WENLONG MOU, University of Toronto

Continuous-time reinforcement learning: blessings of elliptic structures and high-order approximations

Reinforcement learning (RL) for controlling continuous-time diffusion processes has attracted significant research interest in recent years. A key challenge is accurately estimating the value function for an unknown system, given only discretely observed trajectory data. While model-free RL methods offer the flexibility of advanced function approximations, they struggle with long effective horizons and lack the precision of model-based approaches.

In this talk, I present recent developments in the design of continuous-time policy evaluation algorithms, introducing a novel class of Bellman equations. These methods integrate the flexibility of RL techniques with the precision of high-order numerical schemes. Among other results, I will highlight how the underlying elliptic structures provide strong theoretical guarantees, even as the effective horizon extends to infinity. Finally, I will discuss how these theoretical insights inform practical algorithmic design.