



Stochastic Gradients with Adaptive Stepsizes



State-of-art deep learning algorithms are constantly evolving and improving, but the core optimization methods used to train neural networks have remained largely stable. Adaptive gradient variations of stochastic gradient descent such as Adagrad and Adam are popular in practice for demonstrating significantly more stable convergence compared to plain stochastic gradient descent, with less hyperparameter tuning required. The speaker will discuss the current state of theoretical understanding of these algorithms, as well as several open questions inspired by persisting gaps between theory and practice.

Rachel Ward, Ph.D.

W. A. “Tex” Moncrief Distinguished Professor
in Computational Engineering and Sciences
Professor of Mathematics
University of Texas at Austin

Date: Monday, October 24, 2022

Time: 2:00 – 3:00 p.m. US Central Time

Location: **Blocker 220** (in person)

Online: 998 4499 3279 (ID) & 724615 (PWD)

Faculty host: Simon Foucart, MATH

Biography

Dr. Rachel Ward is the W. A. “Tex” Moncrief Distinguished Professor in Computational Engineering and Sciences — Data Science and Professor of Mathematics at UT Austin. She is recognized for her contributions to stochastic gradient descent, compressive sensing, and randomized linear embeddings. From 2017-2018, Dr. Ward was a visiting researcher at Facebook AI Research. Prior to joining UT Austin in 2011, Dr. Ward received the PhD in Computational and Applied Mathematics at Princeton in 2009 and was a Courant Instructor at the Courant Institute, NYU, from 2009-2011. Among her awards are the Sloan research fellowship, NSF CAREER award, 2016 IMA prize in mathematics and its applications, 2020 Simons fellowship in mathematics. She is also an invited speaker at the 2022 International Congress of Mathematicians.

You can also click this link to join the seminar <https://tamu.zoom.us/j/99844993279?pwd=TkJodWFVRURyMmkwakl4SWZGeVJTQT09>

