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 $0 \le s \le t < +\infty$, $x \in \mathfrak{X}$ 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.

The dynamic decision system is interpreted as follows. Suppose a policy $\pi \in \Pi$ is employed. If the system is in a state $x_t \in \mathfrak{X}$ at a time t, then the system at the time $t+\Delta t$ transfers to a new state $x_{t+\Delta t} \in \mathfrak{X}$ which is governed by transition probability measure $p_\pi(t, x_t; t+\Delta t, \cdot)$ and, as a result, the state $x_{t+\Delta t}$ and an action $a \in \mathcal{A}$ determined by probability $\pi(da|t+\Delta t, x_{t+\Delta t})$ cause the controller to incur continuously a loss according to loss rate $r(t+\Delta t, x_{t+\Delta t}, a)$.

We give the following decision's quality. We define a *loss rate function* r_π *under a policy* π defined by $r_\pi(t,x) = \int_{\mathcal{A}} r(t,x,a)\pi(da|t,x)$ for $(t,x) \in T\times\mathfrak{X}$. When the decision system starts from a state x at time t and a policy π is used, then the *total expected discounted loss* is given by

$$V_\pi(t,x) = E_\pi\Big[\int_t^{t^*} e^{-\alpha(s-t)} r_\pi(s, X_s)ds \Big| X_t = x\Big]$$
$$= \int_0^{t^*-t} e^{-\alpha s}\Big\{\int_{\mathfrak{X}} r_\pi(t+s,y)p_\pi(t,x;t+s,dy)\Big\}ds,$$

where E_π is the expectation operator with respect to p_π and $e^{-\alpha}$ means a discount factor. In general, our aim is to get an optimal policy or an ε-optimal policy.

Definition 1.1

(i) The *optimal value function* is given by $V_{\text{opt}}(t,x) = \inf_{\pi\in\Pi} V_\pi(t,x)$;

(ii) A policy $\pi^* \in \Pi$ is called *optimal* if $V_{\pi^*}(t,x) = V_{\text{opt}}(t,x)$ for any $(t,x) \in T\times\mathfrak{X}$;

(iii) For a constant $\varepsilon > 0$, a policy $\pi_\varepsilon \in \Pi$ is called ε-*optimal* if $V_{\pi_\varepsilon}(t,x) \le V_{\text{opt}}(t,x)+\varepsilon$ for any $(t,x) \in T\times\mathfrak{X}$.

2. Optimality equation

The following equation, which is called the *optimality equation*, plays a crucial role in the analysis of the minimization problem.

$$\alpha V(t,x) = \inf_{\pi\in\Pi}\{r_\pi(t,x) + Q_\pi V(t,x)\} + D_t V(t,x), \qquad V(t^*,\cdot) = 0 \text{ on } \mathfrak{X}, \tag{1}$$

where Q_π is a certain operator, transforming a function on $T\times\mathfrak{X}$ into a function on $T\times\mathfrak{X}$, and $D_t V(t,x)$ is the differential coefficient of $V(\cdot,x)$ with respect to t. In order to introduce Q_π, we impose an assumption on the differentiability of p_π.

Assumption A$_1$ For each $\pi \in \Pi$, there exists $q_\pi : T\times\mathfrak{X}\times\mathcal{B}_{\mathfrak{X}} \to \mathbb{R}$ such that each $q_\pi(t,\cdot;\Gamma)$ and $q_\pi(\cdot,x;\Gamma)$ are measurable, and q_π satisfies $\inf_{(t,x)\in T\times\mathfrak{X}} q_\pi(t,x;\{x\}) \ge -c$ for some constant $c > 0$ and

$$\limsup_{h\downarrow 0}\sup_{x\in\mathfrak{X}}\sup_{\Gamma\in\mathcal{B}_{\mathfrak{X}}} |h^{-1}\{p_\pi(t,x;t+h,\Gamma) - \delta_x(\Gamma)\} - q_\pi(t,x;\Gamma)| = 0$$

for each $t \in T$, where $\delta_x(\Gamma) = 1$ if $x \in \Gamma$, $\delta_x(\Gamma) = 0$ otherwise.

It is easy to show that function q_π has the following properties.

Lemma 2.1 (i) *If* $x \in \Gamma$, *then* $-c \le q_\pi(t,x;\Gamma) \le 0$. (ii) *If* $x \notin \Gamma$, *then* $0 \le q_\pi(t,x;\Gamma) \le c$. (iii) $q_\pi(t,x;\mathfrak{X}) = 0$. (iv) *Each* $q_\pi(t,x;\cdot)$ *is a finite signed measure on* $\mathfrak{X}$.

We put $\int_{\mathfrak{X}} f(x)\nu(dx) = \int_{\mathfrak{X}} f(y)\nu^+(dx) - \int_{\mathfrak{X}} f(y)\nu^-(dx)$ for a signed measure ν on $\mathfrak{X}$ and a bounded measurable function f on $\mathfrak{X}$, where ν^+ and ν^- are upper and lower variation measures by the Jordan decomposition of ν. Let $B(T\times\mathfrak{X})$ be the Banach space of all bounded measurable real-valued functions on $T\times\mathfrak{X}$ with the supremum norm $||u||$. We define an operator Q_π on $B(T\times\mathfrak{X})$ by

$$Q_\pi u(t,x) = \int_{\mathfrak{X}} u(t,y)q_\pi(t,x;dy)$$

in the sense of the above integral. By Lemma 2.1, we have

$$|Q_\pi u(t,x)| \le \{q_\pi(t,x;\mathfrak{X}\backslash\{x\}) - q_\pi(t,x;\{x\})\}||u|| \le 2c||u||. \tag{2}$$

For $s \in T$, we define an operator P_s^π from $B(T\times\mathfrak{X})$ into itself by

$$P_s^\pi u(t,x) = \int_{\mathfrak{X}} u(t+s,y)p_\pi(t,x;t+s,dy)$$

if $0 \le t \le t^* - s$ and $x \in \mathfrak{X}$, and $P_s^\pi u(t,x) = 0$ if $t^* - s < t \le t^*$ and $x \in \mathfrak{X}$. It is easy to verify that P_0^π is the identity operator, $||P_s^\pi u|| \le ||u||$ and $P_{s_1}^\pi P_{s_2}^\pi = P_{s_1+s_2}^\pi$. The total loss is represented by

$$V_\pi(t,x) = \int_0^{t^*-t} e^{-\alpha s} P_s^\pi r_\pi(t,x)ds. \tag{3}$$

We must make the preparations for giving useful properties of (1). For each function u on $T\times\mathfrak{X}$, let $D(u)$, $D(u,t)$, $D(u,t,x)$ be the set of all $(t,x) \in T\times\mathfrak{X}$ such that $D_t u(t,x)$ exists, the set of all $x \in \mathfrak{X}$ such that $D_t u(t,x)$ exists, the set of all $s \in [0, t^* - t]$ such that $D_t u(t+s,x)$ exists, respectively.

Let $B_P(T\times\mathfrak{X})$ be the set of all $u \in B(T\times\mathfrak{X})$ which satisfies the following conditions (B_1) and (B_2).

(B_1) $F(s) := P_s^\pi u(t,x)$ is an absolutely continuous function on $[0, t^* - t]$ for each $(t,x) \in [0,t^*)\times\mathfrak{X}$ and $\pi \in \Pi$. See Billingsley (1979, Section 31) for the definition and some properties of absolutely continuous functions.

(B_2) $u(\cdot,x)$ is differentiable at almost all $t \in T$ for all $x \in \mathfrak{X}$. For each $t \in T$, $D_t u(t,\cdot)$ is bounded on $D(u,t)$, and there exists a constant $\overline{u}$ such that

$$|h^{-1}\{u(t+h,x) - u(t,x)\}| \le \overline{u}$$

for all $x \in \mathfrak{X}$ and all non-zero h in some neighborhood of 0. $D_t u$ is measurable on $D(u)$.

Assumption A_2 (i) For each $\pi \in \Pi$, $0 \le t < \tau \le t^*$ and $x \in \mathfrak{X}$, there exist a $\delta > 0$ and a finite measure μ on $\mathfrak{X}$ such that $\sup_{0<\zeta<\delta} p_\pi(t,x;\tau-\zeta,\Gamma) \le \mu(\Gamma)$ for any $\Gamma \in \mathcal{B}_{\mathfrak{X}}$.

(ii) $\limsup_{h\downarrow 0} \sup_{x\in\mathfrak{X}} \sup_{\Gamma\in\mathcal{B}_{\mathfrak{X}}} h^{-1}|p_\pi(t-h,x;t,\Gamma) - p_\pi(t,x;t+h,\Gamma)| = 0$ for each $\pi \in \Pi$ and $t \in (0,t^*]$.

Lemma 2.2 *Suppose that (t,x) is in $[0,t^*)\times\mathfrak{X}$ and $u \in B(T\times\mathfrak{X})$ satisfies condition (B_2). If $s \in D(u,t,x)$, then $F(s) := P_s^\pi u(t,x)$ is differentiable at s and*

$$\frac{\partial}{\partial s} P_s^\pi u(t,x) = P_s^\pi Q_\pi u(t,x) + P_s^\pi D_t u(t,x).$$

Proof. First, we show the differentiability of F from the left. Let s be in $D(u,t,x) \cap (0, t^* - t]$. We put

$$A_{-h} = h^{-1}\{F(s) - F(s-h)\} - P_s^\pi Q_\pi u(t,x) - P_s^\pi D_t u(t,x), \qquad h \in (0,s).$$

We have

$$A_{-h} = \int_{\mathfrak{X}} \left[P_h^\pi [h^{-1}u - Q_\pi u - D_t u] - h^{-1}u\right](t+s-h,y) p_\pi(t,x;t+s-h,dy).$$

By using signed measure

$$\mathcal{L}_{t+s}^h(\Gamma) = p_\pi(t+s-h,y;t+s,\Gamma) - \delta_y(\Gamma) - hq_\pi(t+s,y;\Gamma)$$

on $\mathfrak{X}$ in $P_h^\pi[h^{-1}u - Q_\pi u - D_t u](t+s-h,y)$, we obtain

$$|A_{-h}| \le E + G + \int_{\mathfrak{X}} H(y) p_\pi(t,x;t+s-h,dy),$$

where

$$E = \sup_{y\in\mathfrak{X}} \left| \int_{\mathfrak{X}} [h^{-1}u - Q_\pi u - D_t u](t+s,z) \mathcal{L}_{t+s}^h(dz) \right|,$$

$$G = h \sup_{y\in\mathfrak{X}} \left| Q_\pi \left[Q_\pi u + D_t u\right](t+s,y) \right|,$$

$$H(y) = |h^{-1}\{u(t+s,y) - u(t+s-h,y)\} - (D_t u)(t+s,y)|.$$

By Assumptions A_1 and A_2(ii), $\lim_{h\downarrow 0} \sup_y \sup_\Gamma |h^{-1}\mathcal{L}_t^h(\Gamma)| = 0$. By (2) and condition (B_2), we have $\lim_{h\downarrow 0} E = 0$. By using condition (B_2) and Assumption A_2(i), and applying Lebesgue's convergence theorem, $\int_{\mathfrak{X}} H(y) p_\pi(t,x;t+s-h,dy)$ converges to 0 as $h \downarrow 0$. Since $G \le 2hc\{2c\|u\| + \sup_y |(D_t u)(t+s,y)|\}$, we have $\lim_{h\downarrow 0} G = 0$. Hence $\lim_{h\downarrow 0} A_{-h} = 0$. Thus, we get $\frac{\partial^-}{\partial s} P_s^\pi u(t,x) = P_s^\pi Q_\pi u(t,x) + P_s^\pi D_t u(t,x)$.

Next, we show the differentiability of F from the right. Let s be in $D(u,t,x) \cap [0, t^* - t)$. We put

$$A_h = h^{-1}\{F(s+h) - F(s)\} - P_s^\pi Q_\pi u(t,x) - P_s^\pi D_t u(t,x), \qquad h > 0.$$

Then we have

$$A_h = \int_{\mathfrak{X}} \left[h^{-1}P_h^\pi u - h^{-1}u - Q_\pi u - D_t u\right](t+s,y) p_\pi(t,x;t+s,dy).$$

Using signed measure $\mathcal{R}_{t+s}^h(\Gamma) = p_\pi(t+s,y;t+s+h,\Gamma) - \delta_y(\Gamma) - hq_\pi(t+s,y;\Gamma)$ on $\mathfrak{X}$ in $P_h^\pi u(t+s,y)$, we obtain

$$|A_h| \le B + \left| \int_{\mathfrak{X}} \{C(y) + D(y)\} p_\pi(t,x;t+s,dy) \right|,$$

where

$$B = h^{-1} \sup_{y\in\mathfrak{X}} \left| \int_{\mathfrak{X}} u(t+s+h,z) \mathcal{R}_{t+s}^h(dz) \right|,$$

$$C(y) = h^{-1}\{u(t+s+h,y) - u(t+s,y)\} - (D_t u)(t+s,y),$$

$$D(y) = \int_{\mathfrak{X}} \{u(t+s+h,z) - u(t+s,z)\} q_\pi(t+s,y;dz).$$

By Assumption A_1, $\lim_{h\downarrow 0} \sup_y \sup_\Gamma |h^{-1}\mathcal{R}_{t+s}^h(\Gamma)| = 0$. Therefore, we have $\lim_{h\downarrow 0} B = 0$. By condition (B_2), $\int_{\mathfrak{X}} C(y) p_\pi(t,x;t+s,dy)$ and $D(y)$ converge to 0 as $h \downarrow 0$. Since $|D(y)| \le 4c\|u\|$, we have $\lim_{h\downarrow 0} \int_{\mathfrak{X}} D(y) p_\pi(t,x;t+s,dy) = 0$. Hence $\lim_{h\downarrow 0} A_h = 0$. Thus, we get $\frac{\partial^+}{\partial s} P_s^\pi u(t,x) = P_s^\pi Q_\pi u(t,x) + P_s^\pi D_t u(t,x)$. This completes the proof. □

Theorem 2.1 *Suppose that*

$$\alpha V \leq (resp., \geq)\ r_\pi + Q_\pi V + D_t V + u \text{ on } D(V) \quad \text{and} \quad V(t^*, \cdot) = 0 \text{ on } \mathfrak{X}$$

for $V \in B_P(T\times\mathfrak{X})$ *and* $u \in B(T\times\mathfrak{X})$. *Then,*

$$V(t,x) \leq (resp., \geq)\ V_\pi(t,x) + \int_0^{t^*-t} e^{-\alpha s} P_s^\pi u(t,x)ds, \qquad (t,x) \in T\times\mathfrak{X}. \tag{4}$$

Proof. Obviously, condition (4) holds for $t = t^*$. Let (t,x) be in $[0,t^*)\times\mathfrak{X}$. We define ψ on $[0, t^*-t]$ by

$$\psi(s) = \begin{cases} \alpha e^{-\alpha s} P_s^\pi V(t,x) - e^{-\alpha s} P_s^\pi Q_\pi V(t,x) - e^{-\alpha s} P_s^\pi D_t V(t,x) \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } s \in D(V,t,x), \\ e^{-\alpha s} P_s^\pi r_\pi(t,x) + e^{-\alpha s} P_s^\pi u(t,x) \qquad\qquad \text{otherwise.} \end{cases}$$

By the hypothesis and monotoneity of P_s^π,

$$\psi(s) \leq (\text{resp.}, \geq)\ e^{-\alpha s} P_s^\pi r_\pi(t,x) + e^{-\alpha s} P_s^\pi u(t,x)$$

for all $s \in [0, t^*-t]$. Integrating the inequality from 0 to t^*-t and using (3), we have

$$\int_0^{t^*-t} \psi(s)ds \leq (\text{resp.}, \geq)\ V_\pi(t,x) + \int_0^{t^*-t} e^{-\alpha s} P_s^\pi u(t,x)ds. \tag{5}$$

We put $\Psi(s) = -e^{-\alpha s} P_s^\pi V(t,x)$. By Lemma 2.2, $\Psi' = \psi$ on $D(V,t,x)$. By condition (B_1), Ψ is absolutely continuous on $[0, t^*-t]$. By condition (B_2), the left hand side of inequality (5) is equal to $\Psi(t^*-t) - \Psi(0)$. Since $V(t^*,\cdot) = 0$ on $\mathfrak{X}$, $\Psi(t^*-t)$ vanishes and the left hand side of (5) is equal to $V(t,x)$. Consequently, we get inequality (4). □

We can get a sufficient condition for a policy to be optimal or ε-optimal by Theorem 2.1.

Theorem 2.2 *Suppose that* $V, V_{\pi^*}, V_{\pi^{**}} \in B_P(T\times\mathfrak{X})$ *and* $1(t,x) = 1$.

(i) *If* V *is a solution of the following equation, then* $V = V_\pi$*:*

$$\alpha V = r_\pi + Q_\pi V + D_t V \text{ on } D(V) \text{ and } V(t^*,\cdot) = 0 \text{ on } \mathfrak{X}. \tag{6}$$

(ii) *It is a sufficient condition for* π^* *to be optimal that* V_{π^*} *is a solution of the optimality equation*

$$\alpha V = \inf_{\pi\in\Pi}\{r_\pi + Q_\pi V\} + D_t V \text{ on } D(V) \text{ and } V(t^*,\cdot) = 0 \text{ on } \mathfrak{X}. \tag{7}$$

(iii) *Given any* $\varepsilon > 0$, *it is a sufficient condition for* π^{**} *to be* ε*-optimal that* $V_{\pi^{**}}$ *is a solution of*

$$\alpha V \leq \inf_{\pi\in\Pi}\{r_\pi + Q_\pi V\} + D_t V + \alpha\varepsilon 1 \text{ on } D(V) \text{ and } V(t^*,\cdot) = 0 \text{ on } \mathfrak{X}. \tag{8}$$

3. Existence of solutions of the optimality equation

We focus on establishing the existence of solutions of optimality equation (7). For $g \in B(T\times\mathfrak{X})$, we put

$$J_g(t,x) = \inf_{\pi\in\Pi}\{r_\pi(t,x) + Q_\pi g(t,x) - \alpha g(t,x)\}.$$

We impose the following assumption for measurability.

Assumption A_3 J_g and $I(t,x) := \int_t^{t^*} J_g(s,x)ds$ are measurable on $T\times\mathfrak{X}$ for any $g \in B(T\times\mathfrak{X})$.

For $0 < \lambda < 1$, the discount rate α and c in Assumption A_1, we can construct a strictly increasing finite sequence $\{t_n\}_{0\le n\le N}$ of times such that $t_0 = 0$, $t_N = t^*$ and

$$(t_n - t_{n-1})(\alpha + 2c) \le \lambda$$

for $n = 1, 2, \cdots, N$. For each n, let T_n be interval $[t_{n-1}, t_n]$ and let $B(T_n\times\mathfrak{X})$ be the Banach space of all bounded measurable functions on $T_n\times\mathfrak{X}$ with the supremum norm $\|\cdot\|_n$.

Lemma 3.1 *There exists a unique collection $g_1^* \in B(T_1\times\mathfrak{X}), g_2^* \in B(T_2\times\mathfrak{X}), \ldots, g_N^* \in B(T_N\times\mathfrak{X})$ such that*

$$g_N^*(t,x) = \int_t^{t_N} J_{g_N^*}(s,x)ds$$

for all $(t,x) \in T_N\times\mathfrak{X}$ and

$$g_n^*(t,x) = \int_t^{t_n} J_{g_n^*}(s,x)ds + g_{n+1}^*(t_n,x)$$

for all $(t,x) \in T_n\times\mathfrak{X}$ and $n = 1, 2, \cdots, N-1$.

Proof. We define an operator S_N on $B(T_N\times\mathfrak{X})$ by $S_N g(t,x) = \int_t^{t_N} J_g(s,x)ds$ for $(t,x) \in T_N\times\mathfrak{X}$. By inequality (2), the boundedness of r_π and Assumption A_3, $S_N g \in B(T_N\times\mathfrak{X})$ for every $g \in B(T_N\times\mathfrak{X})$. Taking $f, g \in B(T_N\times\mathfrak{X})$, we have

$$\begin{aligned}|S_N f(t,x) - S_N g(t,x)| &\le \int_t^{t_N} |J_f(s,x) - J_g(s,x)|ds \\ &\le \int_{t_{N-1}}^{t_N} \sup_{\pi\in\Pi} \left|Q_\pi(f-g)(s,x) - \alpha\{f(s,x) - g(s,x)\}\right| ds.\end{aligned}$$

By using (2),

$$\|S_N f - S_N g\|_N \le (t_N - t_{N-1})(\alpha + 2c)\|f - g\|_N \le \lambda\|f - g\|_N.$$

Hence S_N is a contraction mapping. According to the Banach fixed point theorem, S_N has a unique fixed point g_N^* in $B(T_N\times\mathfrak{X})$, that is, $g_N^* = S_N g_N^*$. Thus, we have the first equation in the lemma. Next, we define an operator S_{N-1} from $B(T_{N-1}\times\mathfrak{X})$ into itself by

$$S_{N-1}g(t,x) = \int_t^{t_{N-1}} J_g(s,x)ds + g_N^*(t_{N-1},x)$$

for $(t,x) \in T_{N-1}\times\mathfrak{X}$. Similarly, S_{N-1} has a unique fixed point g_{N-1}^* in $B(T_{N-1}\times\mathfrak{X})$. By the exactly same argument, for $n = N-2, \cdots, 2, 1$, we can define operators S_n by

$$S_n g(t,x) = \int_t^{t_n} J_g(s,x)ds + g_{n+1}^*(t_n,x)$$

for $(t,x) \in T_n \times \mathfrak{X}$, where g^*_{n+1} is the fixed point of S_{n+1}. Thus, we get the second equation in the lemma. □

For g^*_n, $1 \le n \le N$, in Lemma 3.1, we define a function V^* on $T \times \mathfrak{X}$ by

$$V^* = g^*_1 \text{ on } T_1 \times \mathfrak{X} \qquad \text{and} \qquad V^* = g^*_n \text{ on } (t_{n-1}, t_n] \times \mathfrak{X}. \tag{9}$$

We need the following assumption to verify V^* belongs to $B_P(T \times \mathfrak{X})$.

Assumption A_4 For each $(t,x) \in [0,t^*) \times \mathfrak{X}$ and $\pi \in \Pi$, there exist constants $\delta_{t,x} > 0$ and $L_{t,x} \ge 0$ such that $\sup_{t \le s < t^*} \sup_{\Gamma \in \mathcal{B}_{\mathfrak{X}}} h^{-1} |p_\pi(t,x;s+h,\Gamma) - p_\pi(t,x;s,\Gamma)| < L_{t,x}$ for any $0 < h < \delta_{t,x}$.

Lemma 3.2 *For $u \in B(T \times \mathfrak{X})$, if there exists an absolutely continuous function η on T such that*

$$\sup_{x \in \mathfrak{X}} |u(b,x) - u(a,x)| \le |\eta(b) - \eta(a)|$$

for all $a, b \in T$, then $U(s) := P^\pi_s u(t,x)$ is absolutely continuous on $[0, t^* - t]$ for each $(t,x) \in [0,t^*) \times \mathfrak{X}$.

Proof. We suppose that $[a_i, b_i)$, $i = 1, \cdots, k$, is an arbitrary finite collection of disjoint subintervals of $[0, t^* - t]$. Putting $\nu^h_s(\Gamma) = p_\pi(t,x;s+h,\Gamma) - p_\pi(t,x;s,\Gamma)$ for $t \le s < t^*$ and $h > 0$, we have

$$\sum_{i=1}^k |U(b_i) - U(a_i)| \le 2\|u\| \sum_{i=1}^k \sup_{\Gamma \in \mathcal{B}_{\mathfrak{X}}} |\nu^{b_i - a_i}_{t+a_i}(\Gamma)| + \sum_{i=1}^k |\eta(t+b_i) - \eta(t+a_i)|.$$

By Assumption A_4, there exist $\delta_1 > 0$ and $L \ge 0$ such that $\sup_s \sup_\Gamma |\nu^h_s(\Gamma)| < Lh$ whenever $0 < h < \delta_1$. By the absolute continuity of η, given $\varepsilon > 0$, take a $\delta_2 > 0$ such that, for any finite collection $[c_i, d_i)$, $i = 1, \cdots, k$, of disjoint subintervals of $[t, t^*]$ which satisfies $\sum_{i=1}^k (d_i - c_i) < \delta_2$,

$$\sum_{i=1}^k |\eta(d_i) - \eta(c_i)| < 2^{-1}\varepsilon.$$

By putting $\delta_0 = \min\{\varepsilon(4L\|u\|)^{-1}, \delta_1, \delta_2\}$, for each finite collection $[a_i, b_i)$, $i = 1, \cdots, k$, of disjoint subintervals of $[0, t^* - t]$ with $\sum_{i=1}^k (b_i - a_i) < \delta_0$, we obtain

$$\sum_{i=1}^k |U(b_i) - U(a_i)| \le 2L\|u\|\delta_0 + 2^{-1}\varepsilon \le \varepsilon.$$

Thus, U is absolutely continuous on $[0, t^* - t]$. □

Theorem 3.1 *There exists a solution in $B_P(T \times \mathfrak{X})$ which satisfies the optimality equation.*

Proof. Using Lemma 3.1, we have $V^* = g^*_n$ on $T_n \times \mathfrak{X}$ for function V^* defined by (9) and

$$g^*_n(t,x) = \int_t^{t_n} J_{g^*_n}(s,x)ds + \sum_{k=n+1}^N \int_{t_{k-1}}^{t_k} J_{g^*_k}(s,x)ds, \qquad (t,x) \in T_n \times \mathfrak{X}$$

for each n. Hence we obtain

$$V^*(t,x) = \int_t^{t^*} \inf_{\pi\in\Pi}\Big\{r_\pi(s,x) + Q_\pi V^*(s,x) - \alpha V^*(s,x)\Big\}ds, \qquad (t,x)\in T\times\mathfrak{X}. \quad (10)$$

According to Lebesgue's differentiation theorem (Dudley, 1989, Theorem 7.2.1), differential coefficient $D_tV^*(t,x)$ exists for all $x\in\mathfrak{X}$ and almost all $t\in T$, and we obtain $D_tV^* = -\inf_\pi\{r_\pi + Q_\pi V^* - \alpha V^*\}$ on $D(V^*)$ and $V^*(t^*,\cdot) = 0$ on $\mathfrak{X}$. Thus, V^* satisfies optimality equation (7). We show $V^*\in B_P(T\times\mathfrak{X})$. We have $V^*\in B(T\times\mathfrak{X})$ by its definition and $g_n^*\in B(T_n\times\mathfrak{X})$. Using (10) and (2), we get

$$\sup_{x\in\mathfrak{X}}|V^*(b,x) - V^*(a,x)| \le \{M + (2c+\alpha)\|V^*\|\}|b-a|, \qquad a,b\in T.$$

Hence $F(s) := P_sV^*(t,x)$ is absolutely continuous on $[0,t^*-t]$ by Lemma 3.2. By (2) and (7), $D_tV^*(t,\cdot)$ is bounded on $D(V^*,t)$. Moreover, $h^{-1}\{V^*(t+h,x) - V^*(t,x)\}$ is bounded. Finally, by (7) and Assumption A_3, D_tV^* is measurable on $D(V^*)$. Thus V^* belongs to $B_P(T\times\mathfrak{X})$. □

Theorem 3.2 *V_π is the unique solution in $B_P(T\times\mathfrak{X})$ of (6), that is,*

$$\alpha V_\pi = r_\pi + Q_\pi V_\pi + D_tV_\pi.$$

Proof. By Theorem 2.2, any solution in $B_P(T\times\mathfrak{X})$ of (6) must be V_π. Supposing Π is singleton $\{\pi\}$ in Theorem 3.1, we can show the existence of a solution in $B_P(T\times\mathfrak{X})$ of equation (6). □

Corollary 3.1 *If $\inf_{\pi\in\Pi}\{r_\pi + Q_\pi V_{\pi^*}\} = r_{\pi^*} + Q_{\pi^*}V_{\pi^*}$ on $D(V_{\pi^*})$, then π^* is an optimal policy. If $\inf_{\pi\in\Pi}\{r_\pi + Q_\pi V_{\pi_\varepsilon}\} \ge r_{\pi_\varepsilon} + Q_{\pi_\varepsilon}V_{\pi_\varepsilon} - \alpha\varepsilon 1$ on $D(V_{\pi_\varepsilon})$, then π_ε is an ε-optimal policy.*

Proof. By Theorems 3.2 and 2.2, the assertion of the corollary is straightforward. □

Theorem 3.3 *Assume that, for any $\varepsilon > 0$ and a solution $V\in B_P(T\times\mathfrak{X})$ of optimality equation (7), there exists $\pi_\varepsilon\in\Pi$ such that*

$$\inf_{\pi\in\Pi}\{r_\pi + Q_\pi V\} \ge r_{\pi_\varepsilon} + Q_{\pi_\varepsilon}V - \alpha\varepsilon 1 \qquad \text{on } D(V).$$

Then V_{opt} is the unique solution in $B_P(T\times\mathfrak{X})$ of the optimality equation. Moreover, there exists an ε-optimal policy for any $\varepsilon > 0$.

Proof. Let V be a solution in $B_P(T\times\mathfrak{X})$ of the optimality equation. Then we obtain $\alpha V \le r_\pi + Q_\pi V + D_tV$ on $D(V)$ for all $\pi\in\Pi$ and $V(t^*,\cdot) = 0$ on $\mathfrak{X}$. Applying Theorem 2.1, we have $V \le V_{\mathrm{opt}}$. By the hypothesis, for any $\varepsilon > 0$ there exists $\pi_\varepsilon\in\Pi$ such that $\alpha V \ge r_{\pi_\varepsilon} + Q_{\pi_\varepsilon}V + D_tV - \alpha\varepsilon$ on $D(V)$. Applying Theorem 2.1, we get

$$V \ge V_{\pi_\varepsilon} - \varepsilon 1 \ge V_{\mathrm{opt}} - \varepsilon 1.$$

Letting $\varepsilon\downarrow 0$, we get $V \ge V_{\mathrm{opt}}$. Thus, we obtain $V = V_{\mathrm{opt}}$ and $V_{\pi_\varepsilon} \le V_{\mathrm{opt}} + \varepsilon 1$ for every $\varepsilon > 0$. □

Theorem 3.4 *Assume $V_{\rm opt}$ is a solution in $B_P(T\times\mathfrak{X})$ of (7). Then the following assertions are equivalent to each other.*

(i) *π^* is an optimal policy;*

(ii) $\inf_{\pi\in\Pi}\{r_\pi + Q_\pi V_{\rm opt}\} = r_{\pi^*} + Q_{\pi^*}V_{\rm opt}$ *on* $D(V_{\rm opt})$;

(iii) $\inf_{\pi\in\Pi}\{r_\pi + Q_\pi V_{\pi^*}\} = r_{\pi^*} + Q_{\pi^*}V_{\pi^*}$ *on* $D(V_{\pi^*})$.

Proof. By the hypothesis, assertion (ii) implies $\alpha V_{\rm opt} = r_{\pi^*} + Q_{\pi^*}V_{\rm opt} + D_t V_{\rm opt}$ on $D(V_{\rm opt})$. Applying Theorem 2.1, we have assertion (i). By Corollary 3.1, assertion (iii) implies assertion (i). Next, if π^* is optimal, then $V_{\pi^*} = V_{\rm opt}$. By Theorem 3.2 and the hypothesis, V_{π^*} is a solution of both equations (6) and (7). Hence, we have assertion (iii). Consequently, assertion (i) implies assertions (ii) and (iii). □

The preceding theorem shows that a key to the existence of optimal policies is the existence of policies which attain the infimum in assertions (ii) or (iii).

Acknowledgment The author is grateful to the referees for their careful reading of the original manuscript and helpful suggestions.

References

Billingsley, P. (1979), *Probability and Measure* (John Wiley & Sons).

Doshi, B.T. (1976), Continuous time control of Markov processes on an arbitrary state space: discounted rewards, *Ann. Statist.* **4**, 1219-1235.

Dudley, R.M. (1989), *Real Analysis and Probability* (Wadsworth & Brooks).

Kakumanu, P.K. (1971), Continuously discounted Markov decision model with countable state and action spaces, *Ann. Math. Statist.* **42**, No. 3, 919-926.

Lai, H.C. and Tanaka, K. (1991), On continuous-time discounted stochastic dynamic programming, *Appl. Math. Optim.* **23**, 155-169.

Miller, B.L. (1968), Finite state continuous time Markov decision processes with an infinite planning horizon, *J. Math. Anal. Appl.* **22**, 552-569.

Qiying, H. (1993), Nonstationary continuous time Markov decision processes with discounted criterion, *J. Math. Anal. Appl.* **180**, 60-70.

Received September 2, 1997 **Revised January 5, 1998**