503d Input Output Optimal Control of Multi-Scale Systems Using Adaptive Policy Iteration

Eduardo J. Dozal-Mejorada and B. Erik Ydstie

Linear quadratic (LQ) controllers for unconstrained systems can be developed when a plant model is available. The most common method is to generate the state feedback controller by solving the corresponding Riccati equations. In practice this controller has to be combined with a method for optimal state estimation.

In many practical situations the model may have very high order or it may not be known precisely. In these cases it is desirable to develop adaptive methods that find the optimal controller and/or approximations thereof which do not require explicit modeling and solution of the Riccati equations. Dynamic Programming (DP) provides a basis for compiling planning results into reactive strategies for real-time control which can be enhanced with learning. Such strategies allow control even when the system model is not completely known. Barto et. al. (1995) describe various DP-based reinforcement learning algorithms such as Sutton's Temporal Difference methods, Watkins' Q-learning, and Werbos' Heuristic DP. The main application areas for these methods have been for systems with a discrete state space (Markov decision problems).

In 1994, Bradtke introduced the Adaptive Policy Iteration (API) algorithm based on Q-learning and Policy Iteration and successfully solved the problem of adaptive LQ regulation with state feedback. He proved that the algorithm converges to the optimal controller provided that the underlying system is controllable, that there are no disturbances and that a particular signal vector is persistently excited. A Kalman filter or full state information was still needed for implementation. This severely limits the applicability of the method, especially in the multi-scale case when it is not clear which states are important to measure and models are not available for state estimation.

The approach we develop in this paper is motivated by Bradtke's method. However, in order to address the multi-scale and reduced order problems, we have developed methods that use input output data to obtain highly reduced order representations of the optimal feedback strategies and optimal controls that do not require Kalman filtering. This leads to a novel approach for DP-based reinforcement learning which should work well for systems where most modes have dynamics that dissipate quickly. The use of input output data allows us to focus modeling attention on the low order (observable and controllable) dynamics while the fast (poorly observable) dynamics are allowed to drift.

The algorithm itself uses Watkins' Q-learning algorithm implemented with API. The API algorithm estimates a stochastic ARMAX model. The resulting model represents a low order approximation to a high order plant. In this work, we use recursive least squares as our regression tool; however any constrained optimization procedure can be used for this step.

The main contributions of this paper are: (1) The formulation of input output optimal controller using Q-functions and the application of low order control for high- and infinite order problems. (2) Mathematical proofs of stability and convergence. These show that that the algorithm converges asymptotically to the optimal controller within a certain tolerance which can be related to excitation level, magnitude of model mismatch and disturbances. (3) We show how the approach can be generalized to nonlinear (convex) programs and some classes of constrained nonlinear programs. (4) We describe example simulation problems.

References

1. A. G. Barto, S. J. Bradtke, and S. P. Singh, "Learning to act using real-time dynamic programming", Artificial Intelligence 72(1):81--138 (1995)

2. S. J. Bradtke, "Incremental Dynamic Programming for On-line Adaptive Optimal Control", Thesis, University of Massachusetts, Amherst (1994).