Learning of Combinatorial Algorithms

(3) How Does Layer Normalization Improve Deep learning?
Braham Snyder, Hadi Daneshmand, and Chen-Yu Wei
NeurIPS 2025 workshop on Optimization for Machine Learning

Selected Talks

(CPAL) Conference on Parsimony and Learning *Stanford University, 2025*
Title: Learning to Compute

ICTP Seminar Series "Youth in High-Dimensions" *Remote (Italy), 2025*
Title: Understanding Test-time Inference in LLMs

(IOS) INFORMS for optimization *USA, 2024*
Title: In-context learning of linear functions with gradient descent

Talk at University of Edinburgh *UK, 2024*
Title: What makes neural networks statistically powerful, and optimizable?

Finalist Presentation for Vienna Research Groups for Young Investigators Grant *Austria, 2024*
Title: Analyzing Deep Neural Networks Leveraging Stability Theory

Extra Seminar on Artificial Intelligence, University of Groningen *Netherlands, 2024*
Title: What makes neural networks statistically powerful, and optimizable?

Mathematics, Information, and Computation Seminar, New York University *USA, 2023*
Title: Algorithmic View on Neural Information Processing

ISL Colloquium, Stanford University *USA, 2023*
Title: Beyond Theoretical Mean-field Neural Networks

ML Tea Talks, MIT *USA, 2023*
Title: Data representation in deep random neural networks

ML Seminars, Princeton University *USA, 2022*
Title:The power of depth in random neural networks

CIS@MPG Seminar Series, Max Planck Institute for Software Systems *Germany, 2022*
Title: How do neurons learn?

Teaching Experience

Computational Intelligence Lab, ETH Zurich *2015,16,19*
Teaching Assistant for 100+ Students
Recitation and drafting supplementary lecture notes, designing exercises and leading office hours

Deep Learning, ETH Zurich *2017 and 2018*
Teaching Assistant for 100+ Students
Recitation and drafting supplementary lecture notes, grading projects and exams

Machine Learning, ETH Zurich *2016 and 2018*
Teaching Assistant for 100+ Students
Recitation, proposing student projects, writing and grading exams

Machine Learning, Sharif University of Technology *2012*
Teaching Assistant
Recitation and grading exercises

Design and Analysis of Algorithms, Sharif University of Technology *2011*
Teaching Assistant for 100+ Students
Leading a team of 8 teaching assistants, grading student projects and organizing programming workshops

Mentorship

Amir Joudaki, PhD at ETH Zurich *2020-23*
Outputs: papers (17), (14) and (15) in publications, admitted to a postdoc at Broad Institute

Peiyuan Zhan, MS at ETH Zurich *2019-20*
Outputs: papers (13) and (14) in publications, joined Yale for PhD

Antonio Orvieto, PhD at ETH Zurich *2019-20*
Outputs: papers (13) and (14) in publications

Jonas Kohler, PhD at ETH Zurich *2018-20*
Outputs: papers (7), (9), and (11) in publications

Leonard Adolphs, MS at ETH Zurich *2019*
Output: paper (8) in publications

Alexandru Meterez, MS Thesis at ETH Zurich *2023*
Output: paper (19), PhD admissions from Harvard University and MIT

Flowers Alec Massimo, MS Thesis at ETH Zurich *2023*
Joined Invidia

Alexandre Bense, MS Thesis at ETH Zurich *2022*
Alireza Amani, Intern at ETH Zurich *2018*

Academic Service

Area Chair:

- Conference on Neural Information Processing Systems 2023, 2024, and 2025
- International Conference on Machine Learning 2025

Co-organizing:

- ICLR 24 Workshop on Bridging the Gap Between Practice and Theory in Deep Learning
- TILOS & OPTML++ seminars at MIT 2023

Reviewer for Journal of Machine Learning Research, Neurocomputing Journal, IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE Transactions on Signal and Information Processing over Networks, Elsevier Journal on Online Social Networks and Media, Conference on Neural Information Processing Systems, International Conference on Machine Learning, Data Mining and Knowledge Discovery, International Conference on Artificial Intelligence and Statistics, and International Conference on Learning Representations.