A Dynamical Perspective on Accelerated Methods

Ashia Wilson

October 8, 2020

Abstract: Optimization is a core primitive in statistics and machine learning. Within these fields, the rapid growth in the scale and complexity of modern datasets has led to a focus on gradient-based algorithms. Accelerated methods, first proposed by Nesterov in 1983, achieve faster convergence rates than many gradient-based descent methods. In this talk, we discuss accelerated methods from a dynamical systems perspective. We show there is a Lagrangian functional that generates a large class of accelerated methods in continuous time. The continuous-time perspective reveals that most accelerated methods correspond to traveling the same curve in spacetime at different speeds. We also explore how different discretization techniques yield algorithms with optimal complexity guarantees in optimizations using Lyapunov theory. Finally, if time permits, we will discuss the role of acceleration in non-convex optimization and its application to more complex function spaces. This is joint work with Andre Wibisono, Lester Mackey, Benjamin Recht, and Michael Jorda

Biography: Ashia Wilson will start as an assistant professor in the department of Electrical Engineering and Computer Science at MIT in 2021. Before joining the faculty at MIT, she was a postdoctoral researcher in the Machine Learning Group at Microsoft Research, New England. She received undergraduate degrees in Applied Mathematics and Philosophy from Harvard University in 2011. She received her doctorate in Statistics from the University of California, Berkeley in 2018 advised by Benjamin Recht and Michael I. Jordan. Her research interests are in providing rigorous guarantees for algorithmic performance, and in developing frameworks for studying issues of fairness and governance in machine learning.