Contribution ID: **69** Type: **not specified**

Score-Aware Policy-Gradient Methods and Performance Guarantees using Local Lyapunov Conditions

Wednesday, June 19, 2024 2:00 PM (30 minutes)

In this talk, we introduce a policy-gradient method for model-based Reinfocement Learning (RL) that exploits a type of stationary distribution commonly obtained from Markov Decision Processes (MDPs) in stochastic networks, queueing systems and statistical mechanics.
Specifically, when the stationary distribution of the MDP belongs to an exponential family that is parametrized by policy parameters, we can improve existing policy gradient methods for average-reward RL.
Our key identification is a family of gradient estimators, called Score-Aware Gradient Estimators (SAGEs), that in the aforementioned setting, enable policy gradient estimation without relying on value-function approximation.
This contrasts with other common policy-gradient algorithms, such as actor-critic methods.
We first show that policy-gradient with SAGE locally converges, including in cases when the objective function is nonconvex, presents multiple maximizers, and the state space of the MDP is not finite. Under appropriate assumptions such as starting sufficiently close to a maximizer, the policy under stochastic gradient ascent with SAGE has an overwhelming probability of converging to the associated optimal policy.
Other key assumptions are that a local Lyapunov function exists, and a nondegeneracy property of the Hessian of the objective function holds locally around a maximizer.
Furthermore, we conduct a numerical comparison between a SAGE-based policy-gradient method and an actor-critic method.
We specifically focus on several examples inspired from stochastic networks, queueing systems, and models derived from statistical phyiscs, where parameterizable exponential families are commonplace.
Our results demonstrate that a SAGE-based method finds close-to-optimal policies faster than an actor-critic method.

Primary authors: SENEN–CERDA, Albert (IRIT, LAAS–CNRS, and Université de Toulouse); COMTE, Céline (CNRS and LAAS); SANDERS, Jaron (Eindhoven University of Technology); JONCKHEERE, Matthieu (LAAS–CNRS)

Presenter: SENEN–CERDA, Albert (IRIT, LAAS–CNRS, and Université de Toulouse)

Session Classification: Parallel session: Online learning