STOPPED DECISION PROCESS ON COMPACT METRIC SPACES

Iwamoto, Seiichi
Department of Mathematics, Kyushu University

https://doi.org/10.5109/13044

出版情報：統計数理研究. 14 (1/2), pp.51-60, 1970-03. Research Association of Statistical Sciences
バージョン：
権利関係：

STOPPED DECISION PROCESSES ON COMPACT METRIC SPACES

By

Seiichi IWAMOTO*

(Received February 5, 1970)

§1. Introduction.

A stopped decision process is specified by six objects; S, A, q, r, g and α. S, the set of the possible states of some system, is a non-empty Borel subset of a Polish space and A, the set of actions available to you, is a non-empty Borel subset of the Polish space. At each time n, suppose that when the system is in s two alternatives are possible (a) action a is chosen or (b) the next sampling is terminated (stopped). The election (a) moves the system to a new state s' (which will be the state you observe tomorrow) according to the distribution $q(\cdot\,|\,s, a)$ and consequently you receive a reward $r(s, a, s')$, and election (b) terminates the process and you will receive the terminal reward $g(s)$. Since α (<1) is a discount factor, unit reward on the n-th day is worth only α^{n-1}.

A stopped policy is pair (π, τ). The policy π is a sequence $\pi_1, \pi_2, \cdots$ where π_n tells you how to select an action on the n-th day as a function of the previous history $h=(s_1, a_1, \cdots, a_{n-1}, s_n)$ of the system, by associating with each h (Borel measurably) a probability distribution $\pi_n(\cdot\,|\,h)$ on the Borel field of A. τ is an essentially finite, non-anticipative integer valued random variable. Then such a τ defines a stopping rule: If $\tau=n$, then choose (b) and the process is terminated at time n; if $\tau>n$ choose (a) at time n and the system goes to $(n+1)$-st stage.

The stopped policy (π, τ) associates with each initial state s_1 a corresponding n-th day return

$$\sum_{k=1}^{n-1}\alpha^{k-1}r(s_k, a_k, s_{k+1})+\alpha^{n-1}g(s_n)$$

at $\tau=n$, where for each k, s_k is the state of the system, and a_k the action at time k.

Note that only when you take the stopping rule $\tau=n$, you can receive an income which is the sum of the discounted reward by choosing (a) up to date and the discounted one gained by stopping at $\tau=n$.

The stopped decision problem is to maximize your total expected reward over the future

A Baire function f from S into A defines a policy: when in state s at time n, choose an action $f(s)$ independently of when and how you have arrived at state s, where τ is the paired stopping time. Such a policy will be called stationary policy,

* Department of Mathematics, Kyushu University, Fukuoka.

which is denoted by $f^{(\infty)}$. A stationary stopped-policy is one whose policy is stationary and whose stopping time is Markov (Section 3).

Recently the stopped decision problem has been studied extensively by N. Furukawa and S. Iwamoto ([2]), and the existence of stationary stopped-policies, which are optimal in some senses, was shown under some assumptions.

In this paper, we will make three additional assumptions about A, q, r and g. In Section 4, the following assumptions will remain operative; (1) A is a compact metric space, (2) $r(s, a, s')=r^*(s, a)+r^{**}(s')$, where r^*, r^{**} and g are bounded upper semi-continuous and (3) if $s_n \to s$, $a_n \to a$, then $q(\cdot \mid s_n, a_n)$ converges weakly to $q(\cdot \mid s, a)$. Our result is the following: under these assumptions, there always exists an ε-optimal stationary stopped-policy for any $\varepsilon > 0$.

Our method of proof is based on two facts; (A) The optimality equation has at most one bounded solution ([2]), and (B) Selection Theorem, due to Dubins and Savage ([3]), which was proved by A. Maitra in detail.

In Section 2 we give the notations and definitions to be used throughout this paper, and in Section 3 we shall define the stopped decision process. Section 4 is devoted to the existence theorem of ε-optimal stationary stopped policy, preparing the fundamental lemmas.

§ 2. Notations and probabilistic definitions.

This section is devoted to an exposition of the basic notations and probilistic definitions needed for the following sections. We follow [2] as closely as possible.

By a Borel set we mean a Borel subset of some Polish (i. e. complete separable metric) space. By a probability measure on a non-empty Borel set X we mean a probability measure defined on the Borel field of X, and the class of all probability measures on X is denoted by $P(X)$. For any non-empty Borel sets X, Y, a conditional probability measure on Y given X is denoted by $Q(Y \mid X)$. The class of all bounded Baire functions on X is denoted by $M(X)$. $C_1(X)$ is the class of all bounded upper semi-continuous functions on X. It is clear that $C_1(X) \subseteq M(X)$. If $u, v \in M(X)$, $u \geqq v$ means $u(x) \geqq v(x)$ for all $x \in X$. For any $p \in P(X)$ and any $u \in M(X)$, pu denotes the integral of u with respect to p. The product space of X and Y will be denoted by XY. For any $u \in M(XY)$ and any $q \in Q(Y \mid X)$, qu denotes the element of $M(X)$ whose value at $x_0 \in X$ is given by $qu(x_0)=\int_Y u(x_0, y)dq(y \mid x_0)$.

A $p \in P(X)$ is degenerate if $p\{x_0\}=1$ for some $x_0 \in X$, and a $q \in Q(Y \mid X)$ is degenerate if $q(\cdot \mid x)$ is degenerate for each x. The degenerate q are exactly those for which there is a Baire function f mapping X into Y such that

$$q(\{f(x)\} \mid x)=1 \qquad \text{for all } x \in X.$$

Any such function f will also denote its associated degenerate q, so that, for any $u \in M(XY)$,

$$fu(x)=u(x, f(x)) \qquad \text{for all } x \in X.$$

§ 3. Stopped decision processes.

In general the optimization problem in our stopped decision processes is defined by six elements S, A, q, r, g and α. The state space S and the action space A are non-empty Borel sets, and the law of motion q is a sequence of transition probabilities; $q=\{q_1, q_2, \cdots\}$, where each $q_n \in Q(S|H_nA)$ and $H_n=SA\cdots S(2n-1$ factors) is the set of possible histories of the system when the n-th alternative must be chosen. The immediate reward function r is a Baire function on SAS, the terminal reward function g is a Baire function on S and α is a discount factor, $0\leqq\alpha<1$.

But we treat the case only where

$$r\in M(SAS) \quad \text{and} \quad g\in M(S).$$

A policy π is a finite sequence $\{\pi_1, \pi_2, \cdots, \pi_N\}$ or a denumerable sequence $\{\pi, \pi_2, \cdots\}$, where each $\pi_n \in Q(A|H_n)$. The policy π is Markov if each π_n is a degenerate element of $Q(A|S)$, i.e. $\pi=\{f_1, f_2, \cdots\}$, where each f_n is a Baire function from S into A, and stationary if it is Markov and there exists a Baire function f from S into A such that $\pi_n=f$ for all n.

Any policy π, together with q, defines a conditional probability on the countable product set $ASAS\cdots$given that the initial state s, i.e. it defines

$$e_\pi=\pi_1q_1\pi_2q_2\cdots\in Q(ASAS\cdots|S).$$

Let $\mathfrak{S}$ and $\mathfrak{A}$ be the Borel fields of S and A respectively. Let the sample space $\Omega=SASAS\cdots$. The sample point ω is expressed as $\omega=(s_1, a_1, s_2, a_2, \cdots)$, where each $s_n\in S$ and each $a_n\in A$. $\mathfrak{F}_n$ denotes the family of all unions of sets in Ω of the form

$$\{\omega|s_1\in E_1, a_1\in F_1, s_2\in E_2, a_2\in F_2, \cdots, s_n\in E_n\}$$

where $E_i\in\mathfrak{S}$ and $F_i\in\mathfrak{A}$ for $i=1, 2, \cdots, n$. Let π be any policy. Then a stopping time associated with π is a random variable $t=t(\omega)$ with positive integer values such that

$$e_\pi[\{t(\omega)<\infty\}_{s_1}|s_1]=1 \quad \text{for all } s_1\in S$$

and

$$\{t(\omega)=n\}\in\mathfrak{F}_n \quad \text{for each } n$$

where by $\{\cdots\}$ we mean the set of all ω for which the indicated relation holds, and $\{\cdots\}_{s_1}$ means the s_1-section of the ω-set. Let $C(\pi)$, for each π, be the class of all stopping times associated with π. For any policy π and $t\in C(\pi)$, a pair (π, t) is called the stopped-policy (abbreviated as s-policy). $C^N(\pi)$ denotes the class of all stopping times $t\in C(\pi)$ for which

$$e_\pi[\{t(\omega)\leqq N\}_{s_1}|s_1]=1 \quad \text{for all } s_1\in S.$$

An s-policy (π, t) is called the truncated-policy (abbreviated as t-policy) if there exists an integer N such that

$$t\in C^N(\pi).$$

Let $\mathfrak{B}_n$ denote the family of all unions of sets in Ω of the form

$$\{\omega|s_1,\in E_1, s_2\in E_2, \cdots, s_n\in E_n\}$$

where $E_i \in \mathfrak{S}$ and $\mathfrak{G}_n$ the family of all ω-sets of the form

$$\{\omega \mid s_n \in E_n\}$$

where $E_n \in \mathfrak{S}$. A $t \in C(\pi)$ is Markov if for each n

$$\{t(\omega) = n\} \in \mathfrak{B}_n$$

and $\{t(\omega) = n\}$ can be written in the form

$$\{t(\omega) = n\} = \{t(\omega) > n-1\} \cap \Delta_n \qquad \text{for each } n$$

where $\Delta_n \in \mathfrak{G}_n$. An s-policy (π, t) is Markov if both π and t are Markov. A Markov s-policy is stationary, if π is so. A Markov $t \in C(\pi)$ is stationary, if Δ_n does not depend on n.

Let

$$x_n = \sum_{k=1}^{n-1} \alpha^{k-1} r(s_k, a_k, s_{k+1}) + \alpha^{n-1} g(s_n)$$

$$= \sum_{k=1}^{n-1} \alpha^{k-1} r_k + \alpha^{n-1} g_n \qquad \text{for each } n,$$

where $r_k = r(s_k, a_k, s_{k+1})$ and $g_n = g(s_n)$.

E^{π} denotes the expectation with respect to e_{π}. The expected total reward from an s-policy (π, t), as a function of the initial state s_1, is then

$$E^{\pi}(x_t)(s_1) = \int_{ASAS\cdots} x_t \, de_{\pi}(\cdot \mid s_1)$$

$$= \sum_{n=1}^{\infty} \int_{\{t=n\}} x_n \, d(\pi_1 q_1 \pi_2 q_2 \cdots \pi_{n-1} q_{n-1})(\cdot \mid s_1)$$

$$= \sum_{n=1}^{\infty} \int_{\{t=n\}} \left[\sum_{k=1}^{n-1} \alpha^{k-1} r_k + \alpha^{n-1} g_n \right] d(\pi_1 q_1 \pi_2 q_2 \cdots \pi_{n-1} q_{n-1})(\cdot \mid s_1)$$

$$= \int_{ASAS\cdots} \left[\sum_{k=1}^{t-1} \alpha^{k-1} r_k + \alpha^{t-1} g_t \right] de_{\pi}(\cdot \mid s_1).$$

Let Π denote the class of all policies and Λ the class of all stopped policies, i. e.

$$\Lambda = \{(\pi, t) \mid \pi \in \Pi, \ t \in C(\pi)\}.$$

An s-policy (π^*, t^*) is called *ε-optimal* if

$$E^{\pi^*}(x_{t^*}) \geqq E^{\pi}(x_t) - \varepsilon \qquad \text{for every } (\pi, t) \in \Lambda$$

and is called *optimal* if

$$E^{\pi^*}(x_{t^*}) \geqq E^{\pi}(x_t) \qquad \text{for every } (\pi, t) \in \Lambda.$$

§ 4. ε-optimal stationary stopped-policies.

In this section we shall assume that A is a compact metric space, r a bounded upper semi-continuous function on SAS, q a bounded upper semi-continuous function on S and $q_1 = q_2 = \cdots = q \in Q(S \mid SA)$ is continuous, that is, $(s_n, a_n) \to (s, a)$ implies that $q(\cdot \mid s_n, a_n)$ converges weakly to $q(\cdot \mid s, a)$.

We consider the case only where the discount factor $\alpha<1$, and the motion-law of the system is homogeneous Markov We shall use the notations and lemmas closely those of [2].

With any degenerate $f\in Q(A|S)$ we associate the operators T_f and A_f, mapping $M(S)$ into $M(S)$ defined by

$$T_f u(s)=\int [r(s, f(s), \cdot)+\alpha u(\cdot)]\, dq(\cdot \mid s, f(s)) \tag{4.1}$$

$$A_f u(s)=\max [g(s), T_f u(s)] \tag{4.2}$$

respectively. $T_f u(s)$ may be interpretted as the expected return if you are in state s, take an action $f(s)$, and receive a final return $u(s')$ at the resulting state s'. $A_f u(s)$ is the optimal return, when the system allows you to select two possible alternatives in the state s (a) taking an action $f(s)$ or (b) terminating the process. The following properties of the operators T_f, A_f, formulated as a lemma, are immediate from the definition.

LEMMA 4.1. (a) *T_f, A_f are monotone operators.*

(b) *For any $u\in M(S)$, constant c, $T_f(u+c)=T_f u+\alpha c$.*

(c) *For any $u\in M(S)$, constant $c>0$, $A_f(u+c)\leqq A_f u+\alpha c$.*

PROOF. The proofs are stated in Lemma 7.1 of [2].

Here are some basic properties which will take an important role in proving our desired results.

LEMMA 4.2. (a) *With any Baire function $f: S\to A$, the operator A_f is a contraction mapping on $M(S)$ with contraction coefficient α.*

(b) *For any $u, v\in M(S)$ and any Markov policy $\pi=\{f_1, f_2, \cdots\}$, it holds that*

$$\lim_{n\to\infty} \| A_{f_1}A_{f_2}\cdots A_{f_n}u-A_{f_1}A_{f_2}\cdots A_{f_n}v\|=0\,.$$

PROOF. (a), (b) are stated in Lemma 7.2 of [2].

We say that a $u\in M(S)$ satisfies the optimality equation, if it holds that

$$u=\sup_{a\in A} A_a u\,,$$

where A_a is the operator A_f with $f\equiv a$. The principal general result on the optimality equation is contained in the following lemma.

LEMMA 4.3. *The optimality equation has at most one bounded solution.*

PROOF. This lemma is stated as Lemma 7.15 in [2].

But under the assumptions stated at the beginning of this section it will be shown that the optimality equation has a unique bounded solution in $C_1(S)$.

On the other hand we introduce a process $\{\beta_n^N, n=1, 2, \cdots, N\}$ for any fixed integer $N\geqq 1$, which is closely related to the operator A_f. For any policy π we define for each $N\geqq 1$ a finite sequence of random variables $\{\beta_n^N(\pi);\ 1\leqq n\leqq N\}$, by backward induction:

$$\begin{cases} \beta_n^N(\pi)=\max [x_n, \pi_n q_n(\beta_{n-1}^N(\pi))]\,, & n=1, 2, \cdots, N-1\,, \\ \beta_N^N(\pi)=x_N\,. \end{cases} \tag{4.3}$$

We now define the stopping time τ_N

$$\tau_N(\pi) = \text{the first integer } n \text{ such that } \beta_n^N(\pi) = x_n . \tag{4.4}$$

Then it is obvious that

$$\tau_N(\pi) \leqq C^N(\pi) .$$

Let

$$\alpha^{n-1} v_n^N(\pi) = \beta_n^N(\pi) - \sum_{k=1}^{n-1} \alpha^{k-1} r_k$$

(4.3) then turns out to become

$$\begin{cases} v_n^N(\pi) = \max [g_n, \pi_n q_n \{\alpha v_{n+1}^N(\pi) + r_n\}] , & n = 1, 2, \cdots, N-1 , \\ v_N^N(\pi) = g_N . \end{cases}$$

We shall need the following lemma, in particular, for the stationary policy $\pi = f^{(\infty)}$.

LEMMA 4.4. (a) $E^\pi(x_{\tau_N}) = E^\pi(\beta_{\tau_N}^N(\pi)) = \beta_1^N(\pi) = v_1^N(\pi)$.

(b) *Let* $\tau_N(\pi)$ *be defined by* (4.4), *then it follows that*

$$\tau_N(\pi) = \textit{the first integer } n \textit{ such that } v_n^N(\pi) = g_n .$$

(c) *For any Markov* $\pi = \{f_1, f_2, \cdots\}$

$$A_{f_1} A_{f_2} \cdots A_{f_N} g = v_1^{N+1}(\pi) \qquad \textit{for each } N \geqq 1 .$$

PROOF. Lemma 4.5 and 6.2 in [2] show these results.

We have the following lemma, the proof of which is due to [1].

LEMMA 4.5. *Let* u *be a bounded upper semi-continuous on* SA. *Define* $u^* : S \to R$ *by*

$$u^*(s) = \max_{a \in A} u(s, a) .$$

Then u^* *is bounded upper semi-continuous.*

LEMMA 4.6. *Let* $w : S \to R$ *be a bounded upper semi-continuous. The* $h : SA \to R$ *defined by*

$$h(s, a) = \int w(\cdot) dq(\cdot \mid s, a)$$

is bounded upper semi-continuous.

PROOF. If w is continuous, then clearly h is continuous. Now if w is bounded and upper semi-continuous, there exists a sequence of bounded continuous functions $\{w_n, n = 1, 2, \cdots\}$ such that

$$w_n \downarrow w \quad \text{on} \quad S .$$

Let $h_n(s, a) : SA \to R$ be defined by

$$h_n(s, a) = \int w_n(\cdot) dq(\cdot \mid s, a) .$$

Each h_n is continuous and, moreover, by the dominated convergence theorem,

$$h_n \downarrow h \quad \text{on} \quad SA . \tag{4.5}$$

If $(s_n, a_n) \to (s, a)$, then for any $\varepsilon > 0$ there exist integers N_1, N_2 such that

$$n \geqq N_1 \text{ implies } h_n(s, a) \leqq h(s, a) + \frac{\varepsilon}{3} \tag{4.6}$$

and

$$m \geqq N_2 \text{ implies } |h_n(s, a) - h_n(s_m, a_m)| \leqq \frac{\varepsilon}{3} \qquad \text{for } n = 1, 2, \cdots .$$

Therefore

$$m \geqq N_2 \text{ implies } h_n(s_m, a_m) \leqq h_n(s, a) + \frac{\varepsilon}{3} \qquad \text{for } n = 1, 2, \cdots . \tag{4.7}$$

By (4.5),

$$h(s_m, a_m) \leqq h_n(s_m, a_m) + \frac{\varepsilon}{3} \qquad \text{for } n = 1, 2, \cdots . \tag{4.8}$$

Combining these results expressed in (4.6), (4.7) and (4.8) yields the desired relation

$$n \geqq N_1 \text{ and } m \geqq N_2 \text{ imply } h(s_m, a_m) \leqq h(s, a) + \varepsilon .$$

Since ε is arbitrary we obtain

$$\limsup_{m \to \infty} h(s_m, a_m) \leqq h(s, a) .$$

Clearly h is bounded. This completes the proof.

The following theorem due to Dubins and Savage ([3]), was proved in detail by A. Maitra ([1]).

Selection Theorem (A. Maitra [1]): *Let u be a bounded upper semi-continuous on SA. Then there exists a Baire function f from S into A such that*

$$u(s, f(s)) = \max_{a \in A} u(s, a) \qquad \textit{for all} \quad s \in S .$$

For $u, v \in C_1(S)$ we define d_1 by

$$d_1(u, v) = \|u - v\| = \sup_{s \in S} |u(s) - v(s)| ,$$

then the following lemma is established by A. Maitra ([1]).

LEMMA 4.7. *$(C_1(S), d_1)$ is a complete metric space.*

PROOF. It suffices to show that $C_1(S)$ is closed under the uniform convergence. The detailed proof will be omitted here.

Throughout the remainder of this section we will assume that

$$r = r^* + r^{**}$$

where $r^* \in C_1(SA)$, $r^{**} \in C_1(S)$ and $r(s, a, s') = r^*(s, a) + r^{**}(s')$. This decomposition of r into r^* and r^{**} means that the reward $r^{**}(s')$ related to the new state s' is free from the current state s and current action a. Of course r^* and r^{**} are not necessarily constant sign. Therefore this restriction on r is rather slight.

For any $u \in C_1(S)$, let $Au : S \to R$ be the function defined by

$$Au(s) = \max_{a \in A} A_a u(s) . \tag{4.9}$$

Note that the operator T_a ($= T_f$, where $f \equiv a$), which is the contraction mapping on $M(S)$, maps $C_1(S)$ into $C_1(S)$, since for $u \in C_1(S)$

$$\begin{aligned} T_a u(s) &= \int_S [r(s, a, \cdot) + \alpha u(\cdot)] \, dq(\cdot \mid s, a) \\ &= \int_S [r^*(s, a) + r^{**}(\cdot) + \alpha u(\cdot)] \, dq(\cdot \mid s, a) \\ &= r^*(s, a) + \int_S [r^{**}(\cdot) + \alpha u(\cdot)] \, dq(\cdot \mid s, a) \end{aligned}$$

implies $T_a u \in C_1(S)$ for each $a \in A$ because of Lemma 4.6. It is also a contraction mapping on $C_1(S)$ with contraction coefficient α for all $a \in A$. Since for $u, v \in C_1(S)$ $\max\{u, v\} \in C_1(S)$, (4.2) implies that the operator A_a is a contraction mapping on $C_1(S)$ with contraction coefficient α for all $a \in A$.

For any $u \in C_1(S)$, from (4.9)

$$Au(s) = \max_{a \in A} \max\left[g(s), \int_S [r(s, a, \cdot) + \alpha u(\cdot)] dq(\cdot \mid s, a)\right]$$

$$= \max_{a \in A} \max\left[g(s), \int_S [r^*(s, a) + r^{**}(\cdot) + \alpha u(\cdot)] dq(\cdot \mid s, a)\right]. \tag{4.9}'$$

In this expression, $r^{**}(\cdot) + \alpha u(\cdot)$ is upper semi-continuous on S and by virture of Lemma 4.6 integral within square bracket on the right-hand side of (4.9)′ is upper semi-continuous in (s, a). Consequently

$$\max\left[g(s), \int_S [r(s, a, \cdot) + \alpha u(\cdot)] dq(\cdot \mid s, a)\right] \tag{4.10}$$

is also upper semi-continuous in (s, a), and maximum is attained for all $s \in S$ because the action space A is compact. Since (4.10) is obviously bounded, Lemma 4.5 implies that Au is upper semi-continuous. Thus the operator A maps $C_1(S)$ into $C_1(S)$.

LEMMA 4.8. *The operator A defined by* (4.9) *is a contraction mapping on $C_1(S)$, with contraction coefficient α, and consequently, has a unique fixed point.*

PROOF. Let $u_1, u_2 \in C_1(S)$. Clearly $u_1 \leqq u_2 + \|u_1 - u_2\|$. Lemma 4.1 yields

$$A_a u_1 \leqq A_a(u_2 + \|u_1 - u_2\|) \leqq A_a u_2 + \alpha \|u_1 - u_2\| \qquad \text{for all} \quad a \in A.$$

Then

$$\sup_{a \in A} A_a u_1 \leqq \sup_{a \in A} (A_a u_2 + \alpha \|u_1 - u_2\|) = \sup_{a \in A} A_a u_2 + \alpha \|u_1 - u_2\|.$$

In this case 'sup' can be replaced by 'max' because of the Lemma 4.5. Therefore

$$Au_1 - Au_2 \leqq \alpha \|u_1 - u_2\|.$$

Interchanging u_1 and u_2, we get

$$Au_2 - Au_1 \leqq \alpha \|u_1 - u_2\|.$$

Hence, $\|Au_1 - Au_2\| \leqq \alpha \|u_1 - u_2\|$, which proves that operator A is a contraction mapping, as $\alpha < 1$. Since $C_1(S)$ is a complete metric space (Lemma 4.7), it follows from the Banach Fixed Point theorem that the operator A has a unique fixed point in $C_1(S)$. This completes the proof.

We set the following assumptions.

(H1) $r = r' - r''$, where r' and r'' are the real-valued Baire functions on SAS and $r'' \geqq 0$.

(H2) $E^{\pi}(x_t') < \infty$ for all $(\pi, t) \in \Lambda$, where $x_n' = \sum_{k=1}^{n-1} \alpha^{k-1} r'(s_k, a_k, s_{k+1}) + \alpha^{n-1} g(s_n)$.

(H3) $\{(x_n')^-, n \geqq 1\}$ is uniformly integrable $w.r.t.\, e_{\pi}$ for each $\pi \in \Pi$, where $a^- = \max(0, -a)$.

The following lemma is given as Lemma 7.12 in [2].

LEMMA 4.9. *Let* (H1), (H2) *and* (H3) *be satisfied. Any $u \in M(S)$ for which $A_a u \leqq u$*

for all $a \in A$ *satisfies that* $E^{\pi}(x_t) \leqq u$ *for all* $(\pi, t) \in \Lambda$.

Finally we have the desired result on the basis of above lemmas.

THEOREM. *Assume* (H1), (H2) *and* (H3). *Then for any* $\varepsilon > 0$ *there exists an* ε-*optimal stationary s-policy which is t-policy.*

PROOF. By Lemma (4.8) the operator A has a unique fixed point u^* in $C_1(S)$, so that

$$u^* = Au^* = \max_{a \in A} A_a u^*, \quad \text{where} \quad u^* \in C_1(S). \tag{4.11}$$

For this u^*, function $u : SA \to R$ defined by

$$u(s, a) = \max\left[g(s), \int_S [r(s, a, \cdot) + \alpha u^*(\cdot)]\, dq(\cdot \mid s, a)\right]$$

is bounded upper semi-continuous on SA. From the Selection Theorem there exists a Baire function f from S into A such that

$$u(s, f(s)) = \max_{a \in A} u(s, a) \qquad \text{for all} \quad s \in S.$$

Consequently this function defines the operator A_f on $M(S)$ such that

$$u^* = Au^* = \max_{a \in A} A_a u^* = A_f u^*.$$

On the other hand from the Lemma 4.2 the operator A_f is a contraction mapping on $M(S)$. Then by virture of Banach Fixed Point theorem, it has a unique fixed point w^*

$$A_f w^* = w^*, \quad \text{where} \quad w^* \in M(S).$$

The uniqueness of the fixed point of A_f together with the fact $C_1(S) \subseteq M(S)$ yields

$$u^* = w^*.$$

Hence $A_f u^* = u^*$, so that $(A_f)^n u^* = u^*$ for all $n \geqq 1$. From Lemma 4.2, for any $\varepsilon > 0$ there exists an integer N such that $n \geqq N$ implies

$$(A_f)^{n-1} g \geqq (A_f)^{n-1} u^* - \varepsilon = u^* - \varepsilon. \tag{4.12}$$

Since u^* satisfies the optimality equation, $u^* \geqq A_a u^*$ for all $a \in A$.

Then by Lemma 4.9

$$u^* \geqq E^{\pi}(x_t) \qquad \text{for all} \quad (\pi, t) \in \Lambda. \tag{4.13}$$

Combining the results expressed in (4.12) and (4.13) yields the relation

$$(A_f)^{n-1} g \geqq E^{\pi}(x_t) - \varepsilon \text{ for any } (\pi, t) \in \Lambda, \quad \text{where} \quad n \geqq N. \tag{4.14}$$

Lemma 4.4 allows us to write

$$E^{f^{(\infty)}}(x_{\tau_n(f)}) = (A_f)^{n-1} g \tag{4.15}$$

where $f^{(\infty)} = (f, f, \cdots)$ and

$$\tau_n(f) = \text{the first integer } k \text{ such that } v_k^n(f^{(\infty)}) = g_k.$$

Finally, using (4.14) and (4.15) we obtain the final relation

$$E^{f^{(\infty)}}(x_{\tau_n(f)}) \geqq E^{\pi}(x_t) - \varepsilon \text{ for all } (\pi, t) \in \Lambda \quad \text{and} \quad n \geqq N.$$

It is obvious that s-policy $(f^{(\infty)}, \tau_n(f))$ is truncated, i. e.

$$(f^{(\infty)}, \tau_n(f)) \in C^n(f^{(\infty)})$$

and that it is ε-optimal stationary s-policy for $n \geqq N$. This completes the proof.

Moreover if we put the following assumptions in addition to (H1) and (H3) we can find an optimal s-policy which has a stationary stopping time.

(H4) $\lim_{n\to\infty} x_n'' = +\infty$ $w.e_{\pi}.1$ for all $\pi \in \Pi$, where $x_n'' = \sum_{k=1}^{n-1} \alpha^{k-1} r''(s_k, a_k, s_{k+1})$.

(H5) (i) $\{(x_n')^-, n \geqq 1\}$ is uniformly integrable $w.r.t.e_{\pi}$, and (ii) $\sup_N E^{f^{(\infty)}}(x_N' \cdot f) < \infty$ for every stationary $f^{(\infty)}$.

COROLLARY. *Assume* (H1), (H2), (H4) *and* (H5). *Then there exists an optimal stationary s-policy having a stationary stopping time.*

PROOF.

$$\lim_{n\to\infty} (A_f)^n g = v_1(f^{(\infty)}) = u^*, \qquad v_n(f^{(\infty)}) = \lim_{N\to\infty} v_n^N(f^{(\infty)})$$

and τ be the first n such that $v_n(f^{(\infty)}) = g_n$.

The arguments in [2], Corollary 7.2, Theorem 8.2, show us the facts

$$\tau \in C(f^{(\infty)})$$

and

$$\sup_{t \in C(f^{(\infty)})} E^{f^{(\infty)}}(x_t) = \lim_{N\to\infty} E^{f^{(\infty)}}(x_{\tau_N(f)}),$$

which completes the proof.

In conclusion we remark that in paper [1], even if $r(s, a)$ is replaced by $r(s, a, s') = r^*(s, a) + r^{**}(s')$, where $r^*(s, a) \in C_1(SA)$, $r^{**}(s') \in C_1(S)$, the Maitra's result remains true: there exists a stationary plan.

Acknowledgement. I would like to thank Dr. N. Furukawa, under whose guidance this paper was prepared.

References

[1] A. MAITRA, *Discounted dynamic programming on compact metric space*, Sankhyā. series A. vol **30**, (1968), pp. 211-216.

[2] N. FURUKAWA and S. IWAMOTO, *Stopped Decision Processes on Complete Separable Metric Spaces*. to appear. J. Math. Anal. Appl.

[3] L. E. DUBINS and L. J. SAVAGE, *How to Gamble If You Must*. MacGraw.Hill, New York (1965).