Nihonkai Math. J.
Vol.9 (1998), 53–61

On some continuous time discounted Markov decision process

Mitsuhiro Hoshino

Department of Mathematical Science, Graduate School of Science and Technology, Niigata University, Niigata, 950–21, Japan

Abstract

In this paper, we describe a basic minimization problem with respect to a continuous time Markov decision process with non-stationary transition probability rates, a general state space and a general action space. We establish the existence of solutions of the optimality equation which plays an important role in the analysis of the minimization problem.

AMS classification: 90C40
Keywords: Markov decision process, optimality equation, contraction mapping.

1. Introduction and formulation of Markov decision processes

Continuous time Markov decision processes have been investigated by many authors, e.g., Miller (1968), Kakumanu (1971), Doshi (1976), Lai and Tanaka (1991) and Qiying (1993). Miller (1968) investigated the case of a finite state space. Kakumanu (1971) and Qiying (1993) extended Miller's results to the case of a countable state space and a countable action space. Doshi (1976) discussed the case of a general state space and a general action space.

In this paper, a constrained Markov decision process as a dynamic programming model is specified by a set of seven elements $(\mathfrak{X}, \mathcal{A}, T, r, \Pi, p_\pi, \alpha)$. We assume the following. The *state space* $\mathfrak{X}$ is a nonempty Borel subset of a Polish (that is, complete separable metrizable) space with Borel σ-algebra $\mathcal{B}_{\mathfrak{X}}$, the set of states of the dynamic decision system. The *action space* $\mathcal{A}$ is a nonempty Borel subset of a Polish space, the set of actions of the decision system. $T = [0, t^*]$ is the *time set* with $t^* < +\infty$. The decisions are continuously taken on the time set. The *loss rate function* r is a bounded measurable real-valued function on $T \times \mathfrak{X} \times \mathcal{A}$ with a bound M. Throughout this paper, we confine ourselves to Markov policies. A Markov policy $\pi = \pi(A|t,x)$ is a Borel measurable stochastic kernel on $\mathcal{A}$ for fixed $(t,x) \in T \times \mathfrak{X}$, that is, $\pi(\cdot|t,x)$ is a probability measure on $\mathcal{A}$ for each $(t,x) \in T \times \mathfrak{X}$ and $\pi(A|\cdot,\cdot)$ is a measurable function on $T \times \mathfrak{X}$ for each Borel set A. Π is the set of all *admissible policies* and consists of Markov policies. p_π is a *non-stationary transition probability function* under a policy $\pi \in \Pi$, that is, $p_\pi(s,x;t,\Gamma)$ is defined for $0 \leq s \leq t < +\infty$, $x \in \mathfrak{X}$ and $\Gamma \in \mathcal{B}_{\mathfrak{X}}$, each $p_\pi(s,x;t,\cdot)$ is a probability measure on $\mathfrak{X}$ with $p_\pi(s,x;s,\{x\}) = 1$, each $p_\pi(s,\cdot;t,\Gamma)$ is a measurable function, and p_π satisfies the Chapman-Kolmogorov equation. α is a nonnegative constant, the *discount rate* for the loss.