

Evaluating Model Free Policy Optimization Strategies for Non Linear Systems

Aditya Chukka
Computer Science

4/14/2017 at 03:00 pm
CBIM 22

Abstract

The Iterative Linear Quadratic Regulator (ILQR), a variant of Differential Dynamic Programming (DDP) is a tool for optimizing both open-loop trajectories and guiding feedback controllers using dynamics information that can be inferred from data. This technique assumes linear dynamics and quadratic cost functions and improves the control policy iteratively until convergence. We demonstrate the capabilities of this framework in designing controllers for regulating both natural and custom behavior on a simple pendulum, the primitive non linear system. The method's assumptions limit its validity to smaller regions of the state space. Direct Policy Search methods use Reinforcement Learning to develop controllers for such scenarios. Nevertheless, these methods require numerous samples to generate an optimal policy and often converge to poor local optima.

Guided Policy Search (GPS) is a new technique that optimizes complex non-linear policies, such as those represented through deep neural networks, without computing policy gradients in high dimensional parameter space. It trains the policy in a "supervised" fashion using numerous locally valid controllers produced by ILQR. GPS provides appealing improvement and convergence guarantees in simple convex and linear settings and bounds the error in a non-linear setting. We apply Guided Policy Search to generate control policies for locomotion of a tensegrity robot, producing closed-loop motion that could not be achieved with previous methods.