NEARLY OPTIMAL POLICIES AND STOPPING TIMES IN MARKOV DECISION PROCESSES WITH GENERAL REWARDS

Furukawa, Nagata
Department of Mathematics, Faculty of Science, Kyushu University

https://doi.org/10.5109/13142

出版情報：統計数理研究. 19 (1/2), pp.89-109, 1980-03. Research Association of Statistical Sciences
バージョン：
権利関係：

NEARLY OPTIMAL POLICIES AND STOPPING TIMES IN MARKOV DECISION PROCESSES WITH GENERAL REWARDS

By

Nagata FURUKAWA[*]

(Received December 27, 1979)

1. Introduction.

In the previous paper [7], the author has treated non-stationary Markov decision processes associated with optimal choice of stopping times and studied, in connection with the Markov potential theory, a functional equation satisfied by an optimal return, the so-called optimality equation.

The purpose of the present paper is to give some theoretical foundations the previous one [7] based on for its whole part. In this paper non-stationary Markov decision processes with generalized utilities are formulated in the form of the gambling theory in the sense of Dubins and Savage [5]. The optimization is made with respect to a pair of policy and stopping time. In Section 2 we shall prepare the probabilistic definitions and notations which will be used throughout the paper. In Section 3 we shall give the reduction of the fortune space. In Section 4, under the assumption that the utility function is recursive and monotone, we shall show the existence of a $(\bar{p}, \varepsilon)$-optimal bounded pair of policy and stopping time and the existence of a bounded pair which is (p, ε)-optimal. Section 4 will be also concerned with the case of finite horizon. In Section 5 we shall prove under some additional assumptions that there exists a Markov bounded pair which is (p, ε)-optimal. Section 6 gives several examples of recursive and monotone utility functions which satisfy all of the conditions required in this paper.

We like to close this section with refering to the work of Furukawa and Iwamoto [9] in which the utility function has been generalized but the authors had no concern with the stopping rules.

2. Preliminaries.

In this section we shall devolope the basic notation and definitions to be used throughout the paper.

First we shall give general probabilistic notation and definitions following closely those of [2]. By a Borel set we mean a Borel subset of some Polish space. By a

[*] Department of Mathematics, Faculty of Science, Kyushu University, Fukuoka

probability measure on a non-empty Borel set X we mean a countably additive probability measure defined on the Borel σ-field of X, and the set of all probability measures on X is denoted by $P(X)$. For any non-empty Borel sets X, Y, $Q(Y|X)$ is the set of all regular conditional probability measures $q(\cdot|\cdot)$. For any non-empty Borel set X, let $B(X)$ denote the set of all real-valued Borel measurable functions on X. For any $p\in P(X)$, $u\in B(X)$, pu denotes the integral of u with respect to p, supposing that the integral is well-defined. For any $u\in B(X\times Y)$, where $X\times Y$ is the cartesian product of X and Y, and any $q\in Q(Y|X)$, qu denotes the element of $B(X)$ whose value at $x_0\in X$ is given by

$$qu(x_0)=\int_Y u(x_0,\ y)dq(y|x_0)\,,$$

supposing that the integral is well-defined. We can extend the above notation in an obvious way to a finite or countable sequence of non-empty Borel sets. The details are omitted.

For any non-empty Borel sets X, Y, and any non-empty Borel subset Γ of $X\times Y$, let $\Gamma(x)$ denote the x-section of Γ. For any non-empty Borel set Γ in $X\times Y$ such that $\Gamma(x)\neq\emptyset$ for all $x\in X$, $Q(\{\Gamma(x)\}|X)$ denotes the set of all elements of $Q(Y|X)$ satisfying that $q(\Gamma(x)|x)=1$ for all $x\in X$, supposing that $Q(\{\Gamma(x)\}|X)$ is not empty. In general, for any non-empty Borel sets X_1, X_2, $\cdots$, X_{n+1}, and any non-empty Borel subset Γ of $X_n\times X_{n+1}$ such that $\Gamma(x_n)\neq\emptyset$ for all $x_n\in X_n$, $Q(\{\Gamma(x_n)\}|X_1\times X_2\times\cdots X_n)$ denotes the set of all elements of $Q(X_{n+1}|X_1\times X_2\times\cdots\times X_n)$ satisfying that $q(\Gamma(x_n)|x_1, x_2,\cdots,x_n)=1$ for all $(x_1, x_2,\cdots, x_n)\in X_1\times X_2\times\cdots\times X_n$.

For any non-empty Borel set X, $^*\sigma$-algebra on $P(X)$ is the smallest σ-field of subsets of $P(X)$ which makes $p(B)$ measurable in $p\in P(X)$ for every Borel subset B of X. Then $P(X)$ is a Borel set, and $^*\sigma$-algebra on $P(X)$ the σ-field of its Borel subsets ([4]).

Next we shall give the notation and definitions peculiar to our decision problem. Our decision problem is specified by four elements, F, $\tilde{\Gamma}$, g and d. The set of *fortunes* F is the cartesian product of $\boldsymbol{N}$, S and $\boldsymbol{R}$, where the time space $\boldsymbol{N}$ is the set of all nonnegative integers, the set of *states* S of some system is a non-empty Borel set, and $\boldsymbol{R}$ is the space of real numbers. We may interpret $\boldsymbol{R}$, for instance, the space of moneys we get. We shall denote the cartesian product of $\boldsymbol{N}$ and S by $Z=\boldsymbol{N}\times S$. The set of all subsets of $\boldsymbol{N}$ and the σ-field of Borel subsets of S are denoted by $\mathscr{B}_{\boldsymbol{N}}$ and $\mathscr{B}_S$, respectively. $\mathscr{B}_Z$ denotes the product σ-field of $\mathscr{B}_{\boldsymbol{N}}$ and $\mathscr{B}_S$. The *gambling house* $\tilde{\Gamma}$ is a subset of $F\times P(Z)$ such that $\tilde{\Gamma}(f)\neq\emptyset$ for all $f\in F$. We think of $\tilde{\Gamma}(f)$ as the set of *gambles available to us at* f. g is a real-valued Borel measurable function defined on the set $\bigcup_{n=0}^{\infty}[Z_0\times P(Z_1)\times Z_1\times P(Z_2)\times\cdots\times Z_n\times\boldsymbol{R}]$, where $Z_i=Z$ for $i=0, 1,\cdots$, endowing $P(Z)$ with $^*\sigma$-algebra, and is assumed to satisfy that $g(z_0\,; c)=c$ for all $z_0\in Z$ and all $c\in\boldsymbol{R}$. We interpret, for each n, the restriction of g on the set $Z_0\times P(Z_1)\times Z_1\times\cdots\times Z_n\times\boldsymbol{R}$ as the *reward function over* $(n+1)$ *stages*. From this interpretation we shall call g the *generalized reward function*. We shall impose some conditions on the range and integral of g (Condition (B)), and assume, later on, that g has some properties peculiar to the class of dynamic programming problems (Condition (C) and (D)). The terminal reward function d is a bounded Borel measurable function defined on Z.

Let $f_n=(z_n, c_n)$, where $z_n \in Z$ and $c_n \in \boldsymbol{R}$, be a generic element of fortunes at the n-th stage. In this paper we shall assume f_n is of the form: $f_n=(z_n, g(z_0, p_0, z_1, p_1, \cdots, z_n ; 0))$, where $\{p_0, p_1, \cdots p_{n-1}\}$ is the sequence of gambles we have chosen successively up to the $(n-1)$-th stage. When we observe the current fortune $f_n=(z_n, g(z_0, p_0, z_1, p_1, \cdots, z_n ; 0))$ at the n-th stage, we may either stop at f_n or take a gamble p_n from $\tilde{\Gamma}(f_n)$. If we stop at f_n, then we get an output $g(z_0, p_0, z_1, p_1, \cdots, z_n ; d(z_n))$, and if we take p_n, then we move to a new fortune $f_{n+1}=(z_{n+1}, g(z_0, p_0, z_1, p_1, \cdots, p_n, z_{n+1} : 0))$ according to the probability measure extended from p_n onto $(\cdot, g(z_0, p_0, z_1, p, \cdots, p_n, \cdot ; 0))$.

Let $\mathscr{B}_F=\mathscr{B}_Z \times \mathscr{B}^1$ be the product σ-field on F, where $\mathscr{B}^1$ the Borel σ-field of $\boldsymbol{R}$. Throughout this paper we shall impose the following condition on the gambling house.

Condition (A).

(A1) $\tilde{\Gamma}(z, c)=\tilde{\Gamma}(z, c')$ for all $z \in Z$, all $c \in \boldsymbol{R}$ and all $c' \in \boldsymbol{R}$,

(A2) $\tilde{\Gamma}(n) \subset S \times \boldsymbol{R} \times P(\{n+1\} \times S)$ for all $n \in \boldsymbol{N}$, where $\tilde{\Gamma}(n)$ denotes the n-section of $\tilde{\Gamma}$ and $P(\{n+1\} \times S)$ denotes the set $\{p \in P(Z) \mid p(\{n+1\} \times S)=1\}$,

(A3) $\tilde{\Gamma}$ is *measurable*, i.e.,

(a) $\tilde{\Gamma} \in \mathscr{B}_F \times \Sigma$, where Σ is the $*\sigma$-algebra on $P(Z)$,

(b) there exists a Borel Selector of $\tilde{\Gamma}$ from F into $P(Z)$.

In the above condition, (A1) is very natural for the dynamic programming problems, and (A2) implies that any choice of a gamble from $\tilde{\Gamma}(n)$ enables the time to move to $(n+1)$ almost surely, wherever other components of the fortune move to. The concept of measurability of the gambling house stated in (A3) is very similar to that of Strauch ([15]), but not completely same as it. In his definition the measurable gambling house needs to be leavable. In our case, however, the leavability condition for $\tilde{\Gamma}$ contradicts the condition (A2). If we consider a modified leavability condition:

(b)′ $\delta(\{n+1, s, c\}) \in \tilde{\Gamma}(n, s, c)$ for all $n \in \boldsymbol{N}$, all $s \in S$, all $c \in \boldsymbol{R}$, then it is clear that (b)′ implies (b) and does not contradict (A2).

Let $\tilde{H}=F \times P(Z) \times F \times P(Z) \times \cdots$ be the set of all histories $\tilde{h}=(f_0, p_0, f_1, p_1, \cdots)$ and $\tilde{H}_n=F \times P(Z) \times F \times P(Z) \times \cdots \times F$ ($2n+1$ factors) the set of all partial histories $\tilde{h}_n=(f_0, p_0, f_1, p_1, \cdots, f_n)$. *A policy* π *over* $\tilde{H}$ is a sequence $\{\pi_0, \pi_1, \pi_2, \cdots\}$ where each $\pi_n \in Q(P(Z) \mid \tilde{H}_n)$. A policy π over $\tilde{H}$ is called *available in* $\tilde{\Gamma}$, if each $\pi_n \in Q(\{\tilde{\Gamma}(f_n)\} \mid \tilde{H}_n)$, i.e.,for each n, $\pi_n(\tilde{\Gamma}(f_n) \mid \tilde{h}_n)=1$ for all $\tilde{h}_n \in \tilde{H}_n$. A policy over $\tilde{H}$ defined as above may be called a family of random strategies, following Strauch's definition ([15]).

For each $(z_0, p_0, z_1, p_1, \cdots, z_n)$ and any $p \in P(Z)$, let Y denote the function $(z, g(z_0, p_0, z_1, \cdots, z_n, p, z ; 0))$ of z. Define the extension of p to F by $\tilde{p}(E)=p(Y^{-1}(E))$ for every $E \in \mathscr{B}_F$. Then $\tilde{p}$ is well-defined since Y is Borel measurable in z. For each $\sigma \in P(P(Z))$, define the extension of σ to $P(F)$ by $\tilde{\sigma}(B)=\sigma\{p \in P(Z) ; \tilde{p} \in B\}$ for every Borel subset B of $P(F)$ endowed with $*\sigma$-algebra, where $\tilde{p}$ is the extension of p defined above. Then $\tilde{\sigma}$ is well-defined because of the definition of the $*\sigma$-algebra. Let π be an arbitrary policy available in $\tilde{\Gamma}$. For each n and each $(f_0, p_0, f_1, \cdots, f_n)$, then, let the extension of $\pi_n(\cdot \mid f_0, p_0, f_1, \cdots, f_n)$ to $P(F)$ in the above sense be $\tilde{\pi}_n(\cdot \mid f_0, p_0, f_1,$

$\cdots, f_n)$. Let $\tilde{\pi}$ denote the sequence $\{\tilde{\pi}_0, \tilde{\pi}_1, \tilde{\pi}_2, \cdots\}$. We call $\tilde{\pi}$ the extension of π to $P(F)$.

Let $\tilde{\mathfrak{F}}_n$ denote the σ-field generated by Borel cylindrical sets on $\tilde{H}_n$. A mapping t from $\tilde{H}$ into $N\cup\}+\infty\}$ is called a *stopping time with respect to* $\{\tilde{\mathfrak{F}}_n\}$, if $\{\tilde{h}\in\tilde{H}; t(\tilde{h})=n\}\in\tilde{\mathfrak{F}}_n$ for $n=0, 1, 2, \cdots$. A stopping time t(w. r. t.$\{\tilde{\mathfrak{F}}_n\}$) is called *associated with* π, if $P^{\tilde{\pi}}_{f_0}\{t<+\infty\}=1$ for all $f_0\in F$, where $P^{\tilde{\pi}}_{f_0}$ is the probability measure on $\tilde{H}$ induced from the extension $\tilde{\pi}$ of π to $P(F)$ given f_0. Let $\tilde{C}(\pi)$ denote the set of all stopping times (w.r.t. $\{\tilde{\mathfrak{F}}_n\}$) associated with π. A pair (π, t) is called a *stopped policy over* $\tilde{H}$ *available in* $\tilde{\Gamma}$, if π is a policy over $\tilde{H}$ available in $\tilde{\Gamma}$ and $t\in\tilde{C}(\pi)$. Denote the set of all stopped policies over $\tilde{H}$ available in $\tilde{\Gamma}$ by $\tilde{\Lambda}^{\tilde{\Gamma}}$. Let us recall the fortune $f_n=(z_n\,; g(z_0, p_0, z_1, p_1, \cdots, z_n\,; 0))$. Define the utility of f_n by the output of it, i. e., $u(f_n)=g(z_0, p_0, z_1, p_1, \cdots, z_n\,; d(z_n))$. Thus the utility has turned out to be given as a function from fortunes to real numbers like in other literatures. Let $E^{\tilde{\pi}}$ be the expectation operator under the extension $\tilde{\pi}$ of π to $P(F)$. Our optimization problem, then, is to maximize $E^{\tilde{\pi}}[u(f_t)]$ in $\tilde{\Lambda}^{\tilde{\Gamma}}$ (Note that $E^{\tilde{\pi}}[u(f_t)]$ is a function of the initial fortune).

Throughout this paper the following conditions remain valid.

CONDITION (B).

(B1) $\sup\limits_{\pi:\text{ available in }\tilde{\Gamma}} E^{\pi}[\sup\limits_{n}(g(z_0, p_0, z_1, \cdots, z_n\,; d(z_n)))^{+}]<+\infty$,

where $a^{+}=\max(a, 0)$ and E^{π} denotes the expectation operator under π.

(B2) For each n and each bounded $w\in B(Z)$, $g(z_0, p_0, z_1, \cdots, z_n\,; w(z_n))$ is bounded on $\Gamma\times\Gamma\times\cdots\times\Gamma\times Z$.

Let $H=Z\times P(Z)\times Z\times P(Z)\times\cdots$ be the set of all sequences $h=(z_0, p_0, z_1, p_1, \cdots)$, and $H_n=Z\times P(Z)\times Z\times P(Z)\times\cdots\times Z(2n+1$ factors) the set of all finite sequences $h_n=(z_0, p_0, z_1, p_1, \cdots, z_n)$ of length $2n+1$. H_n is considered as the set of finite sequences obtained by deleting the memories of the third component of fortune from $\tilde{h}_n$, i. e., H_n is the space of the time, state and gamble which we have had. A *policy* π *over* H is a sequence $\{\pi_0, \pi_1, \pi_2, \cdots\}$, where each $\pi_n\in Q(P(Z)|H_n)$. Denote the section of $\tilde{\Gamma}$ at $c=0$ by Γ. A policy π over H is called *available in* Γ, if each $\pi_n\in Q(\{\Gamma(z_n)\}\,|H_n)$. Let Π^{Γ} denote the set of all policies over H which are available in Γ. For any $\pi\in\Pi^{\Gamma}$, let ${}^{n}\pi=\{\pi_{n+1}, \pi_{n+2}, \cdots\}$ denote the policy which π defines from the $(n+1)$-th stage onward.

Let $\mathfrak{F}_n$ denote the σ-field generated by Borel cylindrical sets on H_n. A mapping t from H into $N\cup\{+\infty\}$ is called a *stopping time with respect to* $\{\mathfrak{F}_n\}$, if $\{h\in H; t(h)=n\}\in\mathfrak{F}_n$ for $n=0, 1, 2, \cdots$. For any policy π over H, a stopping time (w. r. t. $\{\mathfrak{F}_n\}$) is called *associated with* π, if $P^{\pi}_{z_0}\{t<+\infty\}=1$ for all $z_0\in Z$, where $P^{\pi}_{z_0}$ is the probability measure on H induced from π given z_0. For any policy π over H, let $C(\pi)$ denote the set of all stopping times (w. r. t. $\{\mathfrak{F}_n\}$) associated with π. Let $\Lambda^{\Gamma}=\{(\pi, t)|\pi\in\Pi^{\Gamma}, t\in C(\pi)\}$. A pair $(\pi, t)\in\Lambda^{\Gamma}$ is called a *stopped policy over* H *available in* Γ.

A policy π over H is called *Markov*, if for each n, there is a Borel measurable function φ_n from Z into $P(Z)$ such that $\pi_n(\cdot\,|z_0, p_0, z_1, \cdots, z_n)=\delta(\varphi_n(z_n))$, and a Markov policy is denoted by $\{\varphi_0, \varphi_1, \varphi_2, \cdots\}$. Thus, if a Markov policy $\{\varphi_0, \varphi_1, \varphi_2, \cdots\}$ over H is available in Γ, then for each n, $\varphi_n(z)\in\Gamma(z)$ for all $z\in Z$. Let $\mathcal{H}_n$ denote the σ-

field generated by the family of all sets $\{h\in H;\ z_0\in E_0,\ z_1\in E_1, \cdots, z_n\in E_n\}$ with $E_i\in\mathcal{B}_z$, $i=0, 1, \cdots, n$. Let $\mathcal{G}_n$ denote the family of all sets $\{h\in H;\ Z_n\in E_n\}$ with $E_n\in\mathcal{B}_z$. A stopping time t(w. r. t. $\{\mathcal{F}_n\}$) is called *Markov*, if for each $n\geqq 0$, $\{t=n\}\in\mathcal{H}_n$, and for each $n\geqq 1$, $\{t=n\}=\{t>n-1\}\cap\Delta_n$ for some $\Delta_n\in\mathcal{G}_n$. A stopped policy (π, t) over H is called *Markov*, if π and t are both Markov. A stopped policy (π, t) is called *bounded*, if there exists a nonnegative integer N such that $P_{z_0}^{\pi}\{t\leqq N\}=1$ for all $z_0\in Z$.

For any $\pi\in\Pi^{\Gamma}$, let e_{π} denote the element of $Q(P(Z)\times Z\times P(Z)\times Z\times\cdots|Z)$ induced from π, and E^{π} the expectation operator with respect to e_{π}. For any $p\in P(Z)$ and $\varepsilon>0$, $(\pi^*, t^*)\in\Lambda^{\Gamma}$ is called *(p, ε)-optimal*, if $p\{E^{\pi^*}[u(f_{t^*})]\geqq E^{\pi}[u(f_t)]-\varepsilon\}=1$ for all $(\pi, t)\in\Lambda^{\Gamma}$, and *$(\bar{p}, \varepsilon)$-optimal* if $p(E^{\pi^*}[u(f_{t^*})])\geqq p(E^{\pi}[u(f_t)])-\varepsilon$ for all $(\pi, t)\in\Lambda^{\Gamma}$.

3. Stopped policies over H are enough.

In this section we shall show that a stopped policy over $\tilde{H}$ can be replaced by a stopped policy over H which is equivalent to it.

Let us recall Γ is the section of $\tilde{\Gamma}$ at $c=0$. Then the following proposition is direct from Condition (A).

PROPOSITION 3.1.

(i) *$\Gamma\subset Z\times P(Z)$,*

(ii) *$\Gamma(n)\subset S\times P(\{n+1\}\times S)$ for all $n\in\boldsymbol{N}$,*

(iii) *Γ is measurable, i. e.,*

(a) *$\Gamma\in\mathcal{B}_Z\times\Sigma$,*

(b) *there exists a Borel Selector of Γ from Z into $P(Z)$.*

THEOREM 3.1 *For each $(\pi, t)\in\tilde{\Lambda}^{\tilde{\Gamma}}$, there exists $(\theta, \tau)\in\Lambda^{\Gamma}$ such that*

$$E^{\tilde{\pi}}[u(f_t)]=E^{\theta}[u(f_{\tau})], \tag{3.1}$$

where $\tilde{\pi}$ is the extension of π to $P(F)$.

PROOF. Let $\hat{h}$ denote the infinite sequence $(z_0, g(z_0; 0), p_0, z_1, g(z_0, p_0, z_1; 0), p_1, \cdots, z_n, g(z_0, p_0, z_1, \cdots, z_n; 0), p_n, \cdots)$, where each $z_n\in Z$ and each $p_n\in P(Z)$. Evidently $\hat{h}\in\tilde{H}$. Let $\hat{\mathfrak{F}}_n$ denote the σ-field generated by the family of sets $\{\hat{h};\ z_0\in E_0,\ p_0\in P_0,\ z_1\in E_1,\ p_1\in P_1, \cdots, z_n\in E_n\}$ with $E_i\in\mathcal{B}_z$ and with $p_i\in\Sigma$, $i=0, 1, \cdots, n$.

First we shall show that for each set $\tilde{A}_n\in\tilde{\mathfrak{F}}_n$ we can choose a set $\hat{A}_n\in\tilde{\mathfrak{F}}_n$ such that $\hat{A}_n\subset\tilde{A}_n$ and $P_{z_0}^{\tilde{\pi}}(\tilde{A}_n)=P_{z_0}^{\tilde{\pi}}(\hat{A}_n)$ for all $z_0\in Z$. It sufficies to show that we can choose such set for each element generating $\tilde{\mathfrak{F}}_n$. Let $\tilde{A}_n$ be a set $\{\tilde{h}\in\tilde{H};\ z_0\in E_0,\ c_0\in F_0,\ p_0\in P_0,\ z_1\in E_1,\ c_1\in F_1,\ p_1\in P_1, \cdots, z_n\in E_n,\ c_n\in F_n\}$ where $E_i\in\mathcal{B}_z$, $F_i\in\mathcal{B}^1$ ($i=0, 1, \cdots, n$), $p_i\in\Sigma$ ($i=0, 1, \cdots, n-1$). Let K be the set $\{\tilde{h}\in\tilde{H};\ c_i\neq g(z_0, p_0, z_1, p_1, \cdots, z_i; 0)$ for some i $(0\leqq i\leqq n)$, then K is a Borel set since the graph of g is Borel. Evidently $P_{z_0}^{\tilde{\pi}}(K)=0$ for all $z_0\in Z$ from the definition of $\tilde{\pi}$. Since g is Borel measurable, we can choose a Borel subset $\hat{E}$ of H_n so that $\tilde{A}_n\cap K^c=\{\tilde{h};\ (z_0, p_0, z_1, p_1, \cdots, z_n)\in\hat{E}$ and $c_0=g(z_0; 0),\ c_1=g(z_0, p_0, z_1; 0), \cdots, c_n=g(z_0, p_0, z_1, \cdots, z_n; 0)\}$. Let $\hat{A}_n=\tilde{A}_n\cap K^c$, then $\hat{A}_n$ works.

If we let I_A be the indicator function of $\tilde{h}$, the expectation of the utility can be expressed as follows:

$$E^{\bar{\pi}}[u(f_t)]=\sum_{n=0}^{\infty}E^{\bar{\pi}}[u(f_n)I_{\{t=n\}}]. \tag{3.2}$$

By the above assertion, for each n choose a set $\hat{A}_n\in\hat{\mathfrak{F}}_n$ such that $\hat{A}_n\subset\{t=n\}$ and $P_{z_0}^{\bar{\pi}}(\{t=n\})=P_{z_0}^{\bar{\pi}}(\hat{A}_n)$ for all $z_0\in Z$. Then it follows from (3.2) that

$$E^{\bar{\pi}}[u(f_t)]=\sum_{n=0}^{\infty}E^{\bar{\pi}}[u(f_n)I_{\hat{A}_n}].$$

Define a stopping time $\hat{t}$ by $\{\hat{t}=n\}\equiv\hat{A}_n$ for $n=0, 1, 2, \cdots$, and define a stopping time τ by $\tau(h)=\hat{t}(\hat{h})$, i.e., $\tau(z_0, p_0, z_1, p_1, \cdots, z_n, p_n, \cdots)=\hat{t}(z_0, g(z_0\,;0), p_0, z_1, g(z_0, p_0, z_1\,;0)), \cdots, z_n, g(z_0, p_0, z_1, p_1, \cdots, z_n\,;0), p_n, \cdots)$. Then it is easy to show that

$$\{h\in H;\ \tau(h)=n\}\in\mathfrak{F}_n, \qquad n=0, 1, 2, \cdots,$$

$$P_{z_0}^{\pi}\{\tau<+\infty\}=1 \qquad \text{for all } z_0\in Z,$$

and

$$E^{\bar{\pi}}[u(f_t)]=E^{\pi}[u(f_\tau)]. \tag{3.3}$$

We next define θ_n as follows:

$$\theta_n(\cdot\,|z_0, p_0, z_1, p_1, \cdots, z_n)=\pi_n(\cdot\,|z_0, g(z_0\,;0), p_0, z_1, g(z_0, p_0, z_1\,;0), p_1,$$
$$\cdots, z_n, g(z_0, p_0, z_1, p_1, \cdots, z_n\,;0)).$$

Then for each $B\in\Sigma$, $\theta_n(B|\cdot)$ is Borel measurable, because g is Borel measurable. Since π is available in $\tilde{\Gamma}$, it follows that $\theta_n\in Q(\{\Gamma(z_n)\}|H_n)$ from Condition (A1). Putting $\theta=(\theta_0, \theta_1, \theta_2, \cdots)$, θ becomes a policy over H available in Γ which satisfies

$$E^{\pi}[u(f_t)]=E^{\theta}[u(f_t)]. \tag{3.4}$$

Combining (3.3) and (3.4) leads to (3.1). Evidently $(\theta, \tau)\in\Lambda^{\Gamma}$. This completes the proof.

By virtue of the above theorem, we have only to consider the stopped policies over H as far as we are concerned with the optimization problem defined in Section 2. Hence we shall restrict ourselves to the family of the stopped policies over H in the subsequent sections, and we shall omit the term "over H".

4. (p, ε)-optimal and $(\bar{p}, \varepsilon)$-optimal stopped policies.

In this section we shall show that under some conditions there exist a (p, ε)-optimal stopped policy and a $(\bar{p}, \varepsilon)$-optimal stopped policy both of which are bounded. We shall also give some results in the case of finite stage.

Let N be an arbitrary fixed positive integer, and let $X^N=P(Z)\times Z\times P(Z)\times Z\times\cdots\times Z$ ($2N$ factors). A generic element of X^N will be denoted by $(p_0, z_1, p_1, z_2, \cdots, p_{N-1}, z_N)$. Let Σ^* be the $^*\sigma$-algebra on $P(X^N)$.

Let Π_N^{Γ} denote the set of all initial subsequences $\{\pi_0, \pi_1, \cdots, \pi_{N-1}\}$ of length N of policies available in Γ, and W_N^{Γ} the set of all $(z, \nu)\in Z\times P(X^N)$ for which there exists a $\pi\in\Pi_N^\Gamma$ such that $\nu=e_\pi(z)$, where $e_\pi(z)$ means the probability measure on X^N induced from π given z.

LEMMA 4.1. $W_N^\Gamma\in\mathscr{B}_z\times\Sigma^*$.

PROOF. The lemma will be proved along the same line as Lemma 1 of [15] by making use of a Borel Selector of Γ.

For any $\nu\in P(X^N)$, we have a factorization $\nu=\nu_0\nu_1\cdots\nu_{2N-1}$ such that

$$\nu_0\in P(P(Z))\,,$$

$$\nu_{2n}\in Q(P(Z)\,|\,P(Z)\times Z\times P(Z)\times\cdots\times Z)\qquad (2n \text{ factors}),$$

$$\nu_{2n+1}\in Q(Z\,|\,P(Z)\times Z\times P(Z)\times\cdots\times P(Z))\qquad (2n+1 \text{ factors}).$$

First we shall prove that $(z,\nu)\in W_N^\Gamma$ if and only if (i) $\nu_0(\Gamma(z))=1$ and $\nu_1(\cdot\,|\,p_0)=p_0$ a.s. (ν), and (ii) $\nu_{2n}(\Gamma(z_n)\,|\,p_0, z_1, \cdots, z_n)=1$ for all $(p_0, z_1, \cdots, z_n)$ and $\nu_{2n+1}(\cdot\,|\,p_0, z_1, p_1, \cdots, p_n)=p_n$ a.s. (ν) $(n\geqq 1)$. It sufficies to prove the "if" part, since the "only if" part is trivial. Let (z_0,ν) satisfy both of (i) and (ii). Define a mapping γ^* from Z into $P(P(Z))$ by

$$\gamma^*(z)=\begin{cases}\nu_0 & \text{if } z=z_0\\ \alpha(z) & \text{if } z\neq z_0\,,\end{cases}$$

where α is a Borel Selector of Γ assured in Proposition 3.1. Then $\gamma^*\in Q(\{\Gamma(z)\}\,|\,Z)$, since ν_0 satisfies (i) and Z is a T_1-space. Let $\pi^*=\{\gamma^*, \nu_2, \nu_4, \cdots, \nu_{2(N-1)}\}$, then it follows that $\pi^*\in\Pi_N^\Gamma$ from (ii) and the above result on γ^*. Evidently $\nu=e_{\pi^*}(z_0)$. Hence $(z_0,\nu)\in W_N^\Gamma$.

Denote the probability measures on $(p_0, z_1, p_1, \cdots, z_n)$ and on $(p_0, z_1, p_1, \cdots, p_n)$ induced from ν by $\nu(p_0, z_1, p_1, \cdots, z_n)$ and $\nu(p_0, z_1, p_1, \cdots, p_n)$, respectively. The condition (i) is equivalent to that for all bounded $\phi\in B(X^1)$

$$\int_{X^1}\phi(p_0, z_1)d\nu(p_0, z_1)=\int_{\Gamma(z)}\left[\int_Z\phi(p_0, z_1)dp_0(z_1)\right]d\nu(p_0)\,, \tag{4.2}$$

and the condition (ii) equivalent to that for all bounded $\phi\in B(X^{n+1})$

$$\begin{aligned}(4.2)\qquad &\int_{X^{n+1}}\phi(p_0, z_1, p_1, \cdots, p_n, z_{n+1})d\nu(p_0, z_1, p_1, \cdots, p_n, z_{n+1})\\ &=\int_{[(z_n, p_n)\in\Gamma]}\left[\int_Z\phi(p_0, z_1, p_1, \cdots, p_n, z_{n+1})dp_n(z_{n+1})\right]\\ &\qquad d\nu(p_0, z_1, \cdots, z_n, p_n)\,,\qquad (n\geqq 1)\,.\end{aligned}$$

Each integral in (4.1) and (4.2) is a Borel measurable function of z and ν by virtue of Proposition 3.1, (iii)-(a) and Theorem 2.2 of [4]. Since for each $n\geqq 0$, the Borel σ-field on X^{n+1} is countably generated, we can choose a sequence $\{\phi_{nm}\}_{m=1,2\cdots}$ of bounded Borel measurable functions on X^{n+1} such that if $\mu, \nu\in P(X^{n+1})$ and $\mu\neq\nu$, then

$$\int_{X^{n+1}}\phi_{nj}d\mu\neq\int_{X^{n+1}}\phi_{nj}d\nu$$

for some $j\geqq 1$. For each $m\geqq 1$, let W_{0m} be the set of all (z, ν) such that (4.1) holds for $\phi=\phi_{0m}$, and for each $n\geqq 1$, $m\geqq 1$, W_{nm} the set of all (z, ν) such that (4.2) holds for $\phi=\phi_{nm}$. Then we have

$$W_N^\Gamma=\bigcap_{n=0}^{N-1}\bigcap_{m=1}^{\infty}W_{nm}\,.$$

Hence W_N^Γ is Borel, as each W_{nm} is Borel. This completes the proof.

LEMMA 4.2. *The z-section $W_N^\Gamma(z)$ of W_N^Γ is not empty for all $z\in Z$.*

PROOF. Take a Borel Selector α of Γ, and let $\hat{\pi}=\{\alpha, \alpha, \cdots, \alpha\}$. Evidently $\hat{\pi}\in \Pi_N^\Gamma$ and $e_{\hat{\pi}}(z)\in P(X^N)$ for all $z\in Z$, which completes the proof.

We shall now introduce into the generalized reward function g two important properties, namely, a recursiveness and a monotonicity, whicn can be jointly considered as a characterization of the dynamic programming structure. These properties play an important role leading us to "the principle of optimality" of R. Bellman, as it is seen in the paper [10].

CONDITION (C). g is *recursive*, i.e., for any policy π available in Γ, any i, m and n such that $0\leqq i<m\leqq n$, and any bounded $w\in B(H_{n+1})$, it holds that

$$\begin{aligned}&E^{m-1\pi}[g(z_i, p_i, z_{i+1}, \cdots, z_{n+1}; w(h_{n+1}))]\\ &\quad=g(z_i, p_i, z_{i+1}, \cdots, z_m; E^{m-1\pi}[g(z_m, p_m, z_{m+1}, \cdots, z_{n+1}; w(h_{n+1}))])\\ &\qquad\qquad\qquad\qquad\qquad\text{a.s.}\quad (P^\pi).\end{aligned}$$

CONDITION (D). g is *monotone*, i.e., for any $n\geqq 0$ and any $u, v\in B(H_{n+1})$, it holds that $u(h_{n+1})<v(h_{n+1})$ implies $g(z_n, p_n, z_{n+1}; u(h_{n+1}))\leqq g(z_n, p_n, z_{n+1}; v(h_{n+1}))$, where in the second statement (z_n, p_n, z_{n+1}) is the final subsequence of h_{n+1}.

For each $N\geqq 0$ and any $\pi\in\Pi_N^\Gamma$, let us define a finite sequence of random variables $\{v_n^N(\pi); 0\leqq n\leqq N\}$ by backward induction:

$$(4.3)\qquad \begin{cases} v_n^N(\pi)(h_n)=\max[d(z_n), \pi_n p_n g(z_n, p_n, z_{n+1}; v_{n+1}^N(\pi)(h_{n+1}))] \\ \qquad\qquad\qquad\qquad\qquad\qquad n=0, 1, \cdots, N-1, \\ v_N^N(\pi)(z_N)=d(z_N)\,. \end{cases}$$

Note that for each n, $v_n^N(\pi)$ depends only on $\{\pi_n, \pi_{n+1}, \cdots, \pi_{N-1}\}$ among all components of π and for each π, $v_n^N(\pi)$ is a Borel measurable function of h_n.

Now let

$$v^{N*}(z)=\sup_{\pi\in\Pi_N^\Gamma} v_0^N(\pi)(z)\,,\qquad z\in Z\,,$$

for $N=0, 1, 2, \cdots$. Under Condition (B1), it turns out that $v^{N*}<+\infty$, which we shall prove in Proposition 4.1.

The following lemma is a slight modification of Theorem 2.2 of [4].

LEMMA 4.3. *Let w be a bounded Borel measurable function defined on $\Gamma\times Z$. Then pw is Borel measurable on Γ.*

For any $\nu\in P(X^N)$, let $\nu=\nu_0\nu_1\cdots\nu_{2N-1}$ be the factorization as stated immediately after Lemma 4.1, and define a finite sequence of random variables $\{v_n^N(\nu); 0\leqq n\leqq N\}$ by backward induction:

$$(4.4)\qquad \begin{cases} v_n^N(\nu)(h_n)=\max\left[d(z_n),\ \nu_{2n}\nu_{2n+1}g(z_n,\ \nu_{2n+1},\ z_{n+1}\,;\ v_{n+1}^N(\nu)(h_{n+1}))\right] \\ \qquad\qquad\qquad\qquad\qquad\qquad n=0,\ 1,\ \cdots,\ N-1\,, \\ v_N^N(\nu)(z_N)=d(z_N)\,. \end{cases}$$

LEMMA 4.4. *Let Condition* (C) *be satisfied. Then for each N, v^{N*} is absolutely measurable in z.*

PROOF. Let $\bar{v}(z,\nu)=v_0^N(\nu)(z)$, and let

$$B_\lambda=W_N^\Gamma\cap\{(z,\nu)\in Z\times P(X^N)\,|\,\bar{v}(z,\nu)>\lambda\}\,.$$

Then for each real λ, B_λ is Borel from Condition (C), Lemma 4.1 and Lemma 4.3. By Lemma 4.2 we have

$$v^{N*}(z)=\sup_{\nu\in W_N^\Gamma(z)}\bar{v}(z,\nu)\,,\qquad z\in Z\,.$$

Let

$$C_\lambda=\{z\in Z\,|\,v^{N*}(z)>\lambda\}\,,$$

then it is easy to check that C_λ is equal to the projection of B_λ into Z. Since B_λ is Borel, C_λ is analytic and so absolutely measurable.

THEOREM 4.1. *Let Condition* (C) *be satisfied. Then for each N, any $p\in P(Z)$ and any $\varepsilon>0$, there exists a policy $\hat{\pi}\in\Pi_N^\Gamma$ such that*

$$p\,\{v_0^N(\hat{\pi})>v^{N*}-\varepsilon\}=1\,.$$

PROOF. The proof follows the same line as in Theorem 8.1 of [14] by making use of a Borel Selector of Γ. From Lemma 4.4, we can choose a Borel set $K_1\subset Z$ and a Borel measurable function v_0 on Z such that $p(K_1)=0$ and $v_0(z)=v^{N*}(z)$ for $z\notin K_1$. Let

$$W^\varepsilon=W_N^\Gamma\cap[\{(z,\nu)\,|\,z\notin K_1,\ \bar{v}(z,\nu)>v_0(z)-\varepsilon\}\cup\{(z,\nu)\,|\,z\in K_1\}]\,,$$

then W^ε is Borel from Condition (C), Lemma 4.1 and Lemma 4.3. For each z, the z-section $W^\varepsilon(z)$ of W^ε is not empty because of Lemma 4.2. The $*\sigma$-algebra on $P(Z)$ coincides with the σ-field generated by the weak-topology on $P(Z)$. Hence $P(Z)$ is a Polish space (cf. [13]). Similarly $P(X^N)$ is so. Then applying the absolutely measurable selection theorem by Mackey [12] (originally the analytic selection theorem by Von Neumann [17]), we can find a Borel set $K_2\subset Z$ and a Borel measurable mapping φ from Z into $P(X^N)$ such that $p(K_2)=0$ and $\varphi(z)\in W^\varepsilon(z)$ for $z\notin K_2$. Hence if $z\notin K_2$, then there exists a policy $\hat{\pi}\in\Pi_N^\Gamma$ such that $e_{\hat{\pi}}(z)=\varphi(z)$. Let us define μ by

$$\mu(B\,|\,z)=\varphi(z)(B)\qquad\text{for } z\in Z,\ B\in\mathcal{B}_{X^N}\,,$$

where $\mathcal{B}_{X^N}$ is the Borel σ-field on X^N. Then it follows that $\mu\in Q(X^N\,|\,Z)$ from the definition of φ. Hence we can factor μ:

$$\mu=\mu_0\mu_1\cdots\mu_{2N-1}\,,$$

where

$$\mu_0\in Q(P(Z)\,|\,Z)\,,$$

$$\mu_{2n}\in Q(P(Z)\,|\,Z\times P(Z)\times\cdots\times Z)\qquad(2n+1\text{ factors})$$

$$\mu_{2n+1}\in Q(Z\,|\,Z\times P(Z)\times\cdots\times P(Z))\qquad(2n+2\text{ factors}).$$

Taking a Borel Selector α of Γ, define a policy π^*, which depends on the initial z_0, by

$$\pi^*(z_0)=\begin{cases}\{\mu_0,\ \mu_2,\ \cdots,\ \mu_{2(N-1)}\} & \text{if } z_0\notin K_2\\ \{\alpha,\ \alpha,\ \cdots,\ \alpha\} & \text{if } z_0\in K_2\,.\end{cases}$$

Then it is easy to show that $\pi^*\in\Pi^\Gamma_N$. If $z_0\notin K_1\cup K_2$, then

$$v^N_0(\pi^*)(z_0)=\bar{v}(z_0,\ \varphi(z_0))>v_0(z_0)-\varepsilon=v^{N*}(z_0)-\varepsilon\,.$$

But $p(K_1\cup K_2)=0$. Thus we have $p\{v^N_0(\pi^*)>v^{N*}-\varepsilon\}=1$, which completes the proof.

For each $N\geqq 0$ and any $\pi\in\Pi^\Gamma_N$, define a finite sequence of random variables $\{\beta^N_n(\pi)\,;\ 0\leqq n\leqq N\}$ by

$$\beta^N_n(\pi)(h_n)=g(h_n\,;\ v^N_n(\pi)(h_n))\,,\qquad n=0,\ 1,\ \cdots,\ N\,, \tag{4.5}$$

where $\{v^N_n(\pi)\}$ is defined in (4.3).

LEMMA 4.5. *For each* $N\geqq 0$, $\beta^N_0(\pi)=v^N_0(\pi)$.

PROOF. From (4.5), $\beta^N_0(\pi)(z_0)=g(z_0\,;\ v^N_0(\pi)(z_0))$. From the assumption on g stated in Section 2, $g(z_0\,;\ v^N_0(\pi)(z_0))=v^N_0(\pi)(z_0)$. Hence the lemma follows.

LEMMA 4.6. *Let Conditions* (C) *and* (D) *be satisfied. Then for each* $N\geqq 0$ *and any* $\pi\in\Pi^\Gamma_N$, $\{\beta^N_n(\pi)\}$ *satisfies the recursive relations*:

$$\begin{cases}\beta^N_n(\pi)(h_n)=\max\left[g(h_n\,;\ d(z_n)),\ \pi_n p_n\beta^N_{n+1}(\pi)(h_n)\right]\,, & n=0,\ 1,\ \cdots,\ N-1\,,\\ \beta^N_N(\pi)(h_N)=g(h_N\,;\ d(z_N))\,. & \end{cases} \tag{4.6}$$

PROOF. From Condition (D), it follows that $\max(g(h_n\,;\ a),\ g(h_n\,;\ b))\leqq g(h_n\,;\ \max(a,\ b))$ for any $h_n\in H_n$ and any real numbers $a,\ b$. But the converse inequality is trivial. Hence we get

$$g(h_n\,;\ \max(a,\ b))=\max(g(h_n\,;\ a),\ g(h_n\,;\ b)) \tag{4.7}$$

for any $h_n\in H_n$ and any real numbers $a,\ b$. For each n $(0\leqq n\leqq N-1)$, we get from (4.3), (4.7) and Condition (C) that

$$\begin{aligned} g(h_n\,;\ v^N_n(\pi))&=g(h_n\,;\ \max\left[d(z_n),\ \pi_n p_n g(z_n,\ p_n,\ z_{n+1}\,;\ v^N_{n+1}(\pi))\right])\\ &=\max\left[g(h_n\,;\ d(z_n)),\ g(h_n\,;\ \pi_n p_n g(z_n,\ p_n,\ z_{n+1}\,;\ v^N_{n+1}(\pi)))\right]\\ &=\max\left[g(h_n\,;\ d(z_n),\ \pi_n p_n g(h_{n+1}\,;\ v^N_{n+1}(\pi))\right]\,.\end{aligned}$$

Substituting (4.5) into (4.8) leads to

$$\beta^N_n(\pi)=\max\left[g(h_n\,;\ d(z_n)),\ \pi_n p_n\beta^N_{n+1}(\pi)\right]\,.$$

For $n=N$, we get from (4.3) and (4.5) that

$$\beta^N_N(\pi)=g(h_N\,;\ v^N_N(\pi))=g(h_N\,;\ d(z_N))\,.$$

This completes the proof.

Let $C^N(\pi)$ denote the set of all stopping times $t\in C(\pi)$ for which $P^\pi_{z_0}$ $(\{t\leqq N\})=1$ for all $z_0\in Z$.

DEFINITION 4.1. For any $\pi\in\Pi^\Gamma$, a stopping time $t\in C^N(\pi)$ is called *π-regular*

(w. r. t. $\{\beta_n^N(\pi)\}$), if for all $n=0, 1, \cdots, N$ it holds that

$$(4.9) \qquad E^{n-1\pi}[\beta_t^N(\pi)]\geqq\beta_n^N(\pi) \quad \text{a.s. } (e_\pi) \text{ on } \{t>n\}.$$

The above definition is a modification of the "regularity" given by Chow and Robbins ([3]).

Now, for each N and any $\pi\in\Pi_N^\Gamma$, we define

$$(4.10) \qquad \tau_N(\pi)=\text{the first } n\geqq 0 \text{ such that } v_n^N(\pi)=d(z_n).$$

Note that $(\pi, \tau_N(\pi))$ is bounded by N, since $v_N^N(\pi)=d(z_N)$.

LEMMA 4.7. *Let Condition* (C) *be satisfied. For each N and any $\pi\in\Pi_N^\Gamma$, then, $\tau_N(\pi)$ is π-regular w. r. t. $\{\beta_n^N(\pi)\}$.*

PROOF. From (4.3), (4.5) and Condition (C) it follows that

$$\begin{aligned}
\tau_N(\pi)>n &\Rightarrow v_n^N(\pi)>d(z_n)\\
&\Rightarrow v_n^N(\pi)=\pi_n p_n g(z_n, p_n, z_{n+1}\,;\, v_{n+1}^N(\pi))\\
&\Rightarrow g(h_n\,;\, v_n^N(\pi))=g(h_n\,;\, \pi_n p_n g(z_n, p_n, z_{n+1}\,;\, v_{n+1}^N(\pi)))\\
&\Rightarrow \beta_n^N(\pi)=\pi_n p_n \beta_{n+1}^N(\pi).
\end{aligned}$$

That is, $\{\beta_n^N(\pi)\}$ and $\tau_N(\pi)$ satisfy (3.4) in [8]. Further it is trivial that they satisfy (3.5) in [8], because $\{\beta_n^N(\pi)\}$ is a finite sequence. Hence it follows from Lemma 3.2 of [8] that $\tau_N(\pi)$ is π-regular w. r. t. $\{\beta_n^N(\pi)\}$, which completes the proof.

In the subsequent arguments if there is no fear of confusion, we shall write $\tau_N(\pi)$ as τ_N.

LEMMA 4.8. *Let Conditions* (C) *and* (D) *be satisfied. For each N and any $\pi\in\Pi_N^\Gamma$, then, we have*

$$\text{(a)} \qquad E^\pi[\beta_{\tau_N}^N(\pi)]\geqq E^\pi[\beta_t^N(\pi)] \quad \textit{for all } t\in C^N(\pi),$$

$$\text{(b)} \qquad E^\pi[u(f_{\tau_N})]=\sup_{t\in C^N(\pi)} E^\pi[u(f_t)].$$

PROOF. (a) From (4.6) we have

$$E^{n-1\pi}[\beta_{n+1}^N(\pi)]\leqq\beta_n^N(\pi), \qquad n=0, 1, \cdots, N.$$

The above inequalities mean that $\{\beta_n^N(\pi)\}$ has the supermartingale properties under π. On the other hand $\{\beta_n^N(\pi)\}$ satisfies obviously (3.8) in [8] for every $t\in C^N(\pi)$. Hence it follows from Lemma 3.3 of [8] that

$$E^{n-1\pi}[\beta_t^N(\pi)]\leqq\beta_n^N(\pi) \quad \text{a.s. } (e_\pi) \text{ on } \{\tau_N=n,\ t\geqq n\}$$

for every $n\geqq 0$ and every $t\in C^N(\pi)$. This implies that $\{\beta_n^N(\pi)\}$ and τ_N satisfy (3.2) in [8] for every $t\in C^N(\pi)$. But Lemma 4.7 implies that (3.1) of [8] holds for $\{\beta_n^N(\pi)\}$ and τ_N. Hence from Lemma 3.1 of [8] we get (a).

(b) By the relation that

$$\begin{aligned}
\tau_N=n \Rightarrow v_n^N(\pi)=d(z_n) &\Rightarrow g(h_n\,;\, v_n^N(\pi))=g(h_n\,;\, d(z_n))\\
&\Rightarrow \beta_n^N(\pi)=g(h_n\,;\, d(z_n)),
\end{aligned}$$

we get

$$E^{\pi}[\beta^N_{\tau_N}(\pi)]=\sum_{n=0}^{N}E^{\pi}[\beta^N_n(\pi)I_{(\tau_N=n)}]$$

$$(4.11)\qquad =\sum_{n=0}^{N}E^{\pi}[g(h_n\,;\,d(z_n))I_{(\tau_N=n)}]$$

$$=E^{\pi}[g(h_{\tau_N}\,;\,d(z_{\tau_N}))]=E^{\pi}[u(f_{\tau_N})]\,.$$

(a) together with (4.11) implies that

$$(4.12)\qquad E^{\pi}[u(f_{\tau_N})]\geqq E^{\pi}[\beta^N_t(\pi)]\qquad \text{for all } t\in C^N(\pi)\,.$$

From (4.6) obviously

$$(4.13)\qquad E^{\pi}[\beta^N_t(\pi)]\geqq E^{\pi}[g(h_t\,;\,d(z_t))]$$

$$=E^{\pi}[u(f_t)]\qquad \text{for all } t\in C^N(\pi)\,.$$

Combining (4.12) and (4.13) leads to

$$E^{\pi}[u(f_{\tau_N})]\geqq E^{\pi}[u(f_t)]\qquad \text{for all } t\in C^N(\pi)\,.$$

Since $\tau_N\in C^N(\pi)$, however, it follows that

$$E^{\pi}[u(f_{\tau_N})]=\sup_{t\in C^N(\pi)}E^{\pi}[u(f_t)]\,.$$

Lemma 4.9. *Let Conditions* (C), (D) *be satisfied. For any $\pi\in\Pi^{\Gamma}$ then we have*

$$E^{\pi}[u(f_{\tau_N})]=E^{\pi}[\beta^N_{\tau_N}(\pi)]=v^N_0(\pi)\,.$$

Proof. The first equality has been shown in the proof of Lemma 4.8, (b).

By the same way as Lemma 5.4 we get

$$E^{\pi}[\beta^N_{\tau_N}(\pi)]\leqq\beta^N_0(\pi)\,.$$

But, the converse inequality follows from substituting $t\equiv 0$ in (a) of Lemma 4.8. Hence we obtain the second equality in the lemma.

We now return to the following proposition whose proof has been postponed.

Proposition 4.1. *Let Conditions* (C) *and* (D) *be satisfied. Then $v^{N*}<+\infty$ for each N.*

Proof. By Lemma 4.8, Lemma 4.9 and Condition (B1) we get

$$v^{N*}=\sup_{\pi\in\Pi^{\Gamma}_N}v^N_0(\pi)=\sup_{\pi\in\Pi^{\Gamma}_N}\ \sup_{t\in C^N(\pi)}E^{\pi}[u(f_t)]$$

$$\leqq\sup_{\pi\in\Pi^{\Gamma}_N}E^{\pi}[\sup_{0\leq n\leq N}(u(f_n))^+]<+\infty\,.$$

On the basis of the several results stated above, we can show the existence of a $(p,\,\varepsilon)$-optimal bounded policy in the case of finite horizon.

Theorem 4.2. *Let Conditions* (C) *and* (D) *be satisfied. For each N, any $p\in P(Z)$ and any $\varepsilon>0$, then there exists a bounded policy $(\hat{\pi},\,\hat{t})\in\Lambda^{\Gamma}_N$ such that*

$$(4.14)\qquad p\{E^{\hat{\pi}}[u(f_{\hat{t}})]>E^{\pi}[u(f_t)]-\varepsilon\}=1\qquad \textit{for all } (\pi,\,t)\in\Lambda^{\Gamma}_N\,,$$

where $\Lambda_N^\Gamma=\{(\pi,\ t)\in\Lambda^\Gamma\mid P_{z_0}^\pi(\{t\leqq N\})=1$ *for all* $z_0\in Z\}$.

PROOF. By virtue of Theorem 4.1, for any $p\in P(Z)$ and $\varepsilon>0$ there exists a policy $\hat{\pi}\in\Pi_N^\Gamma$ such that

$$p\{v_0^N(\hat{\pi})>v^{N*}-\varepsilon\}=1\,,$$

which implies from the definition of v^{N*} that

$$p\{v_0^N(\hat{\pi})>v_0^N(\pi)-\varepsilon\}=1 \qquad \text{for all } \pi\in\Pi_N^\Gamma\,. \tag{4.15}$$

By Lemma 4.9 we can rewrite (4.15) as follows.

$$p\{E^{\hat{\pi}}[u(f_{\tau_N(\hat{\pi})})]>E^\pi[u(f_{\tau_N(\pi)})]-\varepsilon\}=1 \qquad \text{for all } \pi\in\Pi_N^\Gamma\,. \tag{4.16}$$

In (4.16), $\tau_N(\hat{\pi})$ and $\tau_N(\pi)$ mean the stopping times defined in (4.10), respectively, corresponding to $\hat{\pi}$ and π. By (b) of Lemma 4.8 and (4.16) we get

$$p\{E^{\hat{\pi}}[u(f_{\tau_N(\hat{\pi})})]<E^\pi[u(f_t)]-\varepsilon\}=1 \qquad \text{for all } (\pi,\ t)\in\Lambda_N^\Gamma\,.$$

Evidently $(\hat{\pi},\ \tau_N(\hat{\pi}))\in\Lambda_N^\Gamma$. This completes the proof.

The following corollary is direct from Theorems 4.1 and 4.2.

COROLLARY 4.1. *Let Conditions* (C) *and* (*D*) *be satisfied. Then for each* N *we have*

(a) *if there exists a policy* $\pi^*\in\Pi_N^\Gamma$ *such that*

$$v_0^N(\pi^*)(z_0)=\sup_{\pi\in\Pi_N^\Gamma} v_0^N(\pi)(z_0)\,, \tag{4.17}$$

then it holds that

$$E^{\pi^*}[u(f_{\tau_N(\pi^*)})](z_0)=\sup_{(\pi,t)\in\Lambda_N^\Gamma} E^\pi[u(f_t)](z_0)\,, \tag{4.18}$$

(b) *if there exists a policy* $\pi^*\in\Pi_N^\Gamma$ *which satisfies* (4.17) *for all* $z_0\in Z$, *then* $(\pi^*,\ \tau_N(\pi^*))$ *satisfies* (4.18) *for all* $z_0\in Z$.

LEMMA 4.10. *Let Conditions* (C) *and* (D) *be satisfied Let* p *be any element of* $P(Z)$ *and* ε *any positive number. For each* N *let* $(\hat{\pi}_N,\ \hat{\tau}_N)$ *denote the bounded policy given by* (4.14). *Then without loss of generality we may assume that* $\{E^{\hat{\pi}^N}[u(f_{\hat{\tau}_N})]\}$ *is a nondecreasing sequence with* p*-probability one.*

PROOF. The proof follows the same way as Lemma 5.5 of [8], so that omitted.

We now impose on g the following condition.

CONDITION (E) For any $(\pi,\ t)\in\Lambda^\Gamma$, if $E^\pi[g(z_0,\ p_0,\ z_1,\ \cdots,\ z_t\,;\ d(z_t))]\neq-\infty$, then

$$\liminf_{N\to\infty}\int_{\{t>N\}}[g(z_0,\ p_0,\ z_1,\ \cdots,\ z_N\,;\ d(z_N))]^- de_\pi=0\,.$$

LEMMA 4.11. *Let Conditions* (C), (D) *and* (E) *be satisfied. Let* p *be any element of* $P(Z)$ *and* ε *any positive number. Let* $(\hat{\pi}^N,\ \hat{\tau}_N)$ *be as in Lemma* 4.10. *Then it holds that*

$$p\{\lim_{N\to\infty} E^{\hat{\pi}^N}[u(f_{\hat{\tau}_N})]\geqq E^\pi[u(f_t)]-\varepsilon\}=1 \qquad \textit{for every } (\pi,\ t)\in\Lambda^\Gamma\,. \tag{4.17}$$

PROOF. Suppose that $E^\pi[u(f_t)]\neq-\infty$. Letting $t_N=\min(t,\ N)$, by Theorem 4.2 we get

$$\int_{\{t\leqq N\}} u(f_t)de_\pi = E^\pi[u(f_{t_N})] - \int_{\{t>N\}} u(f_{t_N})de_\pi \tag{4.18}$$

$$\leqq E^{\hat{\pi}^N}[u(f_{\hat{t}_N})] + \varepsilon - \int_{\{t>N\}} [u(f_{t_N})]^- de_\pi \quad \text{with } p\text{-prop. } 1,$$

for each N. From Condition (B1) and Lemma 4.10, $\lim\limits_{N\to\infty} E^{\hat{\pi}^N}[u(f_{\hat{t}_N})]$ exists as a finite value with p-probability one. Hence letting $N\to\infty$ in (4.18), by Condition (E) we obtain (4.17). When $E^\pi[u(f_t)]=-\infty$, (4.17) is trivially true.

THEOREM 4.3. *Let Conditions* (C), (D) *and* (E) *be satisfied. For anp* $p\in P(Z)$ *and any* $\varepsilon>0$ *then there exists a* ($\bar{p}$, ε)*-optimal stopped policy which is bounded.*

PROOF. From Theorem 4.2, for each N there exists $(\hat{\pi}^N, \hat{t}_N)\in \Lambda_N^\Gamma$ such that

$$p\left\{E^{\hat{\pi}^N}[u(f_{\hat{t}_N})] > E[u(f_t)] - \frac{\varepsilon}{2}\right\} = 1 \qquad \text{for all } (\pi, t)\in \Lambda_N^\Gamma .$$

Hence from Lemma 4.11 it follows that

$$p\left\{\lim_{N\to\infty} E^{\hat{\pi}^N}[u(f_{t_N})] \geqq E^\pi[u(f_t)] - \frac{\varepsilon}{2}\right\} = 1 \qquad \text{for all } (\pi, t)\in \Lambda^\Gamma ,$$

which implies that

$$p(\lim_{N\to\infty} E^{\hat{\pi}^N}[u(f_{\hat{t}_N})]) \geqq p(E^\pi[u(f_t)]) - \frac{\varepsilon}{2} \qquad \text{for all } (\pi, t)\in \Lambda^\Gamma .$$

By the theorem of monotone convergence we get

$$\lim_{N\to\infty} p(E^{\hat{\pi}^N}[u(f_{\hat{t}_N})]) \geqq p(E^\pi[u(f_t)]) - \frac{\varepsilon}{2} \qquad \text{for all } (\pi, t)\in \Lambda^\Gamma . \tag{4.19}$$

Since $p(E^{\hat{\pi}^N}[u(f_{\hat{t}_N})])$ is nondecreasing in N, we can choose a sufficiently large N_0 so that

$$\lim_{N\to\infty} p(E^{\hat{\pi}^N}[u(f_{t_N})]) - \frac{\varepsilon}{2} \leqq p(E^{\hat{\pi}^{N_0}}[u(f_{\hat{t}_{N_0}})]) . \tag{4.20}$$

Combining (4.19) and (4.20) leads us to that $(\hat{\pi}^{N_0}, \hat{t}_{N_0})$ is ($\bar{p}$, ε)-optimal bounded policy, which completes the proof.

The following corollary is an easy consequence of the above theorem.

COROLLARY 4.2. *Let Conditions* (C), (D) *and* (E) *be satisfied. Then we have*

(a) *if* $\{E^{\hat{\pi}^N}[u(f_{\hat{t}_N})]\}$ *converges uniformly on* Z, *for any* $p\in P(Z)$ *and any* $\varepsilon>0$ *then there exists a* (p, ε)*-optimal bounded policy,*

(b) *if* S *is a finite set, for any* $p\in P(\{0\}\times S)$ *and any* $\varepsilon>0$ *then there exists a* (p, ε)*-optimal bounded policy,*

(c) *if* $p\in P(Z)$ *has a finite support, for any* $\varepsilon>0$ *then there exists a* (p, ε)*-optimal bounded policy.*

5. Markov bounded policies.

In this section we shall show that for every bounded policy (π, t) there exists a Markov bounded policy which (p, ε)-dominates (π, t). This fact together with the results obtained in Section 4 will lead us to the existence of a (p, ε)-optimal Markov bounded policy and of a $(\bar{p}, \varepsilon)$-optimal Markov bounded policy under some conditions.

First we need to prepare two lemmas.

LEMMA 5.1. *For any $q \in Q(\{\Gamma(z)\} \mid Z)$, any bounded $w \in B(\Gamma)$ and any $\varepsilon > 0$, there is a Borel measurable function φ from Z into $P(Z)$ such that*

(i) graph $\varphi \subset \Gamma$,

(ii) $q(\{p \in \Gamma(z)\,;\ w(z, p) \leqq w(z, \varphi(z)) + \varepsilon\} \mid z) = 1$ *for all* $z \in Z$.

PROOF. The proof follows same way as Lemma 3.2 of [6].

LEMMA 5.2. *Let π be any element of Π^{Γ}, p any element of $P(Z)$, w any bounded element of $B(\Gamma)$ and let $\varepsilon > 0$. For each $n \geqq 0$ then there exists a Borel measurable function φ_n from Z into $P(Z)$ such that*

(i) graph $\varphi_n \subset \Gamma$,

(ii) $p\pi_0 p_0 \pi_1 \cdots \pi_{n-1} p_{n-1} \{\varphi_n w \geqq \pi_n w - \varepsilon\} = 1$.

PROOF. We can prove the lemma in the same way as Lemma 6.3 of [8] by making use of Lemma 5.1.

Let

$$L_N = \sup_{\pi \in \Pi^{\Gamma}} \max_{0 \leqq n \leqq N} \| v_n^N(\pi) \|$$

for each $N \geqq 0$, where $\|\cdot\|$ means the supremum norm. Note that by virtue of Condition (B), L_N has a finite value for each N.

We now impose on g one more condition.

CONDITION (F) For each $N \geqq 1$ there exists a positive number K_N such that if $\pi \in \Pi^{\Gamma}$, $u \in B(Z)$, $\|u\| \leqq L_N$ and $\eta > 0$, then it holds that

$$E^{\pi N}[g(z_N, p_N, z_{N+1}\,;\ u(z_{N+1}) + \eta)](h_N)$$
$$- E^{\pi N}[g(z_N, p_N, z_{N+1}\,;\ u(z_{N+1}))](h_N) \leqq K_N \eta \quad \text{a.s. } (e_{\pi}) .$$

THEOREM 5.1. *Let Conditions* (C), (D) *and* (F) *be satisfied. Let $(\pi, t) \in \Lambda^{\Gamma}$ be bounded by N. For any $p \in P(Z)$ and any $\varepsilon > 0$ then there exists a Markov $(\pi^*, t^*) \in \Lambda_N^{\Gamma}$ satisfying that*

$$p\{E^{\pi^*}[u(f_{t^*})] \geqq E^{\pi}[u(f_t)] - \varepsilon\} = 1 .$$

PROOF. Let $(\pi, t) \in \Lambda_N^{\Gamma}$. Let $\gamma = \varepsilon/(N-1)$. By Lemma 4.3 and Lemma 5.2 we can find a Borel measurable function φ_{N-1}, whose graph is a subset of Γ, satisfying that

$$\pi_{N-1} p_{N-1} g(z_{N-1}, p_{N-1}, z_N\,;\ d(z_N)) \tag{5.1}$$
$$\leqq \varphi_{N-1} p_{N-1} g(z_{N-1}, p_{N-1}, z_N\,;\ d(z_N)) + \gamma / K_0 K_1 \cdots K_{N-2}$$
$$\text{a.s. } (p\pi_0 p_0 \pi_1 \cdots \pi_{N-2} p_{N-2}) ,$$

where each K_i is defined in Condition (F). From (4.3), (5.1) and the remark under

(4.3) we get

$$
\begin{aligned}
(5.2)\qquad v_{N-1}^N(\pi)&=\max\,[d(z_{N-1}),\ \pi_{N-1}p_{N-1}g(z_{N-1},\ p_{N-1},\ z_N\,;\ d(z_N))]\\
&\leqq\max\,[d(z_{N-1}),\ \varphi_{N-1}p_{N-1}g(z_{N-1},\ p_{N-1},\ z_N\,;\ d(z_N))]+\gamma/K_0K_1\cdots K_{N-2}\\
&=v_{N-1}^N(\varphi_{N-1})(z_{N-1})+\gamma/K_0K_1\cdots K_{N-2}\qquad \text{a. s. } (p\pi_0p_0\pi_1\cdots\pi_{N-2}p_{N-2}).
\end{aligned}
$$

From (5.2) and Condition (D) we have

$$
\begin{aligned}
&g(z_{N-2},\ p_{N-2},\ z_{N-1}\,;\ v_{N-1}^N(\pi))\\
&\quad\leqq g(z_{N-2},\ p_{N-2},\ z_{N-1}\,;\ v_{N-1}^N(\varphi_{N-1})(z_{N-1})+\gamma/K_0K_1\cdots K_{N-2})\\
&\qquad\qquad\text{a. s. }\ (p\pi_0p_0\pi_1\cdots\pi_{N-2}p_{N-2}),
\end{aligned}
$$

which implies that

$$
\begin{aligned}
(5.3)\qquad &\pi_{N-2}p_{N-2}g(z_{N-2},\ p_{N-2},\ z_{N-1}\,;\ v_{N-1}^N(\pi))\\
&\quad\leqq\pi_{N-2}p_{N-2}g(z_{N-2},\ p_{N-2},\ z_{N-1}\,;\ v_{N-1}^N(\varphi_{N-1})(z_{N-1})+\gamma/K_0K_1\cdots K_{N-2})\\
&\qquad\qquad\text{a. s. }\ (p\pi_0p_0\pi_1\cdots\pi_{N-3}p_{N-3}).
\end{aligned}
$$

By virtue of Condition (F), (5.3) yields that

$$
\begin{aligned}
(5.4)\qquad &\pi_{N-2}p_{N-2}g(z_{N-2},\ p_{N-2},\ z_{N-1}\,;\ v_{N-1}^N(\pi))\\
&\quad\leqq\pi_{N-2}p_{N-2}g(z_{N-2},\ p_{N-2},\ z_{N-1}\,;\ v_{N-1}^N(\varphi_{N-1})(z_{N-1}))+\gamma/K_0K_1\cdots K_{N-3}\\
&\qquad\qquad\text{a. s. }\ (p\pi_0p_0\pi_1\cdots\pi_{N-3}p_{N-3}).
\end{aligned}
$$

Applying again Lemma 5.2, we can find a Borel measurable function φ_{N-2}, whose graph is a subset of Γ, satisfying

$$
\begin{aligned}
(5.5)\qquad &\pi_{N-2}p_{N-2}g(z_{N-2},\ p_{N-2},\ z_{N-1}\,;\ v_{N-1}^N(\varphi_{N-1})(z_{N-1}))\\
&\quad\leqq\varphi_{N-2}p_{N-2}g(z_{N-2},\ p_{N-2},\ z_{N-1}\,;\ v_{N-1}^N(\varphi_{N-1})(z_{N-1}))+\gamma/K_0K_1\cdots K_{N-3}\\
&\qquad\qquad\text{a. s. }\ (p\pi_0p_0\pi_1\cdots\pi_{N-3}p_{N-3}).
\end{aligned}
$$

Combining (5.4) and (5.5) leads us to

$$
\begin{aligned}
&\pi_{N-2}p_{N-2}g(z_{N-2},\ p_{N-2},\ z_{N-1}\,;\ v_{N-1}^N(\pi))\\
&\quad\leqq\varphi_{N-2}p_{N-2}g(z_{N-2},\ p_{N-2},\ z_{N-1}\,;\ v_{N-1}^N(\varphi_{N-1})(z_{N-1}))+2\gamma/K_0K_1\cdots K_{N-3}\\
&\qquad\qquad\text{a. s. }\ (p\pi_0p_0\pi_1\cdots\pi_{N-3}p_{N-3}),
\end{aligned}
$$

which further implies that

$$
\begin{aligned}
v_{N-2}^N(\pi)(h_{N-2})&=\max\,[d(z_{N-2}),\ \pi_{N-2}p_{N-2}g(z_{N-2},\ p_{N-2},\ z_{N-1}\,;\ v_{N-1}^N(\pi))]\\
&\leqq v_{N-2}^N(\{\varphi_{N-2},\ \varphi_{N-1}\})(z_{N-2})+2\gamma/K_0K_1\cdots K_{N-3}\\
&\qquad\qquad\text{a. s. }\ (p\pi_0p_0\pi_1\cdots\pi_{N-3}p_{N-3}).
\end{aligned}
$$

Repeating this procedure enables us to choose Borel measurable functions $\varphi_0,\ \varphi_1,\ \cdots,\ \varphi_{N-1}$ such that graph φ_i is a subset of Γ for each i $(0\leqq i\leqq N-1)$ and such that

$$v_0^N(\pi) \leqq v_0^N(\{\varphi_0, \varphi_1, \cdots, \varphi_{N-1}\}) + \varepsilon \qquad \text{a. s. } (p). \tag{5.6}$$

Let π^* denotes the Markov policy $\{\varphi_0, \varphi_1, \cdots, \varphi_{N-1}\}$, and define t^* by

t^*=the first n such that $v_n^N(\pi^*) = d(z_n)$, then $(\pi^*, t^*) \in \Lambda_N^\Gamma$. Since for each n $(0 \leqq n \leqq N)$, $v_n^N(\pi^*)$ does not depend on the history $(z_0, p_0, z_1, p_1, \cdots, p_{n-1})$ but only on z_n, t^* is a Markov stopping time, so that (π^*, t^*) is Markov. From (6) of Lemma 4.8, Lemma 4.9 and (5.6) we obtain

$$p\{E^{\pi^*}[u(f_{t^*})] \geqq E^{\pi}[u(f_t)] - \varepsilon\} = 1,$$

which completes the proof.

COROLLARY 5.1. *Let Conditions* (C), (D), (E) *and* (F) *be satisfied. For any* $p \in P(Z)$ *and any* $\varepsilon > 0$, *then there exists a* $(\bar{p}, \varepsilon)$*-optimal Markov policy which is bounded*

PROOF. The corollary is direct from Theorem 4.3 and Theorem 5.1.

COROLLARY 5.2. *Let Conditions* (C), (D), (E) *and* (F) *be satisfied. Then we have*

(a) *if* $\{E^{\hat{\pi}^N}[u(f_{\hat{t}_N})]\}$ *converges uniformly on* Z, *for any* $p \in P(Z)$ *and* $\varepsilon > 0$, *then there exists a* (p, ε)*-optimal Markov bounded policy,*

(b) *if* S *is a finite set, for any* $p \in P(\{0\} \times S)$ *and any* $\varepsilon > 0$, *then there exists a* (p, ε)*-optimal Markov bounded policy,*

(c) *if* $p \in P(Z)$ *has a finite support, for any* $\varepsilon > 0$, *then there exists a* (p, ε)*-optimal Markov bounded policy.*

PROOF. The corollary is direct from Corollary 4.2 and Theorem 5.1.

6. Examples.

EXAMPLE 1. (Additive process)

$$g(z_m, p_m, z_{m+1}, \cdots, z_n; d(z_n)) = \sum_{k=m}^{n-1} r_k(s_k, p_k, s_{k+1}) + d(z_n),$$

where each r_k is bounded Borel measurable on $\Gamma \times S$ and d bounded Borel measurable on Z. It is clear that in this example g is recursive and monotone, and satisfies Condition (B2) and (F). Let $X_n' = \sum_{k=0}^{n-1} r_k^+$ and let $X_n'' = \sum_{k=0}^{n-1} r_k^-$, where $a^+ = \max(a, 0)$ and $a^- = \max(-a, 0)$, then we have

$$g(z_0, p_0, z_1, \cdots, z_n; d(z_n)) = X_n' - X_n'' + d(z_n). \tag{6.1}$$

For any stopping time $t \in C(\pi)$ we have from (6.1) and the boundedness of d that for $n < t$,

$$g(z_0, p_0, z_1, \cdots, z_n; d(z_n)) \geqq -X_t'' + d(z_n) \geqq -X_t'' - K$$

for some positive K, so that

$$\int_{\{t>n\}} (g(z_0, p_0, z_1, \cdots, z_n; d(z_n)))^- de_\pi \leqq \int_{\{t>n\}} X_t'' de_\pi + KP^\pi\{t > n\}. \tag{6.2}$$

We assume that

$$\sup_{\pi \in \Pi^\Gamma} E^\pi[\sup_n X_n'] < +\infty. \tag{6.3}$$

It is easily checked that (6.3) implies Condition (B1). We next show that Condition (E) also follows from (6.3). Suppose that $E^{\pi}[g(z_0, p_0, z_1, \cdots, z_t; d(z_t))]\neq -\infty$. Since (B1) holds, then $E^{\pi}[g(z_0, p_0, z_1, \cdots, z_t; d(z_t))]$ is finite. By (6.3) $E^{\pi}[X'_t]$ is finite, so that $E^{\pi}[X''_t]$ is also finite. Hence from (6.2) it follows that Condition (E) holds true.

By the following we shall give several typical cases for which (6.3) is satisfied.

Case 1. (Discounted dynamic programming) $r_k=\beta^{k-1}r$ for all k and $d(z_n)=\beta^{n-1}d'(s_n)$, where r is bounded and $0\leqq\beta<1$.

Case 2. (Negative dynamic programming) $r_k\leqq 0$ for all k and $d\leqq 0$.

Case 3. (Terminal control problem) $r_k=0$ for all k.

Case 4. (Statistical sequential analysis in the case when control actions are involved) $r_k=r'_k-c$ for all k, where $r'_k\geqq 0$, $c>0$ and $\sup_{\pi\in\Pi^{\Gamma}} E^{\pi}[\sup_n(\sum_{k=0}^{n} r'_k)]<+\infty$. c is interpreted as a sampling cost.

PROPOSITION 6.1. *Let Condition* (B2) *be satisfied. Assum that there exists an integer* N *such that for every* π, $\{(g(z_0, p_0, z_1, \cdots, z_n; d(z_n)))^-\}$ *is nondecreasing in* $n\geqq N$ *almost surely* (P^{π}). *Then Conditions* (B1) *and* (E) *hold true.*

PROOF. It is clear that (B1) follows from two assumptions given in the proposition.

Suppose that $E^{\pi}[g(z_0, p_0, z_1, \cdots, z_t; d(z_t))]\neq -\infty$. Then from (B1), $E^{\pi}[g(z_0, p_0, z_1, \cdots, z_t; d(z_t))]$ is finite. Since again from (B1), $E^{\pi}[(g(z_0, p_0, z_1, \cdots, z_t; d(z_t)))^+]$ is finite, $E^{\pi}[(g(z_0, p_0, z_1, \cdots, z_t; d(z_t)))^-]$ is finite. For $n\geqq N$, if $t>n$ then $(g(z_0, p_0, z_1, \cdots, z_n; d(z_n)))^-\leqq(g(z_0, p_0, z_1, \cdots, z_t; d(z_t)))^-$ almost surely (P^{π}). Hence we have

$$(6.4)\qquad 0\leqq\int_{\{t>n\}}(g(z_0, p_0, z_1, \cdots, z_n; d(z_n))^- de_{\pi}$$

$$\leqq\int_{\{t>n\}}(g(z_0, p_0, z_1, \cdots, z_t; d(z_t))^- de_{\pi}$$

for $n\geqq N$. Since $E^{\pi}[(g(z_0, p_0, z_1, \cdots, z_t; d(z_t)))^-]$ is finite, from (6.4) we get

$$\liminf_{n\to\infty}\int_{\{t>n\}}(g(z_0, p_0, z_1, \cdots, z_n; d(z_n)))^- de_{\pi}=0,$$

which completes the proof.

EXAMPLE 2. (Additive process)

$$g(z_m, p_m, z_{m+1}, \cdots, z_n; d(z_n))=\sum_{k=m}^{n-1} r_k(s_k, p_k, s_{k+1})+d(z_n),$$

where all components are Borel measurable and

$$(6.5)\qquad r_k=r'_k-\alpha k^2, \qquad k=1, 2, \cdots, \qquad (\alpha>0)$$

$$(6.6)\qquad 0\leqq r'_k\leqq M, \qquad k=1, 2, \cdots, \qquad (0<M<+\infty)$$

and

$$(6.7)\qquad |d|\leqq L \qquad (0<L<+\infty).$$

It is clear that Condition (B2) is satisfied. We have from (6.5), (6.6) and (6.7) that

$$g(z_0, p_0, z_1, \cdots, z_n; d(z_n)) \leqq nM + L - \alpha(1^2 + 2^2 + \cdots + (n-1)^2).$$

Choose N so that

$$(n+1)M + 2L < \alpha n^2 \qquad \text{for all } n \geqq N,$$

then we have

$$g(z_0, p_0, z_1, \cdots, z_{n+1}; d(z_{n+1}))$$
$$< g(z_0, p_0, z_1, \cdots, z_n; d(z_n)) < 0 \qquad \text{for all } n \geqq N+1,$$

which implies that $\{(g(z_0, p_0, z_1, \cdots, z_n; d(z_n)))^-\}$ is increasing in $n \geqq N+1$. Hence from Proposition 6.1, Conditions (B1) and (E) turn out to hold.

EXAMPLE 3. (Multiplicative process)

$$g(z_m, p_m, z_{m+1}, \cdots, z_n; d(z_n)) = \prod_{k=m}^{n-1} r_k(s_k, p_k, s_{k+1}) d(z_n),$$

where all components are Borel measurable and

(6.8) $$0 \leqq r_k \leqq 1, \qquad k = 1, 2, 3, \cdots,$$

(6.9) $$|d| \leqq L \qquad (0 < L < +\infty).$$

Trivially

$$|g(z_0, p_0, z_1, \cdots, z_n; d(z_n))| \leqq L \qquad \text{for all } n.$$

Then it follows that

$$0 \leqq \int_{\{t>n\}} (g(z_0, p_0, z_1, \cdots, z_n; d(z_n)))^- de_\pi \leqq LP^\pi\{t>n\}.$$

Hence Condition (E) holds. It is clear that all other conditions are satisfied.

EXAMPLE 4. (Multiplicative additive process)

$$g(z_m, p_m, z_{m+1}, \cdots, z_n; d(z_n))$$
$$= \sum_{j=m}^{n-1} \prod_{k=m}^{j} r_k(s_k, p_k, s_{k+1}) + \prod_{k=m}^{n-1} r_k(s_k, p_k, s_{k+1}) d(z_n),$$

where all components are Borel measurable, accompanied with (6.8) and (6.9).

Trivially

$$0 \leqq (g(z_0, p_0, z_1, \cdots, z_n; d(z_n)))^- \leqq L \qquad \text{for all } n.$$

Then we can see in the same way as Example 3 that Condition (E) holds. All other conditions are also satisfied.

In the following examples we can easily check that all conditions are satisfied.

EXAMPLE 5. (Divided additive process)

$$g(z_m,\ p_m,\ z_{m+1},\ \cdots,\ z_n\ ;\ d(z_n))$$
$$=h(r_m(s_m,\ p_m,\ s_{m+1}))+\sum_{k=m}^{n-2}\frac{h(r_{k+1}(s_{k+1},\ p_{k+1},\ s_{k+2}))}{\prod_{j=m}^{k} r_j(s_j,\ p_j,\ s_{j+1})}$$
$$+\frac{d(z_n)}{\prod_{j=m}^{n-1} r_j(s_j,\ p_j,\ s_{j+1})},$$

where all components r_j are Borel measurable, h and d bounded Borel measurable and

(6.10) for some constant $K>1$, $r_j \geqq K$ for all j.

EXAMPLE 6. (Exponential additive process)

$$g(z_m,\ p_m,\ z_{m+1},\ \cdots,\ z_n\ ;\ d(z_n))=\sum_{j=m}^{n-1} r_j(s_j,\ p_j,\ s_{j+1})\exp\left[-\sum_{k=m}^{j} v_k(s_k,\ p_k,\ s_{k+1})\right]$$
$$+d(z_n)\exp\left[-\sum_{k=m}^{n-1} v_k(s_k,\ p_k,\ s_{k+1})\right],$$

where all components r_j and v_k are Borel measurable, d bounded Borel measurable and

(6.11) for some positive K_1, $v_k \geqq K_1$ for all k,

(6.12) $r_j = r'_j - r''_j$, $r'_j \geqq 0$ and $r''_j \geqq 0$ for all j,

(6.13) for some positive K_2, $0 \leqq r'_j \leqq K_2$ for all j.

These examples cited above are typical in the non-stationary DP-problems, nevertheless somewhat restrictive in contrast with our general formulation which was intended so as to include a broad class of the DP-problems. Namely, in every example, g has a regularity of its form even if the stage-wise reward is not stationary. Such a regularity is not requisite for the principle of optimality. In reality we can give some mixed types of above examples which keep all of conditions required. But in this paper we shall not illustrate the details.

References

[1] BELLMAN, R. Dynamic Programming. Princeton University Press, Priceton, New Jersey, 1957.

[2] BLACKWELL, D., *Discounted dynamic programming.* Ann. Math. Statist. **36** (1965), 226-235.

[3] CHOW, Y.S. and ROBBINS, H., *On optimal stopping rules.* Z. Wahrscheinlichkeits-theorie und Verw.Gebiete **2** (1963), 33-49.

[4] DUBINS, L. and FREEDMAN, D., *Measurable sets of measures.* Pacific J. Math. **14** (1965), 1211-1222.

[5] DUBINS, L. and SAVAGE, L.J., How to Gamble if you Must. McGraw-Hill, New York, 1965.

[6] FURUKAWA, N., *Markovian decision processes with compact action spaces.* Ann. Math. Statist. **43** (1972), 1612-1622.

[7] FURUKAWA, N., *Functional equations and Markov potential theory in stopped decision processes.* Mem. Fac. Sci., Kyushu University, Ser. A, **29** (1975), 329-347.

[8] FURUKAWA, N. and IWAMOTO, S., *Stopped decision processes on complete separable metric spaces.* J. Math. Anal. Appl. **31** (1970), 615-658.

[9] FURUKAWA, N. and IWAMOTO, S., *Markovian decision processes with recursive reward functions.* Bull. Math. Statist. **15**, Nos. 3-4 (1973), 79-91.

[10] FURUKAWA, N. and IWAMOTO, S., *Dynamic programming on recursive reward systems.* Bull. Math. Statist. **17**, Nos. 1-2 (1976), 103-126.

[11] HINDERER, K., Foundations of Non-stationary Dynamic Programming with Discrete Time Parameter. Lecture Notes in Operations Research and Mathematical Systems, Springer-Verlag, 1970.

[12] MACKY, G.W., *Borel structure in groups and their duals.* Trans. Amer. Math. Soc. **85** (1957), 134-165.

[13] PARTHASARATHY, K.R., Probability Mersures on Metric Spaces. Acad. Press, 1967.

[14] STRAUCH, R.E., *Negative dynamic programming,* Ann. Math. Statist. **37** (1966), 871-890.

[15] STRAUCH, R.E., *Measurable gambling houses.* Trans. Amer. Math. Soc. **126** (1967), 64-72.

[16] SUDDERTH, W.D., *On the existence of good stationary strategies.* Trans. Amer. Math. Soc. **135** (1969), 399-414.

[17] VON NEUMANN, J., *On rings of operators; Reduction theory.* Ann. Math. **50** (1949), 401-485.