EPSTEIN INSTITUTE SEMINAR ▪ ISE 651

Advancing Stochastic Optimization for Reinforcement Learning

ABSTRACT - **Reinforcement Learning (RL)** has attracted considerable interest from both industry and academia recently. The study of RL algorithms with provable rates of convergence, however, is still in its infancy. In this talk, we discuss some recent progresses on the solutions of two fundamental RL problems, i.e., stochastic policy evaluation and policy improvement. It is worth noting that both are related to stochastic variational inequalities.

For **policy evaluation**, prior investigations in the literature focused on temporal difference (TD) learning by employing nonsmooth finite time analysis motivated by stochastic subgradient descent leading to certain limitations. These encompass the requirement of analyzing a modified TD algorithm that involves projection to an a-priori defined Euclidean ball, achieving a non-optimal convergence rate and no clear way of deriving the beneficial effects of parallel implementation. We address all these issues by developing novel analysis of TD, and introducing new algorithms, including conditional TD algorithm (CTD) and fast TD (FTD) algorithm to achieve the best-known so-far convergence rate for policy evaluation.

For **policy improvement**, we present new policy mirror descent (PMD) methods for solving RL problems with either strongly convex or general convex regularizers. By exploring the structural properties of these overall seemly highly nonconvex problems, we show that the PMD methods exhibit fast linear rate of convergence to the global optimality. We develop stochastic counterparts of these methods, and establish an $O(1/\varepsilon)$ (resp., $O(1/\varepsilon^2)$) sampling complexity for solving these RL problems with strongly (resp., general) convex regularizers using different sampling schemes (one involving CTD), where ε denote the target accuracy. We further show the tight complexity for computing the gradients of these regularizers, if necessary. To the best of our knowledge, these complexity bounds, along with our algorithmic developments, appear to be new in both optimization and RL literature. The introduction of convex regularizers also greatly expands the flexibility and applicability of RL models.
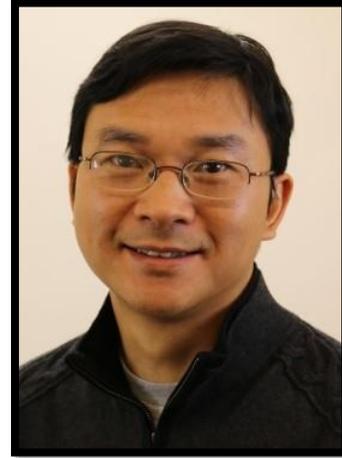
Dr. Guanghui (George) Lan
Associate Professor
The H. Milton Stewart School of
Industrial and Systems Engineering
Georgia Institute of Technology

SPEAKER BIO – Guanghui (George) Lan is an Associate Professor in the H. Milton Stewart School of Industrial and Systems Engineering at Georgia Institute of Technology since January 2016. Dr. Lan was on the faculty of the Department of Industrial and Systems Engineering at the University of Florida from 2009 to 2015, after earning his Ph.D. degree from Georgia Institute of Technology in August 2009. His main research interests lie in optimization and machine learning. The academic honors he received include the Mathematical Optimization Society Tucker Prize Finalist (2012), INFORMS Junior Faculty Interest Group Paper Competition First Place (2012) and the National Science Foundation CAREER Award (2013). Dr. Lan serves as an associate editor for Mathematical Programming, SIAM Journal on Optimization and Computational Optimization and Applications. He is also an associate director of the Center for Machine Learning at Georgia Tech.

USC Viterbi
School of Engineering
Daniel J. Epstein Department of
Industrial and Systems Engineering

TUESDAY, FEBRUARY 16, 2021
3:30 PM – 4:50 PM
ZOOM/ONLINE *PLEASE EMAIL OWH@USC.EDU FOR PASSWORD*