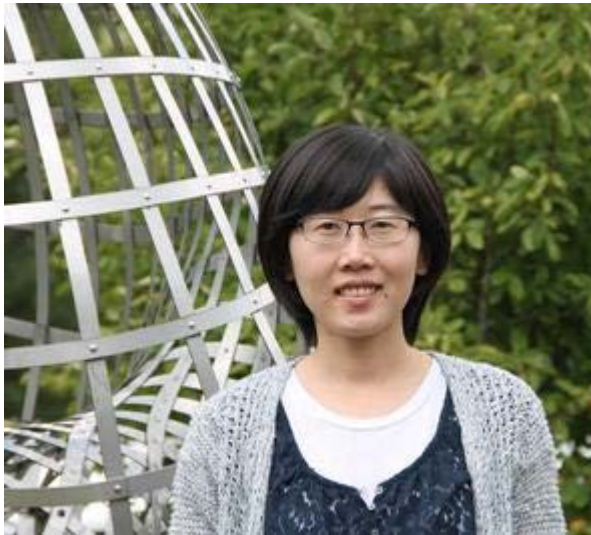


Graduate Seminar



Implicit Regularization in Nonconvex Statistical Estimation

Yuejie Chi

Associate Professor
ECE Department
Carnegie Mellon University

Thursday, March 1st

4:30 PM

WEH 5421

ECE Seminar Committee

Aswin Sankaranarayanan
saswin@ece.cmu.edu

Maysam Chamanzar
mchamanz@andrew.cmu.edu

Swarun Kumar
swarun@cmu.edu

Abstract:

Recent years have seen a flurry of activities in designing provably efficient nonconvex procedures for solving statistical estimation problems. Due to the highly nonconvex nature of the empirical loss, state-of-the-art procedures often require proper regularization (e.g. trimming, regularized cost, projection) in order to guarantee fast convergence. For vanilla procedures such as gradient descent, however, prior theory either recommends highly conservative learning rates to avoid overshooting, or completely lacks performance guarantees.

This talk uncovers a striking phenomenon in nonconvex optimization: even in the absence of explicit regularization, vanilla gradient descent enforces proper regularization implicitly under various statistical models. In fact, gradient descent follows a trajectory staying within a basin that enjoys nice geometry, consisting of points incoherent with the sampling mechanism. This “implicit regularization” feature allows gradient descent to proceed in a far more aggressive fashion without overshooting, which in turn results in substantial computational savings. Focusing on three fundamental statistical estimation problems, i.e. phase retrieval, low-rank matrix completion, and blind deconvolution, we establish that gradient descent achieves near-optimal statistical and computational guarantees without explicit regularization. In particular, by marrying statistical modeling with generic optimization theory, we develop a general recipe for analyzing the trajectories of iterative algorithms via a leave-one-out perturbation argument. As a byproduct, for noisy matrix completion, we demonstrate that gradient descent achieves near-optimal error control — measured entry wise and by the spectral norm — which might be of independent interest.

Bio:

Since January 2018, I am an Associate Professor in the Dept. of Electrical and Computer Engineering at the Carnegie Mellon University, where I hold the Robert E. Doherty Career Development Professorship. I am also an affiliated faculty member in the Machine Learning Department in the School of Computer Science. Previously I was with the Dept. of Electrical and Computer Engineering and the Dept. of Biomedical Informatics at The Ohio State University. I completed my Ph.D. in Electrical Engineering from Princeton University in 2012, where I was fortunate to work with Prof. Robert Calderbank. I received a M.A. in Electrical Engineering from Princeton University in 2009, and a B.Eng. in Electronic Engineering from Tsinghua University in 2007. My research is motivated by the challenge of efficiently extracting information embedded in a large amount of data, as well as collecting data efficiently to gather actionable information. I am interested in the mathematics of data representation that take advantage of structures and geometry to minimize complexity and improve performance. Specific topics include mathematical and statistical signal processing, machine learning, large-scale optimization, sampling and information theory, with applications in sensing, imaging and big data.

I am a recipient of NSF CAREER Award, AFOSR Young Investigator Award, ONR Young Investigator Award, ORAU Ralph E. Powe Junior Faculty Enhancement Award, Google Faculty Research Award, IEEE Signal Processing Society Young Author Best Paper Award, IEEE ICASSP Best Student Paper Award, among others.

SEMINAR NOTES: (REFRESHMENTS SERVED AT 4:00 PM)