ON THE EXISTENCE OF OPTIMAL CONTROL IN CONTINUOUS TIME MARKOV DECISION PROCESSES

Yasuda, Masami
Department of Mathematics, Kagoshima University

https://doi.org/10.5109/13058

出版情報：統計数理研究. 15 (1/2), pp.7-17, 1972-03. Research Association of Statistical Sciences
バージョン：
権利関係：

ON THE EXISTENCE OF OPTIMAL CONTROL IN CONTINUOUS TIME MARKOV DECISION PROCESSES

By

Masami YASUDA*

(Received November 1, 1971)

§ 1. Introduction.

In this paper we shall treat the optimal control problem in continuous time Markov decision processes having a Borel state space and a compact action space varying with both the time and the state. The cost functional we consider here is the sum of the integral over the finite horizon of a return rate which depends on both the controller and the corresponding response, and the expected return of the system at the final fixed time. Our optimal control problem is to find a controller which maximzie the cost functional over the given planning horizon.

Main results are a necessary and sufficient condition for an optimality, and an algorithm for finding the optimal controller. B. L. Miller [1] treated the problem similar to ours, but his paper was restricted to the case of the finite state space and action space. Our situation is succeeded owing to the implicit function's lemma of K. Tsuji and N. Furukawa [3]. The method of construction of our algorithm is often used in Dynamic Programming problem, for example in [4].

§ 2. Notations, definitions and problem formulations.

The state space S is an abstruct set. $\mathfrak{B}(S)$ means a σ-algebra of subsets of S. In this paper we shall confine ourselves to the case when for each $x \in S$, the single-point set $\{x\} \in \mathfrak{B}(S)$. Hence the assumption that $(S, \mathfrak{B}(S))$ is a separable Borel space is sufficient to the case. Let μ denote a finite measure on $\mathfrak{B}(S)$. Hereafter we assume $\mathfrak{B}(S)$ is completed with the measure μ.

The action space A is a non-empty compact set in R^n. Let 2^A denote the familty of all non-empty closed subsets contained in A. Hausdorff distance h between A_1 and A_2 in 2^A is a function

$$h(A_1, A_2) = \max_{a_1 \in A_1, a_2 \in A_2} (d(a_1, A_2), d(a_2, A_1)),$$

where d is the Euclidean distance of a point from a set. The function h is known

* Department of Mathematics, Kagoshima University, Kagoshima.

to be a metric in 2^A and the space $(2^A, h)$ is complete. Let J be a fixed finite closed interval $[t_0, T]$ in R^1, λ be a Lebesgue measure on J, and $\mathfrak{B}(J)$ be a family of Lebesgue measurable sets in J. A set-valued mapping $A(\cdot, \cdot): JS \to 2^A$ is $\lambda \otimes \mu$-measurable if there is a sequence $\{A_n(\cdot, \cdot)\}$ of a $\lambda \otimes \mu$-simple set-valued mapping for which $A_n(\cdot, \cdot)$ converges in $\lambda \otimes \mu$-measure to $A(\cdot, \cdot)$, where $\lambda \otimes \mu$-simple means that it assumes only a finite number of values in 2^A and each of them on a $\lambda \otimes \mu$-measurable set.

A mapping π from JS to A is called an *admissible controller*, if π is measurable in the pair (t, x), and if for each t, x, $\pi(t, x) \in A(t, x)$.

In order to determine the transition probability of the system, we shall introduce p and Π defined on JSA and $JS\mathfrak{B}(S)A$ respectively. Let p and Π satisfy the following conditions:

(C1) p is measurable in the pair (t, x) for each fixed $\xi \in A$.

(C2) p is continuous in ξ for each fixed $(t, x) \in JS$.

(C3) There exists $p_0(t)$ such that $0 < p(t, x, \xi) \leqq p_0(t)$ for all $(x, \xi) \in SA$, and $\int_J |p_0(t)|^a dt < \infty$ for some $a > 1$.

(C4) Π is measurable in the pair (t, x) for each fixed $(\Lambda, \xi) \in \mathfrak{B}(S)A$.

(C5) Π is a completely additive set function on $\mathfrak{B}(S)$ such that

$$0 \leqq \Pi(t, x, \Lambda, \xi) \leqq 1, \qquad \Pi(t, x, \{x\}, \xi) = 0$$

$$\Pi(t, x, S, \xi) = 1, \qquad \text{for all } (t, x, \Lambda, \xi) \in JS\mathfrak{B}(S)A.$$

(C6) Π is weakly continuous in ξ in the sense that for every bounded measurable function g on JS, $\int_S g(t, x)\Pi(t, x, dz, \xi)$ is continuous in ξ for each fixed $(t, x) \in JS$.

The *transition probability* P^π of the system corresponding to every admissible controller π is uniquely determined by the following proposition, the proof of which is the same as the one in [2].

PROPOSITION 2.1. *If π is an admissible controller, and if p and Π satisfy the assumptions* (C1)-(C6), *then there exists a unique solution P^π of the equation:*

$$\text{(1)} \qquad \frac{\partial P^\pi(\tau, x, t, \Lambda)}{\partial t} = -\int_\Lambda p(t, y, \pi(t, y)) P^\pi(\tau, x, t, dy)$$

$$+ \int_S p(t, y, \pi(t, y))\Pi(t, y, \Lambda, \pi(t, y)) P^\pi(\tau, x, t, dy) \quad \textit{for almost all } t$$

and

$$\text{(2)} \qquad P^\pi(\tau, x, \tau, \Lambda) = \delta(x, \Lambda)$$

where

$$\delta(x, \Lambda) = \begin{cases} 1 & \text{if } x \in \Lambda \\ 0 & \text{if } x \notin \Lambda, \end{cases}$$

satisfying that P^π is a probability measure on $\mathfrak{B}(S)$, for each fixed τ, x, Λ, absolutely continuous in t, and measurable in the pair (τ, x), that

$$P^{\pi}(\tau, x, t, \{x\}) \geqq \exp\left\{-\int_{\tau}^{t} p(s, x, \pi(s, x))ds\right\} > 0 \quad \textit{for all } x, \tau, t, \tag{3}$$

and that

$$\frac{\partial P^{\pi}(\tau, x, t, \Lambda)}{\partial \tau} = p(\tau, x, \pi(\tau, x))\Big\{P^{\pi}(\tau, x, t, \Lambda) - \int_{S} P^{\pi}(\tau, y, t, \Lambda)\Pi(\tau, x, dy, \pi(\tau, (x)))\Big\} \quad \textit{for almost all } \tau, \tag{4}$$

and

$$P^{\pi}(t, x, t, \Lambda) = \delta(x, \Lambda).$$

Let $r: SA \to R^1$ be a given function such that for each fixed $a \in A$, $r(x, a)$ is bounded measurable in $x \in S$, and for each fixed $x \in S$, $r(x, a)$ is continuous in $a \in A$. Let $\varphi: S \to R^1$ be a given function which is bounded measurable on S. And let $\alpha \geqq 0$ be a given real number, we includ also the discounted case of finite horizon. Then the *reward function* I^{π} corresponding to an admissible controller π is given by

$$I^{\pi}(t, x) = \int_{t}^{T} e^{-\alpha(s-t)}ds \int_{S} r(y, \pi(s, y))P^{\pi}(t, x, s, dy) + e^{-\alpha(T-t)}\int_{S} \varphi(y)P^{\pi}(t, x, T, dy). \tag{5}$$

Thus $I^{\pi}(t, x)$ means the *total expected reward* over the time interval $[t, T]$, starting from x at the time t and following the admissible controller π. A controller π^* is called *optimal*, if π^* is an admissible controller and if for every admissible controller π, $I^{\pi^*}(t, x) \geqq I^{\pi}(t, x)$ for all (t, x) in JS.

§ 3. Preliminary results.

DEFINITION. Let F, U and V be set functions, and let φ_1 and φ_2 be bounded measurable functions. Then we write

$$F(dx) = \varphi_1(x)U(dx) + \varphi_2(x)V(dx),$$

iff

$$F(A) = \int_{A} \varphi_1(x)U(dx) + \int_{A} \varphi_2(x)V(dx)$$

for any measurable set A.

Following Lemma 3.1 is immediate from the above definition and Proposition 2.1.

LEMMA 3.1. (i) *Suppose that*

$$F(dx) = \varphi_1(x)U(dx) + \varphi_2(x)V(dx).$$

Then for any bounded measurable function v it holds that

$$v(x)F(dx) = v(x)\varphi_1(x)U(dx) + v(x)\varphi_2(x)V(dx).$$

(ii) *P^{π} satisfies that*

$$\frac{\partial P^{\pi}(\tau, x, t, dy)}{\partial t} = -p(t, y, \pi(t, y))P^{\pi}(\tau, x, t, dy) + \int_{S} p(t, z, \pi(t, z))\Pi(t, z, dy, \pi(t, z))P^{\pi}(\tau, x, t, dz) \quad \textit{for almost all } t, \tag{6}$$

and

$$\frac{\partial P^{\pi}(\tau, x, t, dy)}{\partial \tau} = p(\tau, x, \pi(\tau, x))\left\{P^{\pi}(\tau, x, t, dy) - \int_S P^{\pi}(\tau, z, t, dy)\Pi(\tau, x, dz, \pi)(\tau, x))\right\} \quad \text{for almost all } \tau. \tag{7}$$

LEMMA 3.2. *Let $(E_i, \mathfrak{B}_i)$, $i=1, 2$, be Borel spaces. Let F be a complete additive set function on $\mathfrak{B}_1$ such that $0 \leqq F(\Lambda_1) \leqq 1$ for every $\Lambda_1 \in \mathfrak{B}_1$. Let G defined on $E_1\mathfrak{B}_2$, be such that G is complete additive on $\mathfrak{B}_1$ for each fixed $x_1 \in E_1$, and integrable with respect to F for each fixed $\Lambda_2 \in \mathfrak{B}_2$. Then, for any $\Gamma_1 \in \mathfrak{B}_1$ and any $\Gamma_2 \in \mathfrak{B}_2$, it holds that*

$$\int_{\Gamma_1} f(x_1)F(dx_1)\int_{\Gamma_2} g(x_2)G(x_1, dx_2) = \int_{\Gamma_2} g(x_2)\int_{\Gamma_1} f(x_1)G(x_1, dx_2)F(dx_1). \tag{8}$$

This lemma is due to the results in Feller [2].

LEMMA 3.3. *If $f(t, y)$, $t \in J$, $y \in S$ is bounded measurale in y for each fixed t, and absolutely continuous in t for each fied y, if $P(t, \Lambda)$ is a probability measure on $\mathfrak{B}(S)$ for each t, and absolutely continuous in t for each Λ, and if, for almost all t, its derivative P_t is completely additive on $\mathfrak{B}(S)$, then it holds that*

$$\int_{\tau}^{t}\int_S f(s, y)P_s(s, dy)ds + \int_{\tau}^{t}\int_S f_s(s, y)P(s, dy)ds = \int_S f(t, y)P(t, dy) - \int_S f(\tau, y)P(\tau, dy) \tag{9}$$

where f_s is a derivative of f.

PROOF. Since f is absolutely continuous,

$$\int_{\tau}^{t}\int_S f(s, y)P_s(s, dy)ds = \int_{\tau}^{t} ds\int\left\{\int_{\tau}^{s} f_u(u, y)du + f(\tau, y)\right\}P_s(s, dy)$$

Changing the order of integration by Lemma 3.2 and also using the absolute continuity of P, this leads to

$$\int_S f(t, y)P(t, dy) - \int_{\tau}^{t}\int_S f_s(s, y)P(s, dy)ds - \int_S f(\tau, y)P(\tau, dy).$$

Thus the lemma is proved.

DEFINITION. Let $f(t, y)$ be bounded measurable in $y \in S$ for each fixed $t \in J$, and absolutely continuous in $t \in J$ for each fixed $y \in S$. Then the operator A^{π} associated with an admissible controller π is given by

$$A^{\pi}f(t, y) = \frac{\partial f(t, y)}{\partial t} + p(t, y, \pi(t, y))\int_S \{f(t, z) - f(t, y)\}\Pi(t, y, dz, \pi(t, y)).$$

LEMMA 3.4. *If $f(t, y)$ is bounded measurable in $y \in S$ for each fixed $t \in J$, and for each fixed $y \in S$, absolutely continuous in $t \in J$, then it holds that for all τ, x and t,*

$$\int_S f(t, y)P^{\pi}(\tau, x, t, dy) - f(\tau, x) = \int_{\tau}^{t}\int_S A^{\pi}f(s, y)P^{\pi}(\tau, x, s, dy)ds. \tag{10}$$

PROOF. By using Lemma 3.1 (ii) and Lemma 3.3 we obtain

$$\int_\tau^t\int_S f(s,y)\frac{\partial P^\pi(\tau,x,s,dy)}{\partial s}ds \tag{11}$$

$$=\int_S f(t,y)P^\pi(\tau,x,t,dy)-f(\tau,x)-\int_\tau^t\int_S f_s(s,y)P^\pi(\tau,x,s,dy)ds\,.$$

On the other hand it holds that

$$-\int_\tau^t\int_S p(s,y,\pi(s,y))f(s,y)P^\pi(\tau,x,s,dy)ds \tag{12}$$

$$+\int_\tau^t\int_S\int_S p(s,z,\pi(s,z))f(s,y)\Pi(s,z,dy,\pi(s,z))P^\pi(\tau,x,s,dz)ds$$

$$=\int_\tau^t\int_S\{-p(s,y,\pi(s,y))f(s,y)+p(s,y,\pi(s,y))\int_S f(s,z)$$

$$\times\Pi(s,y,dz,\pi(s,y))\}P^\pi(\tau,x,s,dy)ds\,.$$

Combining (11) and (12), we get (10).

LEMMA 3.5. *The reward function I^π is equal to the unique solution f of the following equation;*

$$\begin{cases} A^\pi f(t,x)-\alpha f(t,x)=-r(x,\pi(t,x)) \\ f(T,x)=\varphi(x) \end{cases} \tag{13}$$

for almost all $t\in J$ and for all $x\in S$.

PROOF. Let

$$F_1=I^\pi(\tau,x)-e^{-\alpha(T-\tau)}\int_S\varphi(z)P^\pi(\tau,x,T,dz)$$

$$=\int_\tau^T e^{-\alpha(t-\tau)}\int_S r(z,\pi(t,z))P^\pi(\tau,x,t,dz)dt\,.$$

Since r is bounded, using Lemma 3.2 and the relation

$$P^\pi(\tau,x,t,dz)=\delta(x,dz)-\int_\tau^t\frac{\partial P^\pi(s,x,t,dz)}{\partial s}ds\,,$$

we get

$$F_1=\int_\tau^T e^{-\alpha(t-\tau)}r(x,\pi(t,x))dt-F_2$$

where

$$F_2=\int_\tau^T e^{-\alpha(t-\tau)}\int_S\int_\tau^t r(z,\pi(t,z))\frac{\partial P^\pi(s,x,t,dz)}{\partial s}dsdt\,.$$

By Lemma 3.1 and Lemma 3.2 we have

$$F_2=F_3-F_4\,,$$

where

$$F_3=\int_\tau^T e^{-\alpha(t-\tau)}\int_S\int_\tau^t r(z,\pi(t,z))p(s,x,\pi(s,x))P^\pi(s,x,t,dz)dsdt$$

$$=\int_\tau^T e^{-\alpha(s-\tau)}p(s,x,\pi(s,x))[I^\pi(s,x)-e^{-\alpha(T-s)}\int_S\varphi(y)P^\pi(s,x,T,dy)]ds$$

and

$$F_4=\int_\tau^T e^{-\alpha(t-\tau)}\int_S r(z,\pi(t,z))\int_\tau^t p(s,x,\pi(s,x))\int_S P^\pi(s,y,t,dz)$$
$$\times\Pi(s,x,dy,\pi(s,x))dsdt$$
$$=\int_\tau^T e^{-\alpha(s-\tau)}p(s,x,\pi(s,x))\int_S[I^\pi(s,y)-e^{-\alpha(T-s)}$$
$$\times\int_S\varphi(z)P^\pi(s,y,T,dz)]\Pi(s,x,dy,\pi(s,x))ds\,.$$

Again Lemma 3.1 gives

$$\int_S\varphi(y)P^\pi(\tau,x,T,dy)-\varphi(x)$$
$$=-\int_\tau^T p(s,x,\pi(s,x))\int_S\varphi(y)P^\pi(s,x,t,dy)ds$$
$$+\int_\tau^T p(s,x,\pi(s,x))\int_S\Big\{\int_S\varphi(y)P^\pi(s,z,t,dy)\Big\}\Pi(s,x,dz,\pi(s,x))ds\,.$$

Hence, from these relations we have

$$I^\pi(\tau,x)-e^{-\alpha(T-\tau)}\varphi(x)$$
$$=\int_\tau^T e^{-\alpha(t-\tau)}r(x,\pi(t,x))dt$$
$$-\int_\tau^T e^{-\alpha(t-\tau)}p(t,x,\pi(t,x))I^\pi(t,x)dt$$
$$+\int_\tau^T e^{-\alpha(t-\tau)}p(t,x,\pi(t,x))\int_S I^\pi(t,z)\Pi(t,x,dz,\pi(t,x))dt\,.$$

Obviously

$$I^\pi(\tau,x)-\varphi(x)=-\alpha\int_\tau^T I^\pi(t,x)dt+\int_\tau^T r(x,\pi(t,x))dt$$
$$-\int_\tau^T p(t,x,\pi(t,x))I^\pi(t,x)dt$$
$$+\int_\tau^T p(t,x,\pi(t,x))\int_S I^\pi(t,z)\Pi(t,x,dz,\pi(t,x))dt\,,$$

which follows that

$$A^\pi I^\pi(t,x)-\alpha I^\pi(t,x)=-r(x,\pi(t,x))$$
$$I^\pi(T,x)=\varphi(x)\qquad\text{for almost all } t \text{ and for all } x\,.$$

To show the uniqueness of the solution of (13). Let f_1 and f_2 be two solutions of (13). Then put $f=f_1-f_2$. Thus it follows that f satisfies the equation;

$$\text{(14)}\qquad\begin{cases} A^\pi f(t,x)-\alpha f(t,x)=0 \\ f(T,x)=0 \qquad \text{for almost all } t \text{ and for all } x\,. \end{cases}$$

We shall now show that $f(t,x)=0$ for all t, x. It follows directly from (10) that

$$\int_S f(T, y)P^\pi(t, x, T, dy)-f(t, x)$$

$$=\int_t^T\int_S A^\pi f(s, y)P^\pi(t, x, s, dy)ds\,,$$

which together with (14) yields

$$-f(t, x)=\alpha\int_t^T\int_S f(s, y)P^\pi(t, x, s, dy)ds\,.$$

Let $g(s, x)=\int_S f(s, y)P^\pi(t, x, s, dy)$ with arbitrary fixed t. As $P^\pi(t, x, t, dy)=\delta(x, dy)$, $g(t, x)=f(t, x)$. Hence $g(t, x)=0$ follows from $g(t, x)+\alpha\int_t^T g(s, x)dt=0$.

LEMMA 3.6. *Let $(X, \mathfrak{B}(X), \sigma)$ be a probability space, and let A be a compact subset in R^n. Suppose that $f(x, a)$; $XA\to R^1$ is measurable in x for each fixed a, and continuous in a for each fixed x, and that $A(x)$; $X\to 2^A$ is a σ-measurable set-valued mapping. Then $f(x, A(x))=\{y\,;\, y=f(x, a),\ a\in A(x)\}$ is a σ-measurable set-valued mapping from S to 2^A, $\max_{a\in A(x)} f(x, a)$ is measurable, and there exists a measurable function $u(x)$; $X\to R^n$ such that*

$$f(x, u(x))=\max_{a\in A(x)} f(x, a) \qquad \textit{for all } x\in X\,.$$

PROOF. These results can be obtained by a slight modification of those in Tsuji-Furukawa [3].

§ 4. A necessary and sufficient condition for the optimality.

With any admissible controller π, we associate the operators L^π and L^a, given by

$$L^\pi f(t, x)=r(x, \pi(t.\ x))-p(t, x, \pi(t, x))f(t, x)$$

$$+p(t, x, \pi(t, x))\int_S f(t, z)\Pi(t, x, dz, \pi(t, x))-\alpha f(t, x)\,,$$

$$L^a f(t, x)=r(x, a)-p(t, x, a)f(t, x)$$

$$+p(t, x, a)\int_S f(t, z)\Pi(t, x, dz, a)-\alpha f(t, x)\,,$$

respectively, provided that $a\in A(t, x)$ and f is bounded measurable.

COROLLARY 4.1. *The reward function I^π is the unique solution of the equation:*

$$-\frac{\partial I^\pi(t, x)}{\partial t}=L^\pi I^\pi(t, x)$$

$$I^\pi(T, x)=\varphi(x) \qquad \textit{for almost all } t \textit{ and for all } x\,.$$

PROOF. This is a re-statement of Lemma 3.5.

LEMMA 4.1. *For any admissible controllers π, π' it holds that*

$$I^{\pi}(t,x)-I^{\pi'}(t,x)+\alpha\int_t^T\int_S I^{\pi}(s,y)P^{\pi'}(t,x,s,dy)ds \tag{15}$$

$$-\alpha\int_t^T\int_S I^{\pi'}(s,y)P^{\pi'}(t,x,s,dy)ds$$

$$=\int_t^T\int_S[L^{\pi}I^{\pi}(s,y)-L^{\pi'}I^{\pi}(s,y)]P^{\pi'}(t,x,s,dy)ds\,.$$

PROOF. From Lemma 3.1 (ii), Lemma 3.3 and Lemma 3.5 we have

$$\int_S\varphi(y)P^{\pi}(t,x,T,dy)-I^{\pi}(t,x) \tag{16}$$

$$=\int_t^T\int_S[-r(y,\pi(s,y))+\alpha I^{\pi}(s,y)]P^{\pi}(t,x,s,dy)ds$$

and

$$\int_S\varphi(y)P^{\pi'}(t,x,T,dy)-I^{\pi}(t,x) \tag{17}$$

$$=\int_t^T\int_S[L^{\pi'}I^{\pi}(s,y)-L^{\pi}I^{\pi}(s,y)]P^{\pi'}(t,x,s,dy)ds$$

$$+\int_t^T\int_S[-r(y,\pi'(s,y))+\alpha I^{\pi}(s,y)]P^{\pi'}(t,x,s,dy)ds\,.$$

On the other hand it holds that

$$\int_S\varphi(y)P^{\pi'}(t,x,T,dy) \tag{18}$$

$$=I^{\pi'}(t,x)-\int_t^T\int_S[r(y,\pi'(s,y))-\alpha I^{\pi'}(s,y)]P^{\pi'}(t,x,s,dy)ds\,.$$

Finally substituting (18) into (17), (16) and (17) yield the relation (15).

DEFINITION. We say that a bounded measurable function $f(t,x)$, $t\in J$, $x\in S$ satisfies the *optimal equation*, if it holds that

$$\frac{\partial f(t,x)}{\partial t}\max_{a\in A(t,x)}L^a f(t,x)=0 \qquad \text{for almost all } t \text{ and for all } x$$

with

$$f(T,x)=\varphi(x) \qquad \text{for all } x\,.$$

THEOREM 4.1. *An admissible controller π^* is optimal if and only if its reward function I^{π^*} satisfies the optimal equation.*

PROOF. The "if" part is an immediate consequence of Lemma 4.1 by virtue of the equaties in Corollary 4.1.

We shall next prove the "only if" part. Suppose that the reward function I^{π^*} corresponding to the optimal controller π^* fails to satisfy the optimal equation, that is,

$$-\frac{\partial I^{\pi^*}(t,x_0)}{\partial t}\neq\max_{a\in A(t,x_0)}L^a I^{\pi^*}(t,x_0)$$

on a subset $J^*\subset J$ which measure is positive and on a point $x_0\in S$. Let π be an adimissible controller satisfying

$$L^{\pi}I^{\pi^*}(t, x) = \max_{a \in A(t,x)} L^a I^{\pi^*}(t, x) \qquad \text{for almost all } t \text{ and for all } x \in S.$$

The existence of such an admissible controller is guaranteed by Lemma 3.6. Clearly we have

$$L^{\pi}I^{\pi^*}(t, x_0) > L^{\pi^*}I^{\pi^*}(t, x_0) \qquad \text{for } t \in J^* \text{ and for a point } x_0 \in S.$$

Hence from (3) in Proposition 2.1 and Lemma 4.1, for $t \in J^*$ and for a point $x_0 \in S$,

$$I^{\pi^*}(t, x_0) < I^{\pi}(t, x_0)$$

which contradicts the optimality of π^*.

§ 5. Iteration procedure for finding an optimal controller.

THEOREM 5.1. *If we define the sequence* $\{R_n\}$ *by the following iteration procedure* (i)-(iii), *then* $\{R_n\}$ *is a monotone increasing and* $\lim R_n(t, x)$ *exists for each* $(t, x) \in JS$.

(i) *Let* $\pi^{(1)}(t, x)$ *be any admissible controller.*

(ii) *For each* $n \geqq 1$ *compute* $R_n(t, x) = I^{\pi^{(n)}}(t, x)$, *or* $R_n(t, x)$ *be the unique solution of the equation:*

$$-\frac{\partial R_n(s, x)}{\partial s} = L^{\pi^{(n)}} R_n(s, x), \qquad R_n(T, x) = \varphi(x), \quad \text{for almost all } s \text{ and for all } x. \tag{19}$$

(iii) *For each* $n \geqq 1$ *find an admissible controller* $\pi^{(n+1)}$ *such that*

$$L^{\pi^{(n+1)}} R_n(s, x) = \max_{a \in A(s,x)} L^a R_n(s, x) \qquad \text{for almost all } s \text{ and for all } x. \tag{20}$$

Start from (i), *and then repeat* (ii) *and* (iii) *by turns.*

NOTE. In step (ii), the existence of the unique solution of (19) is due to Corollary 4.1. In the step (iii), the existence of an admissible controller satisfying (20) is assured by Lemma 3.6.

PROOF. Because of the boundedness of the return rate r in the reward function, it is sufficient to verify the monotony of the sequence $\{R_n\}$.

From the definition of $\pi^{(n+1)}$ it follows that

$$-\frac{\partial R_n(s, x)}{\partial s} - L^{\pi^{(n+1)}} R_n(s, x) \leqq -\frac{\partial R_n(s, x)}{\partial t} + L^{\pi^{(n)}} R_n(s, x) = 0$$

for almost all s and for all x. Hence it holds that

$$\frac{\partial R_{n+1}(s, x)}{\partial s} - \frac{\partial R_n(s, x)}{\partial s} \leqq -L^{\pi^{(n+1)}} R_{n+1}(s, x) + L^{\pi^{(n+1)}} R_n(s, x). \tag{21}$$

Putting

$$\psi(t, x) = R_{n+1}(t, x) - R_n(t, x),$$

(21) is re-written in terms of the operator A^{π} as

$$\begin{cases} A^{\pi^{(n+1)}}\psi(s, x) - \alpha\psi(s, x) \leqq 0 \\ \psi(T, x) = 0 \qquad \text{for almost all } s \text{ and for all } x. \end{cases} \tag{22}$$

Forthermore according to Lemma 3.4 and (22), we have

$$\int_S \phi(T, y)P^{\pi^{(n+1)}}(t, x, T, dy)-\phi(t, x)$$

$$\leqq \alpha\int_t^T\int_S \phi(s, x)P^{\pi^{(n+1)}}(t, x, s, dy)ds \qquad \text{for each } t,\ x.$$

A similar argument in the later part of the proof of Lemma 3.5 implies

$$\phi(t, x)\geqq 0 \qquad \text{for each } t,\ x.$$

This is nothing but to say

$$R_{n+1}(t, x)\geqq R_n(t, x) \qquad \text{for each } t,\ x.$$

COROLLARY 5.1. *If, in the above iteration procedure,* $\pi^{(n+1)}(t, x)=\pi^{(n)}(t, x)$ *for almost all* t *and for all* x, *hold for* n, *then* $\pi^{(n)}$ *is an optimal controller.*

PROOF. The corollary is straightforward from the fact that the assumption implies the reward function $I^{\pi^{(n)}}$ corresponding to $\pi^{(n)}$ satisfies the optimal equation.

COROLLARY 5.2. *For any admissible controllers* π *and* π',

$$I^{\pi'}(t, x)\geqq I^{\pi}(t, x) \qquad \textit{for each } t,\ x$$

according as

$$\frac{\partial I^{\pi}(s, x)}{\partial s}+L^{\pi'}I^{\pi}(s, x)\geqq 0 \qquad \textit{for almost all } s \textit{ and for all } x.$$

PROOF. The corollary is obvious from the proof of the above theorem.

THEOREM 5.2. *Let* $\{R_n\}$ *be as in Theorem* 5.1. *Then the limit function* $R^*(t, x)=\lim_{n\to\infty} R_n(t, x)$, $t\in J$, $x\in S$ *satisfies the optimal equation. Furthermore the optimal controller* π^* *exists and is characterized by*

$$L^{\pi^*}R^*(t, x)=\max_{a\in A(t,x)} L^aR^*(t, x) \qquad \textit{for almost all } t \textit{ and for all } x.$$

PROOF. We shall show R^* satisfies the optimal equation. Firstly, putting

$$M=R^*(t, x)-R^*(s, x)+\int_s^t L^{\pi^*}R^*(u, x)du$$

we shall prove $M\leqq 0$. From the definition of $\pi^{(n+1)}$ it follows that

$$\int_s^t L^{\pi^*}R_n(u, x)du\leqq\int_s^t L^{\pi^{(n+1)}}R_n(u, x)du.$$

Therefore we get, by the monotony of $\{R_n\}$,

$$M\leqq R^*(t, x)-R_n(t, x)+\alpha\int_s^t (R_{n+1}(u, x)-R_n(u, x))du$$

$$+\int_s^t p(x, \pi^{(n+1)}(u, x))(R_{n+1}(u, x)-R_n(u, x))du,$$

which yields by the monotone convergence

$$M\leqq 0 \qquad \text{as } n\to\infty.$$

We shall next verify $M \geqq 0$. Clearly

$$\begin{aligned} M &\geqq R^*(t, x)-R^*(s, x)+\int_s^t L^{\pi^{(n)}}R^*(u, x)du \\ &\quad -R_n(t, x)+R_n(s, x)-\int_s^t L^{\pi^{(n)}}R_n(u, x)du \\ &\geqq -(R^*(s, x)-R_n(s, x)) \\ &\quad -\alpha\int_s^t (R^*(u, x)-R_n(u, x))du \\ &\quad -\int_s^t p(x, \pi^{(n)}(u, x))(R^*(u, x)-R_n(u, x))du\,. \end{aligned}$$

Therefore we have

$$M \geqq 0 \qquad \text{as } n \to \infty\,.$$

Thus we have proved that

$$R^*(t, x)-R^*(s, x)+\int_s^t L^{\pi^*}R^*(u, x)du = 0 \tag{23}$$

for each t, $s \in J$ and $x \in S$.

Put $t=T$ in (23), by virture of the trivial equality

$$R^*(T, x) = \lim_n R_n(T, x) = \varphi(x) \qquad \text{for each } x \in S,$$

then it follows that

$$\varphi(x)-R^*(s, x)+\int_s^T L^{\pi^*}R^*(u, x)du = 0 \qquad \text{for each } s \text{ and } x\,.$$

This is what we wanted to verify.

Acknowledgment.

The author would like to thank Professor N. Furukawa of Kyushu Univ. for his helpful advice and encouragement in preparing this paper.

References

[1] Miller, B.L., *Finite state continuous time Markov decision processes with a finite planning horizon.* SIAM J. Control Vol. 6, No. 2, (1968).

[2] Feller, W., *On the integro-differential equations of purely disconiinuous Markoff processes.* Trans. Amer. Math. Soc., Vol. 48, (1940); Errata Vol. 58, (1945).

[3] Tsuji, K. and N. Furukawa, *On the existence of optimal controls in a nonlinear differential system.* Memo. Fac. Sci., Kyushu Univ., Vol. 22, No. 2, (1968).

[4] Fleming, W.H., *Some Markovian optimization problems.* J. Math. Mech., Vol. 12, Nc. 1, (1963).