# Hadi Daneshmand

32 Vassar St. Cambridge, Massachusetts, USA

https://www.mit.edu/~hdanesh/index.html ◇ hadi.daneshmand@gmail.com ◇ hdanesh@mit.edu


## Research Interests

*General:* Machine learning and Artificial Intelligence, Optimization, Stochastic Processes
*Specific:* Foundations of representation learning (algorithmic, functional, and computational aspects)
*Applications:* In-context learning with large language models, Generative models, Image processing with convolutional networks

## Academic Appointments

**Massachusetts Institute of Technology**, Postdoctoral Associate — *Cambridge, USA, Since 2022*
**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** — *USA, Since 2022*
Postdoctoral associate, mentor: *Suvrit Sra*
Recipient of a FODSI (Foundations Of Data Science Institute) postdoctoral fellowship

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

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

**Boston University** — *USA, Since 2022*
Visiting researcher hosted by *Francesco Orabona*

**Princeton University** — *USA, 2022*
Postdoctoral fellow hosted by *Chi Jin*
Recipient of early postdoc mobility grant of Swiss National Science Foundation

**Vector Institute at the University of Toronto** — *Canada, 2019*
Research intern, mentor: *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: *Bernhard Scholkopf*
Research on sample complexity of graph inference from information cascade

## Award

### Research

**Postdoctoral Fellowship of FODSI (Foundations Of Data Science Institute)** *2023*
*Outputs*: papers (17) and (18) in publications, and preprint (2)

**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, and preprint (1)

**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, Google Scholar

▷ **Analyzing the Data Processing in Deep Neural Networks**

(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

▷ **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

## Publications

### ▷ 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

### ▷ 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

## Publications

▷ **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

## Reprints

(2) Towards Training Without Depth Limits: Batch Normalization Without Gradient Explosion
Alexandru Meterez, Amir Joudaki, Francesco Orabona, Alexander Immer, Gunnar Rätsch and Hadi Daneshmand

(1) Polynomial-time sparse measure recovery
Hadi Daneshmand and Francis Bach

## Selected Talks

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

**Machine Learning Seminars, Rensselaer Polytechnic Institute** *Virtual, 2023*
Title: Dynamical isometry — Beyond a mean field theory

**An Invited Talk at The Australian National University** *Virtual, 2023*
Title: Dynamical isometry of data representations in random deep 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

**Winter Seminar Series, Sharif University of Technology** *Virtual, 2022*
Title: Representations in Random Deep Neural Networks

**Spotlight Presentation, Conference on Neural Information Processing Systems** *Virtual, 2022*
Title: Batch normalization orthogonalizes representations in deep random neural networks

**ML Seminars, National Institute for Research in Digital Science and Technology** *France, 2021*
Title: Representations in Random Deep Neural Networks

## 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

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

**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

**Kwangjun Ahn, PhD at MIT** *2022-23*
Output: paper (18) in publications

**Ashkan Soleymani, PhD at MIT** *2023*
In progress

**Alexandru Meterez, MS Thesis at ETH Zurich** *2023*
In progress

**Flowers Alec Massimo, MS Thesis at ETH Zurich** *2023*
In progress

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

## Academic Service

**Area Chair** for Conference on Neural Information Processing Systems 2023
**Organizing** 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.