CC PUB #296
November 1999

Piecewise Stationary Markov Decision Processes, I: Constant Gain

M. Jacobson, N. Shimkin and A. Shwartz
Department of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 32000, Israel

Abstract

We consider a class of non-stationary dynamic decision-making models to which we refer as Piecewise Stationary Markov Decision Processes (PSMDPs). In these models, the decision making horizon can be partitioned into intervals, called *renewal cycles*, of $N+1$ epochs. The transition law and reward function are identical over the first N epochs of each renewal cycle, but distinct at the final epoch. The motivation for these models is in applications where decisions of different nature are taken at different time scales, i.e. many "low-level" decisions are made between "high-level" ones.

Our aim is to characterize solutions of the discounted reward optimality problem for large values of N, with the effective discount rate over an entire renewal cycle held fixed. In this model, *initially stationary* policies are natural candidates for optimal policies. Similar to turnpike policies, an initially stationary policy uses a fixed decision rule for some large number of epochs in each renewal cycle, followed by a relatively short planning horizon of time-varying decision rules.

Our analysis relates the PSMDP to an average reward MDP defined by the stationary part of the system. We focus here on the constant gain case, where this MDP's optimal average reward is independent of the initial state. We find that the optimal value of the PSMDP can be fully characterized. It is shown that initially stationary policies are ϵ-optimal under weak conditions and require a planning horizon whose length is bounded in N. We further identify conditions under which these policies are precisely optimal. The non-constant gain case is briefly considered here via examples, and further analyzed in the companion paper.