Kybernetika

Karel Sladký

On the set of optimal controls for Markov chains with rewards

Kybernetika, Vol. 10 (1974), No. 4, (350)--367

Persistent URL: http://dml.cz/dmlcz/125693

KYBERNETIKA — VOLUME 10 (1974), NUMBER 4

On the Set of Optimal Controls for Markov Chains with Rewards

KAREL SLADKÝ

On the base of a recurrence formula for the expected reward of a finite Markov chain (that is an extension of Laurent expansion for expected discounted reward obtained in [6], [9] as well as the recurrence relation for expected non-discounted reward inferred in [5], [7]) necessary and sufficient optimality conditions for controlled Markov chains with sensitive discounted and sensitive averaging criteria are established. Employing the obtained optimality conditions criterion equivalence between these two types of optimality criteria is shown.

0. INTRODUCTION AND NOTATIONS

In a fundamental paper on finite state controlled Markov chains Blackwell [1] defines an optimal policy (control) as β optimal for discount factor β sufficiently near 1 and establishes (nonconstructively) the existence of a stationary optimal policy. In [1] Blackwell also introduces a nearly optimal policy (control) as that for which the difference betweeen the total expected discounted reward for that policy and β optimal policy tends to 0 as $\beta \to 1$ and gives an algorithm that in special cases enables to find a nearly optimal policy. A general form of the algorithm for finding a nearly optimal policy was found by Veinott in [8].

In [6] Miller and Veinott gave a constructive proof of the existence and an algorithm for finding an optimal policy. Their approach is based on the Laurent expansion for expected discounted reward for β near 1. A similar approach was also used by Veinott in [9] for investigating sensitive discounted optimal policies.

Necessary and sufficient optimality conditions for Markov chains were derived in [3] for the case with discounting and in [5], [7] for the long range average reward of Markov chains. In [2] Denardo and Miller verified Veinott's conjecture on the existence of a stationary average overtaking optimal policy and established necessary and sufficient optimality conditions for that optimality criterion (another form of these optimality conditions was inferred in [7]).

In the present paper we shall investigate necessary and sufficient optimality conditions for sensitive discounted and sensitive averaging optimality criteria. First (compare Theorem 1.7 and Corollary 1.8) we shall infer a recurrence formula for the expected reward of a controlled Markov chain for an arbitrary control (policy). This formula is an extension of Laurent expansion for expected discounted reward obtained in [6], [9] as well as the recurrence relation for expected non-discounted reward inferred in [5], [7] (compare Remark 1.10 and Corollary 1.5). Using the obtained recurrence formula we shall establish (see Theorem 2.2 and 2.4) the necessary and sufficient optimality conditions for sensitive discounted optimality criteria. The (sufficient) optimality conditions obtained in [6] for optimal policy and in [9] for optimal policies with sensitive discounted optimality criteria immediately follow from Theorem 2.2 and 2.4.

In the last paragraph of the paper we shall infer necessary and sufficient optimality conditions for sensitive averaging optimal controls (policies) (compare Theorem 3.5). This part of the paper extends some author's results obtained in [7] and also establishes a criterion equivalence (compare Theorem 3.6) between sensitive discounted and sensitive averaging optimality criteria. Only a partial result in this direction was obtained by Lippman [4] concerning the criterion equivalence between nearly optimal and average overtaking optimal policies.

Throughout the paper notations and terminology used in [5], [7] will be followed as close as possible.

We shall consider a controlled Markov chain with state space $I = \{1, 2, \ldots, r\}$ and the set of control parameter values (actions) $z \in J = \{1, 2, \ldots, s\}$ in any of the state. Choosing control parameter value $z \in J$ in state $j \in I$ state $k \in I$ will be reached in the next transition with given probability $p(j, k; z)$, and one stage reward $c(j, k; z)$ will be obtained from such a transition. The values $p(j, k; z)$, $c(j, k; z)$ are supposed to be known for any pair $j, k \in I$ and any $z \in J$. For the sake of brevity we shall introduce the expected one stage reward $\bar{c}(j;z)$ in state j if the control parameter value takes value z. Obviously, $\bar{c}(j; z) = \sum_{k \in I} p(j, k; z) \,.\, c(j, k; z)$. A control (policy) ω of the chain is given by a sequence of control parameter values (actions) $z \in J$ chosen with respect to the complete history of the chain. So we write $\omega = \{z_n(j_0, j_1, \ldots, j_n),$ $n = 0, 1, \ldots\}$ where $z_n(j_0, j_1, \ldots, j_n)$ is the control parameter value (action) chosen at the n-th transition following the occurrence of states $j_0, j_1, \ldots, j_n$. ω is called a Markovian (memoryless) control (policy) if $z_n(j_0, j_1, \ldots, j_n) = z_n(j_n)$ for $n = 0, 1, \ldots$. A Markovian control is called homogeneous if $z_n(j_n) = z(j_n)$. A Markovian homogeneous control will be called stationary. For stationary control we write $\omega \sim z(j)$.

Denoting X_n (for $n = 0, 1, \ldots$) the random state of the considered Markov chain at the n-th transition the probability distribution of a sequence $\{X_n; n = 0, 1, \ldots\}$ is determined by the chosen control ω and the initial state $j \in I$ (of course, $X_0 = j$). The symbol E_j^ω is used for mathematical expectation with respect to this probability distribution. For shortening we shall often delete the arguments $X_0, X_1, \ldots, X_k$ and

$j_0, j_1, \ldots, j_k$ in $z_k(\ldots)$ e.g. we shall write z_2 instead of $z_2(X_0, X_1, X_2)$ or instead of $z_2(j_0, j_1, j_2)$.

The (random) β-discounted reward up to the N following transitions $C_N(\beta)$ is given by $C_N(\beta) = \sum_{n=0}^{N-1} \beta^n . c(X_n, X_{n+1}; z_n)$ (we set $C_0(\beta) = 0$). Here $\beta \in \langle 0, 1)$ is a given discount factor and $\varrho = (1 - \beta)/\beta > 0$ denotes the associated rate of interest. Obviously, for any control $\omega \sim z_n$

$$E_j^{\omega} C_N(\beta) = \sum_{n=0}^{N-1} \beta^n . E_j^{\omega} \bar{c}(X_n; z_n)$$

where $X_0 = j$ and

$$E_j^{\omega} C(\beta) = \lim_{N \to \infty} E_j^{\omega} C_N(\beta) < \infty$$

always exists.

Following Veinott (compare [9]) a control ω_l will be called l-discount optimal (for $l = -1, 0, 1, \ldots$) if

$$\liminf_{\beta \to 1-} \left(\frac{1-\beta}{\beta}\right)^{-l} \left[E_j^{\omega_l} C(\beta) - E_j^{\omega} C(\beta)\right] \geqq 0$$

for any control ω and all $j \in I$. Note that in the terminology introduced by Blackwell [1] 0-discount optimal control is a nearly optimal one and an optimal control is l-discount optimal for $l = \infty$.

For the investigation of sensitive averaging controls we shall introduce the concept of l-order average reward optimal controls. If control $\omega \sim z_n$ is used, for the expected value of the (random) undiscounted reward up to the N following transitions C_N obviously holds $E_j^{\omega} C_N = \sum_{n=0}^{N-1} E_j^{\omega} \tilde{c}(X_n; z_n)$. Let us set $v_0^{(N)}(j; \omega) = E_j^{\omega} C_N$ and let us for $l = 0, 1, \ldots$ define the values $v_{l+1}^{(N)}(j; \omega) = \sum_{n=0}^{N-1} v_l^{(n)}(j; \omega)$ (obviously, $v_0^{(0)}(j; \omega) = = v_1^{(0)}(j; \omega) = v_1^{(1)}(j; \omega) = \ldots = v_l^{(0)}(j; \omega) = v_l^{(1)}(j; \omega) = \ldots = v_l^{(l)}(j; \omega) = 0$). A control $\hat{\omega}_l$ will be called l-order average reward optimal (for $l = 0, 1, \ldots$) if for any other control ω and all $j \in I$

$$\liminf_{N \to \infty} \frac{1}{N - l} \left[v_l^{(N)}(j; \hat{\omega}_l) - v_l^{(N)}(j; \omega)\right] \geqq 0 .$$

Note that any average reward optimal control is 0-order average reward optimal and 1-order average reward optimal control is usually called an average overtaking optimal control.

1. RECURRENCE RELATION FOR EXPECTED DISCOUNTED REWARD

In this paragraph a recurrence formula for expected discounted reward will be inferred. This recurrence formula will be very useful for investigating expected discounted reward in case that the discount factor β (resp. the interest rate ϱ) is near 1 (resp. 0).

First, we shall need some preliminary lemmas. Let for a stationary control $\tilde{\omega} \sim \tilde{z}(j)$

$$\mathbf{g} = \|g_j\|_{j=1}^r\,, \quad \mathbf{c}^{\tilde{\omega}} = \|c(j;\,\tilde{z}(j))\|_{j=1}^r\,, \quad \mathbf{u}_m = \|u_{m,j}\|_{j=1}^r \quad (m = 0, 1, \ldots)$$

be column vectors and let us denote by $\pi(j, k; \tilde{\omega})$ the limit probability of transition probability matrix $\mathbf{P}^{\tilde{\omega}} = \|p(j, k; \tilde{z}(j))\|_{j,k=1}^r$ defined as

$$\mathbf{\Pi}^{\tilde{\omega}} = \|\pi(j, k; \tilde{\omega})\|_{j,k=1}^r = \lim_{m\to\infty} \frac{1}{m+1} \sum_{n=0}^{m} (\mathbf{P}^{\tilde{\omega}})^n$$

(note that $(\mathbf{P}^{\tilde{\omega}})^0 = \mathbf{I}$ where $\mathbf{I}$ is $(r \times r)$ unit matrix).

1.1. Lemma. *For any stationary control $\tilde{\omega} \sim \tilde{z}(j)$ there exists a (unique) solution (denoted $\mathbf{g}^{\tilde{\omega}}, \mathbf{u}_0^{\tilde{\omega}}$) of equations*

$$(1.1) \qquad \mathbf{g} = \mathbf{P}^{\tilde{\omega}} . \mathbf{g}\,,$$

$$(1.2) \qquad \mathbf{u}_0 + \mathbf{g} = \mathbf{c}^{\tilde{\omega}} + \mathbf{P}^{\tilde{\omega}} . \mathbf{u}_0\,,$$

$$(1.2') \qquad \mathbf{\Pi}^{\tilde{\omega}} . \mathbf{u}_0 = \mathbf{0}\,.$$

Proof. The existence of a unique $\mathbf{g}^{\tilde{\omega}}$ satisfying (1.1), (1.2) immediately follows after multiplying (1.2) by $\mathbf{\Pi}^{\tilde{\omega}}$ (using the equality $\mathbf{P} . \mathbf{\Pi} = \mathbf{\Pi}$ we obtain $\mathbf{\Pi}^{\tilde{\omega}} . \mathbf{g}^{\tilde{\omega}} = \mathbf{\Pi}^{\tilde{\omega}} . \mathbf{c}^{\tilde{\omega}}$ as rank $(\mathbf{I} - \mathbf{P})$ + rank $\mathbf{\Pi} = r$). Then from (1.2), (1.2') immediately follows the existence of a unique solution $\mathbf{u}_0^{\tilde{\omega}}$ (compare also [1] § 3). □

1.2. Lemma. *Let $\tilde{\omega} \sim \tilde{z}(j)$ be a stationary control and let $\mathbf{u}_0^{\tilde{\omega}}$ be the solution to (1.1), (1.2), (1.2'). Then there exists a sequence $\{\mathbf{u}_m^{\tilde{\omega}}, m = 1, 2, \ldots\}$ where $\mathbf{u}_m^{\tilde{\omega}}$ is determined from $\mathbf{u}_{m-1}^{\tilde{\omega}}$ as a (unique) solution of the equations*

$$(1.3) \qquad \mathbf{u}_m + \mathbf{u}_{m-1}^{\tilde{\omega}} = \mathbf{P}^{\tilde{\omega}} . \mathbf{u}_m\,,$$

$$(1.3') \qquad \mathbf{\Pi}^{\tilde{\omega}} . \mathbf{u}_m = \mathbf{0}\,.$$

Proof. As rank $(\mathbf{I} - \mathbf{P})$ + rank $\mathbf{\Pi} = r$ the existence of a (unique) solution to (1.3), (1.3') is obvious. □

Introducing the fundamental matrix $\mathbf{Z} = (\mathbf{I} - \mathbf{P} + \mathbf{\Pi})^{-1}$ (that always exists, see e.g. [1]) from (1.1), (1.2), (1.2') we obtain

$$(1.4) \qquad \mathbf{u}_0^{\tilde{\omega}} = \mathbf{Z}^{\tilde{\omega}}(\mathbf{c}^{\tilde{\omega}} - \mathbf{g}^{\tilde{\omega}})$$

where

$$(1.4') \qquad \boldsymbol{g}^{\tilde{\omega}} = \boldsymbol{\Pi}^{\tilde{\omega}} . \mathbf{c}^{\tilde{\omega}} .$$

Similarly, from (1.3), (1.3′) we have for $m = 1, 2, \ldots$

$$(1.5) \qquad \boldsymbol{u}_m^{\tilde{\omega}} = -\boldsymbol{Z}^{\tilde{\omega}} . \boldsymbol{u}_{m-1}^{\tilde{\omega}} .$$

In virtue of (1.4), (1.4′) and (1.5) we have for $m = 0, 1, 2, \ldots$

$$(1.6) \qquad \boldsymbol{u}_m^{\tilde{\omega}} = -(-\boldsymbol{Z}^{\tilde{\omega}})^{m+1} (\boldsymbol{I} - \boldsymbol{\Pi}^{\tilde{\omega}}) . \mathbf{c}^{\tilde{\omega}} .$$

1.3. Remark. At most r vectors $\boldsymbol{u}_m^{\tilde{\omega}}$ can be linearly independent. Moreover, if the vectors $\boldsymbol{u}_0^{\tilde{\omega}}, \boldsymbol{u}_1^{\tilde{\omega}}, \ldots, \boldsymbol{u}_L^{\tilde{\omega}}$ (where $L \leqq r - 1$) are linearly independent and $\boldsymbol{u}_{L+1}^{\tilde{\omega}}$ is their linear combination using (1.5) we can easily verify by induction that for any $m > L$ $\boldsymbol{u}_m^{\tilde{\omega}}$ must be a linear combination of $\boldsymbol{u}_0^{\tilde{\omega}}, \boldsymbol{u}_1^{\tilde{\omega}}, \ldots, \boldsymbol{u}_L^{\tilde{\omega}}$ (resp. of $\boldsymbol{u}_1^{\tilde{\omega}}, \ldots, \boldsymbol{u}_L^{\tilde{\omega}}, \boldsymbol{u}_{L+1}^{\tilde{\omega}}$) and that any $L + 1$ vectors $\boldsymbol{u}_{m_1}^{\tilde{\omega}}, \boldsymbol{u}_{m_2}^{\tilde{\omega}}, \ldots, \boldsymbol{u}_{m_{L+1}}^{\tilde{\omega}}$ must be linearly dependent.

Let us denote

$$(1.7) \qquad \psi_{-1}^{\tilde{\omega}}(j; z) = \sum_{k \in I} p(j, k; z) . g_k^{\tilde{\omega}} - g_j^{\tilde{\omega}} ,$$

$$(1.8) \qquad \psi_0^{\tilde{\omega}}(j; z) = \bar{c}(j; z) + \sum_{k \in I} p(j, k; z) . u_{0,k}^{\tilde{\omega}} - u_{0,j}^{\tilde{\omega}} - g_j^{\tilde{\omega}}$$

and for $m = 1, 2, \ldots$

$$(1.9) \qquad \psi_m^{\tilde{\omega}}(j; z) = -u_{m-1,j}^{\tilde{\omega}} + \sum_{k \in I} p(j, k; z) . u_{m,k}^{\tilde{\omega}} - u_{m,j}^{\tilde{\omega}} .$$

We shall denote by $J_m^{\tilde{\omega}}(j)$ (for $m = -1, 0, 1, \ldots$) the set of all control parameter values in state j for which $\psi_{-1}^{\tilde{\omega}}(j; z) = \psi_0^{\tilde{\omega}}(j; z) = \ldots = \psi_m^{\tilde{\omega}}(j; z) = 0$. Of course, $J_m^{\tilde{\omega}}(j) \subset J_{m-1}^{\tilde{\omega}}(j) \subset \ldots \subset J_{-1}^{\tilde{\omega}}(j) \subset J$ for any $j \in I$.

Now we shall try to find a Laurent expansion of an arbitrary control $\omega \sim z_n$. The partial form of this expansion brings

1.4. Lemma. *Let $\tilde{\omega} \sim \tilde{z}(j)$ be a stationary control. Then for an arbitrary control $\omega \sim z_n$ and all $j \in I$; $N = 0, 1, 2, \ldots$*

(1.10)

$$E_j^{\omega} C_N(\beta) = (1 - \beta)^{-1} \left[(1 - \beta^N) . g_j^{\tilde{\omega}} + \sum_{n=0}^{N-2} (\beta^{n+1} - \beta^N) . E_j^{\omega} \psi_{-1}^{\tilde{\omega}}(X_n; z_n)\right] +$$
$$+ \left[\frac{1}{\beta} u_{0,j}^{\tilde{\omega}} - \beta^N . E_j^{\omega} u_{0,X_N}^{\tilde{\omega}} + \sum_{n=0}^{N-1} \beta^n . E_j^{\omega} \psi_0^{\tilde{\omega}}(X_n; z_n)\right] -$$
$$- \frac{1 - \beta}{\beta} . \sum_{n=0}^{N} \beta^n . E_j^{\omega} u_{0,X_n}^{\tilde{\omega}} .$$

Proof. Using (1.7), (1.8) for the expected discounted reward up to the N following transitions we can write

$$(1.11) \qquad E_j^\omega C_N(\beta) = \sum_{n=0}^{N-1} \beta^n . E_j^\omega \bar{c}(X_n; z_n) =$$

$$= \sum_{n=0}^{N-1} \beta^n . E_j^\omega \psi_0^{\tilde{\omega}}(X_n; z_n) + \sum_{n=0}^{N-1} \beta^n . E_j^\omega g_{X_n}^{\tilde{\omega}} - \sum_{n=0}^{N-1} \beta^n . E_j^\omega(u_{0,X_{n+1}}^{\tilde{\omega}} - u_{0,X_n}^{\tilde{\omega}}) .$$

For the second term on the righthand side of (1.11) we have (as $X_0 = j$)

$$(1.12) \qquad \sum_{n=0}^{N-1} \beta^n . E_j^\omega g_{X_n}^{\tilde{\omega}} = \sum_{n=0}^{N-1} \beta^n . E_j^\omega[g_{X_0}^{\tilde{\omega}} + \sum_{m=0}^{n-1} (g_{X_{m+1}}^{\tilde{\omega}} - g_{X_m}^{\tilde{\omega}})] =$$

$$= \sum_{n=0}^{N-1} \beta^n[g_{X_0}^{\tilde{\omega}} + \sum_{m=0}^{n-1} E_j^\omega \psi_{-1}^{\tilde{\omega}}(X_m; z_m)] = \frac{1 - \beta^N}{1 - \beta} . g_j^{\tilde{\omega}} +$$

$$+ \frac{1}{1 - \beta} \sum_{n=0}^{N-2} (\beta^{n+1} - \beta^N) . E_j^\omega \psi_{-1}^{\tilde{\omega}}(X_n; z_n) .$$

For the last term on the righthand side of (1.11) we can write

$$(1.13) \qquad - \sum_{n=0}^{N-1} \beta^n . E_j^\omega(u_{0,X_{n+1}}^{\tilde{\omega}} - u_{0,X_n}^{\tilde{\omega}}) = u_{0,j}^{\tilde{\omega}} - \beta^N . E_j^\omega u_{0,X_N}^{\tilde{\omega}} -$$

$$- (1 - \beta) \sum_{n=0}^{N-1} \beta^n . E_j^\omega u_{0,X_{n+1}}^{\tilde{\omega}} .$$

Setting (1.12), (1.13) into (1.11) we obtain (1.10). □

1.5. Corollary. *As*

$$(1 - \beta)^{-1} . [(1 - \beta^N) . g_j^{\tilde{\omega}} + \sum_{n=0}^{N-2} (\beta^{n+1} - \beta^N) . E_j^\omega \psi_{-1}^{\tilde{\omega}}(X_n; z_n)] =$$

$$= g_j^{\tilde{\omega}} \sum_{n=0}^{N-1} \beta^n + \sum_{n=0}^{N-2} \sum_{m=n+1}^{N-1} \beta^m . E_j^\omega \psi_{-1}^{\tilde{\omega}}(X_n; z_n)$$

letting $\beta \to 1$ from (1.10) we obtain

$$(1.14) \qquad E_j^\omega C_N = N . g_j^{\tilde{\omega}} + \sum_{n=0}^{N-1} [(N - 1 - n) . E_j^\omega \psi_{-1}^{\tilde{\omega}}(X_n; z_n) +$$

$$+ E_j^\omega \psi_0^{\tilde{\omega}}(X_n; z_n)] + u_{0,j}^{\tilde{\omega}} - E_j^\omega u_{0,X_N}^{\tilde{\omega}} .$$

Recurrence relation (1.14) was more directly inferred in [7].

Now we shall try to find a series in the powers of $\varrho = (1 - \beta)/\beta$ for the last term on the righthand side of (1.10). Using (1.9) we can infer

1.6. Lemma. *Let $\tilde{\omega} \sim \tilde{z}(j)$ be a stationary control. Then for an arbitrary control $\omega \sim z_n$ and all $m, N = 1, 2, \ldots$ it holds*

$$\sum_{n=0}^{N+m-1} \beta^n \,.\, E_j^\omega\, u_{m-1,X_n}^{\tilde{\omega}} = -\frac{1}{\beta}\, u_{m,j}^{\tilde{\omega}} + \beta^{N+m} E_j^\omega\, u_{m,X_{N+m}}^{\tilde{\omega}} - \tag{1.15}$$

$$- \sum_{n=0}^{N+m-1} \beta^n \,.\, E_j^\omega\, \psi_m^{\tilde{\omega}}(X_n; z_n) + \frac{1-\beta}{\beta} \sum_{n=0}^{N+m} \beta^n \,.\, E_j^\omega\, u_{m,X_n}^{\tilde{\omega}} \,.$$

Proof. We can write

$$\sum_{n=0}^{N+m-1} \beta^n \,.\, E_j^\omega\, u_{m-1,X_n}^{\tilde{\omega}} = - \sum_{n=0}^{N+m-1} \beta^n \,.\, E_j^\omega\, \psi_m^{\tilde{\omega}}(X_n; z_n) + \tag{1.16}$$

$$+ \sum_{n=0}^{N+m-1} \beta^n \,.\, E_j^\omega(u_{m,X_{n+1}}^{\tilde{\omega}} - u_{m,X_n}^{\tilde{\omega}}) \,.$$

On the other hand

$$\sum_{n=0}^{N+m-1} \beta^n \,.\, E_j^\omega(u_{m,X_{n+1}}^{\tilde{\omega}} - u_{m,X_n}^{\tilde{\omega}}) = -u_{m,j}^{\tilde{\omega}} + \beta^{N+m} \,.\, E_j^\omega\, u_{m,X_{N+m}}^{\tilde{\omega}} + \tag{1.17}$$

$$+ (1-\beta) \,.\, \sum_{n=0}^{N+m-1} \beta^n \,.\, E_j^\omega\, u_{m,X_{n+1}}^{\tilde{\omega}} = -\frac{1}{\beta}\, u_{m,j}^{\tilde{\omega}} +$$

$$+ \beta^{N+m} E_j^\omega\, u_{m,X_{N+m}}^{\tilde{\omega}} + \frac{1-\beta}{\beta} \sum_{n=0}^{N+m} \beta^n \,.\, E_j^\omega\, u_{m,X_n}^{\tilde{\omega}} \,.$$

Setting (1.17) into (1.16) we obtain (1.15). □

From Lemma 1.4 and Lemma 1.6 we can derive a general form of a Laurent expansion for the expected discounted reward.

1.7. Theorem. *Let $\tilde{\omega} \sim \tilde{z}(j)$ be a stationary control. Then for an arbitrary control $\omega \sim z_n$ and all $j \in I$; $M, N = 0, 1, 2, \ldots$; $M \leqq N$ holds*

(1.18)

$$E_j^\omega\, C_N(\beta) = \left(\frac{1-\beta}{\beta}\right)^{-1} \,.\, \left[\frac{1-\beta^N}{\beta}\, g_j^{\tilde{\omega}} + \sum_{n=0}^{N-2} (\beta^n - \beta^{N-1}) \,.\, E_j^\omega\, \psi_{-1}^{\tilde{\omega}}(X_n; z_n)\right] +$$

$$+ \sum_{m=0}^{M-1} \left(\frac{1-\beta}{\beta}\right)^m \,.\, \left[\frac{1}{\beta}\, u_{m,j}^{\tilde{\omega}} - \beta^{N+m} \,.\, E_j^\omega\, u_{m,X_{N+m}}^{\tilde{\omega}} +\right.$$

$$\left. + \sum_{n=0}^{N+m-1} \beta^n \,.\, E_j^\omega\, \psi_m^{\tilde{\omega}}(X_n; z_n)\right] - \left(\frac{1-\beta}{\beta}\right)^M \sum_{n=0}^{N+M-1} \beta^n \,.\, E_j^\omega\, u_{M,X_n}^{\tilde{\omega}} \,.$$

Proof. The proof immediately follows if we set (1.15) into (1.10). □

1.8. Corollary. *For* $N \to \infty$ *from* (1.18) *we obtain*

$$E_j^{\omega}\, C(\beta) = \left(\frac{1-\beta}{\beta}\right)^{-1} \cdot \left[\frac{1}{\beta}\, g_j^{\tilde{\omega}} + \sum_{n=0}^{\infty} \beta^n \,.\, E_j^{\omega}\, \psi_{-1}^{\tilde{\omega}}(X_n;\, z_n)\right] + \tag{1.19}$$

$$+ \sum_{m=0}^{M-1} \left(\frac{1-\beta}{\beta}\right)^{m} \left[\frac{1}{\beta}\, u_{m,j}^{\tilde{\omega}} + \sum_{n=0}^{\infty} \beta^n \,.\, E_j^{\omega}\, \psi_m^{\tilde{\omega}}(X_n;\, z_n)\right] +$$

$$- \left(\frac{1-\beta}{\beta}\right)^{M} \sum_{n=0}^{\infty} \beta^n \,.\, E_j^{\omega} u_{M,X_n}^{\tilde{\omega}} \,.$$

1.9. Remark. As from (1.19)

$$E_j^{\tilde{\omega}}\, C(\beta) = \left(\frac{1-\beta}{\beta}\right)^{-1} \frac{1}{\beta}\, g_j^{\tilde{\omega}} + \sum_{m=0}^{M-1} \left(\frac{1-\beta}{\beta}\right)^{m} \cdot \frac{1}{\beta}\, u_{m,j}^{\tilde{\omega}} - \tag{1.20}$$

$$- \left(\frac{1-\beta}{\beta}\right)^{M} \sum_{n=0}^{\infty} \beta^n \,.\, E_j^{\tilde{\omega}} u_{M,X_n}^{\tilde{\omega}}$$

and

$$\left(\frac{1-\beta}{\beta}\right)^{M} \sum_{n=0}^{\infty} \beta^n \,.\, E_j^{\omega} u_{M,X_n}^{\tilde{\omega}} = o(1-\beta)^{M-2} \tag{1.20'}$$

(a function $f(\beta)$ is said to be $o(1-\beta)^L$ if $\lim_{\beta \to 1-} [f(\beta)]/(1-\beta)^L = 0)$ we can write

(1.21)

$$E_j^{\omega}\, C(\beta) = E_j^{\tilde{\omega}}\, C(\beta) + \sum_{n=0}^{\infty} \beta^n \left[\sum_{m=-1}^{M-1} \left(\frac{1-\beta}{\beta}\right)^{m} \cdot E_j^{\omega}\, \psi_m^{\tilde{\omega}}(X_n;\, z_n)\right] + o(1-\beta)^{M-2} \,.$$

Note that in (1.19), (1.21) the interchange of $\sum_{n=0}^{\infty} \sum_{m=-1}^{M-1}$ is justified as all $E_j^{\omega}\, \psi_m^{\tilde{\omega}}(X_n;\, z_n)$ are uniformly bounded and

$$\sum_{n=0}^{\infty} \sum_{m=-1}^{M-1} \beta^n \left(\frac{1-\beta}{\beta}\right)^{m} < \infty$$

for any $\beta \in (0, 1)$.

1.10. Remark. Setting $E_j^{\omega}\, C(\beta) = v(j;\, \beta;\, \omega)$ and introducing a (column) vector $\mathbf{v}(\beta;\, \omega) = \|v(j;\, \beta;\, \omega)\|_{j=1}^{r}$ from (1.20) we can write for a stationary control $\tilde{\omega} \sim \tilde{z}(j)$

$$\mathbf{v}(\beta;\, \tilde{\omega}) = \left(\frac{1-\beta}{\beta}\right)^{-1} \cdot \frac{1}{\beta}\, \mathbf{g}^{\tilde{\omega}} + \sum_{m=0}^{M-1} \left(\frac{1-\beta}{\beta}\right)^{m} \cdot \frac{1}{\beta}\, \mathbf{u}_m^{\tilde{\omega}} + \tag{1.22}$$

$$+ \left(\frac{1-\beta}{\beta}\right)^{M} \sum_{n=0}^{\infty} (\beta \mathbf{P}^{\tilde{\omega}})^n \,.\, \mathbf{u}_M^{\tilde{\omega}} \,.$$

Setting from (1.4′), (1.6) into (1.22) and introducing the interest rate $\varrho = (1 - \beta)/\beta$ instead of the discount factor β we obtain for $\varrho < |\mathbf{Z}^{\tilde{\omega}}|^{-1}$ (where $|\mathbf{Z}|$ is the norm of $\mathbf{Z} = \|z_{ij}\|_{i,j=1}^{r}$ defined as $|\mathbf{Z}| = \max_{i \in I} \sum_{j \in I} |z_{ij}|$) letting $M \to \infty$

$$(1.23) \quad \mathbf{v}(\beta; \tilde{\omega}) = (\varrho + 1) \,.\, [\varrho^{-1} \,.\, \boldsymbol{\Pi}^{\tilde{\omega}} \,.\, \mathbf{c}^{\tilde{\omega}} - \sum_{m=0}^{\infty} \varrho^m (-\mathbf{Z}^{\tilde{\omega}})^{m+1} \,.\, (\mathbf{I} - \boldsymbol{\Pi}^{\tilde{\omega}}) \,.\, \mathbf{c}^{\tilde{\omega}}] \,.$$

As $(\mathbf{I} - \boldsymbol{\Pi})^m = \mathbf{I} - \boldsymbol{\Pi}$ (1.23) can be also written as

$$(1.23') \qquad \mathbf{v}(\beta; \tilde{\omega}) = (\varrho + 1) \,.\, [\varrho^{-1} \,.\, \boldsymbol{\Pi}^{\tilde{\omega}} \,.\, \mathbf{c}^{\tilde{\omega}} - \sum_{m=0}^{\infty} \varrho^m (-\mathbf{H}^{\tilde{\omega}})^{m+1} \, \mathbf{c}^{\tilde{\omega}}]$$

where

$$\mathbf{H}^{\tilde{\omega}} = \mathbf{Z}^{\tilde{\omega}}(\mathbf{I} - \boldsymbol{\Pi}^{\tilde{\omega}}) \,.$$

(1.23′) is the Laurent expansion for a stationary control found by Miller and Veinott in [6].

2. NECESSARY AND SUFFICIENT l-DISCOUNT OPTIMALITY CONDITIONS

In this paragraph the recurrence formula for expected discounted reward inferred in Theorem 1.7 will be employed for establishing necessary and sufficient optimality conditions if l-discount optimality criteria are considered.

First, we shall investigate the optimality conditions in the class of stationary controls. We shall prove the following lemma.

2.1. Lemma. *There exists a stationary control $\omega^* \sim z^*(j)$ such that for all $j \in I$*

$$(2.1) \qquad \psi_{-1}^{\omega^*}(j; z) \leqq 0 \quad \textit{for any} \quad z \in J$$

and

$$(2.1') \qquad \psi_m^{\omega^*}(j; z) \leqq 0 \quad \textit{for any} \quad z \in J_{m-1}^{\omega^*}(j) \,; \quad m = 0, 1, 2, \ldots$$

Proof. Setting from (1.5) into (1.9) we can write for $m = 1, 2, \ldots$ and any stationary control $\omega \sim z(j)$

$$(2.2) \qquad \psi_m^{\omega}(j; z) = \sum_{k \in I} [p(j, k; z) - p(j, k; z(j)) + \pi(j, k; \omega)] \,.\, u_{m,k}^{\omega} \,.$$

As the vectors $\mathbf{u}_m^{\omega}$ for $m > r$ are linear combinations of $\mathbf{u}_1^{\omega}, \ldots, \mathbf{u}_r^{\omega}$ (compare Remark 1.3) if $\psi_m^{\omega}(j; z) = 0$ for $m = 1, 2, \ldots, r$ then also $\psi_m^{\omega}(j; z) = 0$ for any $m > r$. So $z \in J_r^{\omega} \Rightarrow z \in J_m^{\omega}$ and (as $J_{m-1}^{\omega} \supset J_m^{\omega}$) $J_m^{\omega} \equiv J_r^{\omega}$ for any $m > r$.

Thus it is sufficient to show the existence of a stationary control $\omega^* \sim z^*(j)$ for which (2.1), (2.1′) will hold for $m = 0, 1, \ldots, r$. The proof can be performed by control (policy) iteration method.

Let ω_1 be an arbitrary stationary control. If conditions (2.1), (2.1′) for $m = 0, 1, \ldots, r$ are not fulfilled for all $j \in I$ let us construct a stationary control $\omega_2 \sim z'(j)$ in such a way that for any $j \in I$ the first non-zero element (if exists) of the sequence $\{\psi_{-1}^{\omega_1}(j; z'(j)), \psi_0^{\omega_1}(j; z'(j)), \ldots, \psi_r^{\omega_1}(j; z'(j))\}$ is positive. So for β sufficiently near 1 and any $j \in I$

$$\sum_{m=-1}^{r+1} \left(\frac{1-\beta}{\beta}\right)^m \psi_m^{\omega_1}(j; z'(j)) \geqq 0$$

and a strict inequality holds at least for one $j \in I$. Setting ω_2, ω_1 into (1.21) we have for $M = r + 2$

$$(2.3) \quad E_j^{\omega_2} C(\beta) - E_j^{\omega_1} C(\beta) = \sum_{n=0}^{\infty} \beta^n \left[\sum_{m=-1}^{r+1} \left(\frac{1-\beta}{\beta}\right)^m E_j^{\omega_2} \psi_m^{\omega_1}(X_n; z'(X_n))\right] + $$
$$+ o(1-\beta)^r \geqq \sum_{m=-1}^{r+1} \left(\frac{1-\beta}{\beta}\right)^m \psi_m^{\omega_1}(j; z'(j)) + o(1-\beta)^r .$$

From (2.3) for β sufficiently near 1 $E_j^{\omega_2} C(\beta) - E_j^{\omega_1} C(\beta) \geqq o(1-\beta)^r$ and a strict inequality is fulfilled at least for one $j \in I$. Using the above procedure we can construct a sequence of stationary controls $\{\omega_1, \omega_2, \ldots\}$ the terms of which cannot recur. As there exists only a finite number of stationary controls repeating the above procedure a stationary control ω^* must be found. □

Now we are in a position to formulate necessary and sufficient l-discount optimality conditions.

2.2. Theorem. *A control $\omega_l \sim z_{l,n}$ is l-discount optimal (for $l = -1, 0, 1, \ldots$) if and only if for all $j \in I$ and any $n = 0, 1, \ldots$ the following conditions are fulfilled*

$$(2.4) \qquad E_j^{\omega_l} \psi_m^{\omega^*}(X_n; z_{l,n}) = 0 \quad \textit{for} \quad m = -1, 0, \ldots, l$$

and

$$(2.5) \qquad \lim_{\beta \to 1-} \left(\frac{1-\beta}{\beta}\right) \sum_{n=0}^{\infty} \beta^n \, . \, E_j^{\omega_l} \psi_{l+1}^{\omega^*}(X_n; z_{l,n}) = 0 \, .$$

Proof. For any control $\omega \sim z_n$ and any $M = -1, 0, \ldots$ it holds (compare (1.21))

$$(2.6) \qquad E_j^{\omega} C(\beta) - E_j^{\omega^*} C(\beta) =$$
$$= \sum_{n=0}^{\infty} \beta^n \left[\sum_{m=-1}^{M} \left(\frac{1-\beta}{\beta}\right)^m E_j^{\omega} \psi_m^{\omega^*}(X_n; z_n)\right] + o(1-\beta)^{M-1} .$$

Let us suppose that

$$(2.7)\qquad E_j^{\omega}\,\psi_m^{\omega^*}(X_n;\,z_n) = 0 \quad \text{for any} \quad m < m_0 \quad \text{and all} \quad n = 0, 1, \ldots$$

and

$$(2.7')\qquad E_j^{\omega}\,\psi_{m_0}^{\omega^*}(X_n;\,z_n) < 0 \quad \text{for certain} \quad n = n_0\,.$$

(Note that under conditions (2.7) it cannot happen $E_j^{\omega}\,\psi_{m_0}^{\omega^*}(X_n;\,z_n) > 0$ and that there exists $h_{m_0} < 0$ such that $h_{m_0} \,.\, E_j^{\omega}\,\psi_{m_0+1}^{\omega^*}(X_n;\,z_n) \geqq E_j^{\omega}\,\psi_{m_0}^{\omega^*}(X_n;\,z_n)$ for any n.) If (2.7), (2.7′) hold then from (2.6) we obtain for any $1 > \beta > h_{m_0}/(1 + h_{m_0})$

$$(2.8)\qquad E_j^{\omega}\,C(\beta) - E_j^{\omega^*}\,C(\beta) = \sum_{n=0}^{\infty} \beta^n \left[\sum_{m=m_0}^{m_0+1} \left(\frac{1-\beta}{\beta}\right)^m E_j^{\omega}\,\psi_m^{\omega^*}(X_n;\,z_n)\right] +$$

$$+\, o(1-\beta)^{m_0} < \beta^{n_0} \left(\frac{1-\beta}{\beta}\right)^{m_0} E_j^{\omega}\,\psi_{m_0}^{\omega^*}(X_{n_0};\,z_{n_0}) + o(1-\beta)^{m_0}\,.$$

1. Necessity. Let condition (2.4) not hold for certain $m = m_0 \leqq l$, $n = n_0$ and let m_0, n_0 be chosen such that conditions (2.7), (2.7′) are fulfilled (we set $\omega \equiv \omega_l$). Then from (2.8) we obtain

$$(2.9)\qquad \liminf_{\beta\to 1-} \left(\frac{1-\beta}{\beta}\right)^{-m_0} [E_j^{\omega_l}\,C(\beta) - E_j^{\omega^*}\,C(\beta)] \leqq E_j^{\omega_l}\,\psi_{m_0}^{\omega^*}(X_{n_0};\,z_{l,n_0}) < 0$$

and ω_l cannot be m_0-discount optimal (and even l-discount optimal). So condition (2.4) must be fulfilled for any l-discount optimal control.

Let condition (2.4) be fulfilled but (2.5) let not hold. Then (as under condition (2.4) $E_j^{\omega}\,\psi_{l+1}^{\omega^*}(X_n;\,z_{l,n}) \leqq 0$)

$$(2.10)\qquad \liminf_{\beta\to 1-} \left(\frac{1-\beta}{\beta}\right) \sum_{n=0}^{\infty} \beta^n \,.\, E_j^{\omega_l}\,\psi_{l+1}^{\omega^*}(X_n;\,z_{l,n}) < 0$$

and from equation (2.6) for $\omega \equiv \omega_l$, $M = l + 1$ we obtain

$$(2.11)\qquad \left(\frac{1-\beta}{\beta}\right)^{-l} [E_j^{\omega_l}\,C(\beta) - E_j^{\omega^*}\,C(\beta)] =$$

$$= \left(\frac{1-\beta}{\beta}\right) \sum_{n=0}^{\infty} \beta^n \,.\, E_j^{\omega_l}\,\psi_{l+1}^{\omega^*}(X_n;\,z_{l,n}) + \left(\frac{1-\beta}{\beta}\right)^{-l} o(1-\beta)^l\,.$$

Setting (2.10) into (2.11) we obtain

$$(2.12)\qquad \liminf_{\beta\to 1-} \left(\frac{1-\beta}{\beta}\right)^{-l} [E_j^{\omega_l}\,C(\beta) - E_j^{\omega^*}\,C(\beta)] < 0$$

and the control ω_l cannot be l-discount optimal. So conditions (2.4) and (2.5) must be fulfilled for any l-discount optimal control.

2. Sufficiency. As for any control $\omega \sim z_n$, any $n = 0, 1, \ldots$ and all $j \in I$ the first non-zero element (if exists) of a finite sequence

$$\left\{\left(\frac{1-\beta}{\beta}\right)^{-1} E_j^{\omega}\, \psi_{-1}^{\omega^*}(X_n;\, z_n);\; E_j^{\omega}\, \psi_0^{\omega^*}(X_n;\, z_n);\; \left(\frac{1-\beta}{\beta}\right) E_j^{\omega}\, \psi_1^{\omega^*}(X_n;\, z_n);\right.$$
$$\left.\ldots;\; \left(\frac{1-\beta}{\beta}\right)^{l+1} E_j^{\omega}\, \psi_{l+1}^{\omega^*}(X_n;\, z_n)\right\}$$

must be negative using the expansion (2.6) for $E_j^{\omega^*} C(\beta) - E_j^{\omega} C(\beta)$, $M = l + 1$ we can easily see that

$$\liminf_{\beta \to 1-} \left(\frac{1-\beta}{\beta}\right)^{-l} \left[E_j^{\omega^*} C(\beta) - E_j^{\omega} C(\beta)\right] \geqq 0\,. \tag{2.13}$$

Using the expansion (2.6) for $E_j^{\omega_l} C(\beta) - E_j^{\omega^*} C(\beta)$, $M = l + 1$ under conditions (2.4), (2.5) we obtain

$$\lim_{\beta \to 1-} \left(\frac{1-\beta}{\beta}\right)^{-l} \left[E_j^{\omega_l} C(\beta) - E_j^{\omega^*} C(\beta)\right] = 0\,. \tag{2.14}$$

From (2.13), (2.14) immediately follows

$$\liminf_{\beta \to 1-} \left(\frac{1-\beta}{\beta}\right)^{-l} \left[E_j^{\omega_l} C(\beta) - E_j^{\omega} C(\beta)\right] \geqq 0 \tag{2.15}$$

for any $j \in I$ and any control $\omega \sim z_n$.

So any control ω_l fulfilling conditions (2.4) and (2.5) must be l-discount optimal. □

2.3. Corollary. *Let condition* (2.4) *hold. Then* (2.5) *is fulfilled if and only if*

$$\lim_{N \to \infty} \frac{1}{N+1} \sum_{n=0}^{N} E_j^{\omega_l}\, \psi_{l+1}^{\omega^*}(X_n;\, z_{l,n}) = 0\,. \tag{2.5'}$$

Proof. As under condition (2.4) $E_j^{\omega_l}\, \psi_{l+1}^{\omega^*}(X_n;\, z_{l,n}) \leqq 0$ for all $j \in I$ and any $n = 0, 1, \ldots$ then in virtue of Hardy-Littlewood theorem (compare [10] Theorem 97)

$$\lim_{N \to \infty} \frac{1}{N+1} \sum_{n=0}^{N} E_j^{\omega_l}\, \psi_{l+1}^{\omega^*}(X_n;\, z_{l,n}) = \tag{2.16}$$
$$= \lim_{\beta \to 1-} (1-\beta) \sum_{n=0}^{\infty} \beta^n \,.\, E_j^{\omega_l}\, \psi_{l+1}^{\omega^*}(X_n;\, z_{l,n})$$

if at least one of the limit exists.

So (2.5) (resp. (2.5′)) together with (2.16) immediately imply (2.5′) (resp. (2.5)). □

In case that $l \geqq r$ for l-discount optimal controls the following theorem is valid.

2.4. Theorem. *A control* $\omega \sim z_n$ *is r-discount optimal if* (2.4) *is fulfilled for* $m = -1, 0, 1, \ldots, r$. *For* $l > r$ *a control* ω_l *is l-discount optimal if and only if it is r-discount optimal.*

Proof. In Theorem 2.2 it was established that condition (2.4) for $m = -1, 0, 1, \ldots, r$ must be fulfilled for any r-discount optimal control.

As for $m > r$ the vectors $\boldsymbol{u}_m^{\omega^*}$ are linear combinations of $\boldsymbol{u}_1^{\omega^*}, \ldots, \boldsymbol{u}_r^{\omega^*}$ and (compare the proof of Lemma 2.1) if $\psi_m^{\omega^*}(j; z) = 0$ for $m = 1,2, \ldots, r$ then also $\psi_m^{\omega^*}(j; z) = 0$ for any $m > r$ if condition (2.4) is fulfilled for $m = -1, 0, \ldots, r$ then for any $l > r$ and any $n = 0, 1, \ldots$ also $E_j^\omega\, \psi_l^{\omega^*}(X_n; z_n) = 0$ (and, of course,

$$\left.\lim_{N\to\infty} \frac{1}{N+1} \sum_{n=0}^{N} E_j^\omega\, \psi_j^{\omega^*}(X_n; z_n) = 0\right). \qquad \square$$

From Theorem 2.2 and Theorem 2.4 we obtain an important corollary concerning the existence of a stationary l-discount optimal control.

2.5. Corollary. Any stationary control ω_l^* (for $l = -1, 0, 1, \ldots$) fulfilling conditions (2.1), (2.1′) for $m = l + 1$ is l-discount optimal. If a stationary control ω_r^* is r-discount optimal then it is also l-discount optimal for $l > r$.

3. EQUIVALENCE BETWEEN DISCOUNTED AND AVERAGE REWARD OPTIMALITY CRITERIA

In this paragraph we shall investigate average reward type optimality criteria corresponding to the concept of l-discount optimal control.

First, using the results of Corollary 1.5 and Lemma 1.6 (compare (1.14), (1.15)) we shall infer a recurrence formula for the difference $v_l^{(N)}(j; \omega) - v_l^{(N)}(j; \omega^*)$ where $\omega \sim z_n$ is an arbitrary control and the stationary control $\omega^* \sim z^*(j)$ fulfils conditions (2.1) and (2.1′) (the values $v_l^{(N)}(j; \omega)$ for $l = 0, 1, \ldots$ are given by a recurrence relation $v_{l+1}^{(N)}(j; \omega) = \sum_{n=0}^{N-1} v_l^{(n)}(j; \omega)$ where $v_0^{(N)}(j; \omega) = E_j^\omega\, C_N$ and $v_0^{(0)}(j; \omega) = 0$). It holds

3.1. Lemma. *For any* $l = 0, 1, \ldots; N = 0, 1, 2, \ldots$

$$v_l^{(N)}(j; \omega) - v_l^{(N)}(j; \omega^*) = \Phi_l^{(N)}(j; \omega) + E_j^{\omega^*} u_{l,X_N}^{\omega^*} - E_j^\omega u_{l,X_N}^{\omega^*} \tag{3.1}$$

where

$$\Phi_0^{(N)}(j; \omega) = \sum_{n=0}^{N-1} [(N - 1 - n) \,.\, E_j^\omega\, \psi_{-1}^{\omega^*}(X_n; z_n) + E_j^\omega\, \psi_0^{\omega^*}(X_n; z_n)] \tag{3.2}$$

and

$$\Phi_{l+1}^{(N)}(j; \omega) = \sum_{n=0}^{N-1} [\Phi_l^{(n)}(j; \omega) + E_j^\omega\, \psi_{l+1}^{\omega^*}(X_n; z_n)]\,. \tag{3.2′}$$

Proof. By induction with respect to l. For $l = 0$ (3.1) holds (compare (1.14)). Let (3.1) hold for certain $l \geqq 0$. Summing (3.1) for $N = 0, 1, \ldots, M - 1$ we obtain

$$(3.3) \qquad v_{l+1}^{(M)}(j; \omega) - v_{l+1}^{(M)}(j; \omega^*) = \sum_{N=0}^{M-1} [v_l^{(N)}(j; \omega) - v_l^{(N)}(j; \omega^*)] =$$

$$= \sum_{N=0}^{M-1} \Phi_l^{(N)}(j; \omega) + \sum_{N=0}^{M-1} (E_j^{\omega^*} u_{l,X_N}^{\omega^*} - E_j^{\omega} u_{l,X_N}^{\omega^*}) .$$

But from Lemma 1.6 (compare (1.15)) we have

$$(3.4) \qquad \sum_{N=0}^{M-1} (E_j^{\omega^*} u_{l,X_N}^{\omega^*} - E_j^{\omega} u_{l,X_N}^{\omega^*}) =$$

$$= \sum_{N=0}^{M-1} E_j^{\omega} \psi_{l+1}^{\omega^*}(X_N; z_N) + E_j^{\omega^*} u_{l+1,X_M}^{\omega^*} - E_j^{\omega} u_{l+1,X_M}^{\omega^*} .$$

Setting (3.4) into (3.3) and using (3.2′) we immediately obtain

$$v_{l+1}^{(M)}(j; \omega) - v_{l+1}^{(M)}(j; \omega^*) = \Phi_{l+1}^{(M)}(j; \omega) + E_j^{\omega^*} u_{l+1,X_M}^{\omega^*} - E_j^{\omega} u_{l+1,X_M}^{\omega^*} . \qquad \square$$

For the investigation of the optimality properties of l-order average reward optimal controls the following lemma will be very useful.

3.2. Lemma. *Let for a given parameter* $m = -1, 0, 1, \ldots$ $\{a_m(n), n = 0, 1, \ldots,\}$ (*where* $a_m(n) \leqq H_m < \infty$) *be a given sequence with the following properties*:

For given integers $m_0 \geqq -1$, $n_0 \geqq 0$ *and real* $h < 0$

i) $a_m(n) = 0$ *for all* $m = -1, 0, \ldots, m_0 - 1$ *and any* $n = 0, 1, \ldots$;

ii) $a_{m_0}(n) = 0$ *for any* $n < n_0$; $a_{m_0}(n_0) < 0$ *and* $a_{m_0}(n) \leqq 0$ *for any* $n > n_0$;

iii) $a_{m_0}(n) \leqq h \cdot a_{m_0+1}(n)$ *for any* n.

Let us denote for $m = 0, 1, \ldots$

$$(3.5) \qquad A_{m+1}(N) = \sum_{n=0}^{N-1} [A_m(n) + a_{m+1}(n)]$$

where

$$(3.5') \qquad A_0(N) = \sum_{n=0}^{N-1} [(N - 1 - n)\, a_{-1}(n) + a_0(n)] .$$

Then it holds

$$(3.6) \qquad \limsup_{N \to \infty} \frac{1}{N - m_0 - 1} A_{m_0+1}(N) \leqq a_{m_0}(n_0) < 0$$

and

$$(3.7) \qquad \limsup_{N \to \infty} \frac{1}{N - m} A_m(N) = -\infty \quad \textit{for any} \quad m > m_0 + 1 .$$

Proof. As

$$A_{m_0+1}(N) = \sum_{M=0}^{N-1} \Big[\sum_{n=0}^{M-1} a_{m_0}(n) + a_{m_0+1}(M) \Big] = \sum_{n=0}^{N-1} \big[(N-1-n) \, . \, a_{m_0}(n) + a_{m_0+1}(n) \big]$$

from i), ii), iii) we obtain

$$\begin{aligned} (3.8) \qquad A_{m_0+1}(N) &\leqq (N-1-n_0) \, . \, a_{m_0}(n_0) + a_{m_0+1}(n_0) + \\ &+ \sum_{n=n_0+1}^{N-1} \big[(N-1-n) \, . \, a_{m_0}(n) + a_{m_0+1}(n) \big] \cdot \chi_{m_0+1}(n) \leqq \\ &\leqq (N-1-n_0) \, . \, a_{m_0}(n_0) + \\ &+ a_{m_0+1}(n_0) + \sum_{n=n_0+1}^{N-1} \big[(N-1-n) \, . \, h + 1 \big] \, . \, a_{m_0+1}(n) \cdot \chi_{m_0+1}(n) \end{aligned}$$

where $\chi_m(n) = 0$ (resp. $= 1$) if $a_m(n) \leqq 0$ (resp. $a_m(n) > 0$). As for any $n \leqq N - 1 + h^{-1} = n'$, $\big[(N-1-n) \, . \, h + 1\big] \, . \, a_{m_0+1}(n) \cdot \chi_{m_0+1}(n) \leqq 0$

$$(3.9) \qquad \limsup_{N \to \infty} \frac{1}{N} \sum_{n=n_0+1}^{N-1} \big[(N-1-n) \, . \, h + 1 \big] \, . \, a_{m_0+1}(n) \cdot \chi_{m_0+1}(n) \leqq 0 \, .$$

From (3.8) and (3.9) immediately follows (3.6).

In virtue of (3.6) and the boundeness of $a_m(n)$ $\limsup_{n \to \infty} \big[A_{m_0+1}(n) + a_{m_0+2}(n)\big] = - \infty$. Then from (3.5) immediately follows (3.7) for $m = m_0 + 2$.

Repeating the above procedure we can establish (3.7) for any $m > m_0 + 1$. □

3.3. Remark. As the set J of control parameter values is finite if for an arbitrary control $\omega \sim z_n$ $E_j^\omega \, \psi_m^{\omega^*}(X_n; z_n) = 0$ for $m = -1, 0, \ldots, m_0 - 1$; $n = 0, 1, \ldots$ and $E_j^\omega \, \psi_{m_0}^{\omega^*}(X_{n_0}; z_{n_0}) < 0$ then there exists $h_{m_0} < 0$ such that for any $n = 0, 1, \ldots$ $h_{m_0} \, . \, E_j^\omega \, \psi_{m_0+1}^{\omega^*}(X_n; z_n) \geqq E_j^\omega \, \psi_{m_0}^{\omega^*}(X_n; z_n)$. Obviously, $E_j^\omega \, \psi_m^{\omega^*}(X_n; z_n) < H_m < \infty$. Setting $a_m(n) = E_j^\omega \, \psi_m^{\omega^*}(X_n; z_n)$ and choosing n_0 such that $E_j^\omega \, \psi_{m_0}^{\omega^*}(X_n; z_n) = 0$ for any $n < n_0$, $E_j^\omega \, \psi_{m_0}^{\omega^*}(X_{n_0}; z_{n_0}) < 0$ it can be easily seen that the values $a_m(n)$ satisfy the assumptions of Lemma 3.2.

The optimality properties of the control ω^* (fulfilling conditions (2.1), (2.1′)) are contained in the next lemma.

3.4. Lemma. *The control $\omega^* \sim z^*(j)$ is l-order average reward optimal for $l = 0, 1, \ldots$.*

Proof. In virtue of the recurrence formula obtained in Lemma 3.1 (compare (3.1)) we have for any control $\omega \sim z_n$

$$(3.10) \qquad \liminf_{N \to \infty} \frac{1}{N - l} \big[v_l^{(N)}(j; \omega^*) - v_l^{(N)}(j; \omega) \big] =$$

$$= - \limsup_{N \to \infty} \frac{1}{N - l} \left[v_l^{(N)}(j; \omega) - v_l^{(N)}(j; \omega^*) \right] = - \limsup_{N \to \infty} \frac{1}{N - l} \Phi_l^{(N)}(j; \omega) .$$

As the values $a_m(n) = E_j^{\omega} \psi_m^{\omega^*}(X_n; z_n)$ satisfy the assumptions of Lemma 3.2 (compare Remark 3.3) using the results of Lemma 3.2 (compare (3.5) and (3.2'))

$$\limsup_{N \to \infty} \frac{1}{N - l} \Phi_l^{(N)}(j; \omega) \leqq 0 .$$

So for any control $\omega \sim z_n$ and all $j \in I$

$$(3.11) \qquad \liminf_{N \to \infty} \frac{1}{N - l} \left[v_l^{(N)}(j; \omega^*) - v_l^{(N)}(j; \omega) \right] \geqq 0$$

and the control ω^* must be l-order average reward optimal for any l. □

Now we are in a position to formulate the main result of this paragraph.

3.5. Theorem. *A control $\omega_l \sim z_{l,n}$ is l-order average reward optimal (for $l = 0, 1, \ldots$) if and only if the following conditions are fulfilled for any $m = -1, 0, 1, \ldots, l - 1$ and any $n = 0, 1, \ldots$*

$$(3.12) \qquad E_j^{\omega_l} \psi_m^{\omega^*}(X_n; z_{l,n}) = 0$$

and

$$(3.13) \qquad \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} E_j^{\omega_l} \psi_l^{\omega^*}(X_n; z_{l,n}) = 0 .$$

Proof. Note (compare Remark 3.3) that for any control $\omega \sim z_n$ and all $j \in I$ $a_m(n) = E_j^{\omega} \psi_m^{\omega^*}(X_n; z_n)$ satisfy the assumptions of Lemma 3.2.

1. Necessity. Let condition (3.12) not hold for certain $m = m_0$ and $n = n_0$ and let m_0, n_0 be chosen such that $E_j^{\omega_l} \psi_{m_0}^{\omega^*}(X_n; z_{l,n}) = 0$ for any $n < n_0$ and $E_j^{\omega_l} \psi_m^{\omega^*}(X_n; z_{l,n}) = 0$ for any n and $m < m_0$. Using the recurrence formula inferred in Lemma 3.1 we have

$$(3.14) \qquad v_m^{(N)}(j; \omega_l) - v_m^{(N)}(j; \omega^*) = E_j^{\omega^*} u_{m,X_N}^{\omega^*} - E_j^{\omega_l} u_{m,X_N}^{\omega^*} + \Phi_m^{(N)}(j; \omega_l) .$$

From Lemma 3.2 immediately follows for any $m > m_0$

$$(3.14') \qquad \limsup_{N \to \infty} \frac{1}{N - m} \Phi_m^{(N)}(j; \omega_l) \leqq E_j^{\omega_l} \psi_{m_0}^{\omega^*}(X_{n_0}; z_{l,n_0}) < 0 .$$

So from (3.14), (3.14')

$$(3.15) \liminf_{N \to \infty} \frac{1}{N - l} \left[v_l^{(N)}(j; \omega_l) - v_l^{(N)}(j; \omega^*) \right] \leqq \limsup_{N \to \infty} \frac{1}{N - l} \Phi_l^{(N)}(j; \omega_l) < 0$$

and the control ω_l cannot be l-order average reward optimal. Under condition (3.12) recurrence formula (3.1) reads

$$(3.16)\ v_l^{(N)}(j;\omega_l) - v_l^{(N)}(j;\omega^*) = E_j^{\omega^*} u_{l,X_N}^{\omega^*} - E_j^{\omega_l} u_{l,X_N}^{\omega^*} + \sum_{n=0}^{N-1} E_j^{\omega_l}\,\psi_l^{\omega^*}(X_n; z_{l,n})\,.$$

As under condition (3.12) $E_j^{\omega_l}\,\psi_l^{\omega^*}(X_n; z_{l,n}) \leqq 0$

$$\limsup_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} E_j^{\omega_l}\,\psi_l^{\omega^*}(X_n; z_{l,n}) \leqq 0$$

and if (3.13) does not hold then from (3.16)

$$(3.16') \qquad \liminf_{N\to\infty} \frac{1}{N-l}\,[v_l^{(N)}(j;\omega_l) - v_l^{(N)}(j;\omega^*)] =$$

$$= \liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} E_j^{\omega_l}\,\psi_l^{\omega^*}(X_n; z_{l,n}) < 0$$

and the control ω_l cannot be l-order average reward optimal.

2. Sufficiency. Let conditions (3.12), (3.13) be fulfilled. Then from (3.1), (3.16) we obtain

$$(3.17) \qquad \lim_{N\to\infty} \frac{1}{N-l}\,[v_l^{(N)}(j;\omega_l) - v_l^{(N)}(j;\omega^*)] = 0\,.$$

But in Lemma 3.4 we have shown that ω^* is l-order average reward optimal. So for an arbitrary control $\omega \sim z_n$ and any $j \in I$, $l = 0, 1, \ldots$

$$(3.18) \qquad \liminf_{N\to\infty} \frac{1}{N-l}\,[v_l^{(N)}(j;\omega^*) - v_l^{(N)}(j;\omega)] \geqq 0\,.$$

Combining (3.17), (3.18) we obtain

$$(3.19) \qquad \liminf_{N\to\infty} \frac{1}{N-l}\,[v_l^{(N)}(j;\omega_l) - v_l^{(N)}(j;\omega)] \geqq 0\,.$$

So any control ω_l fulfilling conditions (3.12), (3.13) must be l-order average reward optimal. □

Comparing the necessary and sufficient optimality conditions for l-discount optimal and l-order average reward optimal controls we can derive the following theorem.

3.6. Theorem. *A control ω_l is l-order average reward optimal (for $l = 0, 1, \ldots$) if and only if it is $(l - 1)$-discount optimal.*

Proof. The proof trivially follows from the results of Theorem 2.2 and Theorem 3.5 and Corollary 2.3 (using Corollary 2.3 the necessary and sufficient optimality conditions are the same in both of the Theorems). □

3.7. Remark. A partial form of Theorem 3.5 (for $l = 1$) was obtained by Lippman [4] by a slightly different approach. Lippman's approach uses well known Abelian theorems and rests on the existence of 1-order average reward optimal control (called average overtaking optimal policy) established in [2].

(Received February 6, 1974.)

REFERENCES

[1] Blackwell, D.: Discrete Dynamic Programming. Annals Mathem. Statistics *33* (1962), 2, 719—726.

[2] Denardo, E. V., Miller, B. L.: An Optimality Condition for Discrete Dynamic Programming with No Discounting. Annals Mathem. Statistics *39* (1968), 4, 1220—1227.

[3] Lippman, S. A.: On the Set of Optimal Policies in Discrete Dynamic Programming. Journal Mathem. Analysis Applic. *24* (1968), 2, 440—445.

[4] Lippman, S. A.: Criterion Equivalence in Discrete Dynamic Programming. Operations Research *17* (1969), 5, 920—923.

[5] Mandl, P.: On the Variance in Controlled Markov Chains. Kybernetika *7* (1971), 1, 1—12.

[6] Miller, B. L., Veinott, A. F.: Discrete Dynamic Programming with a Small Interest Rate. Annals Mathem. Statistics *40* (1969), 2, 366—370.

[7] Sladký, K.: Necessary and Sufficient Optimality Conditions for Average Reward of Controlled Markov Chains. Kybernetika *9* (1973), 2, 124—137.

[8] Veinott, A. F.: On Finding Optimal Policies in Discrete Dynamic Programming with No Discounting. Annals Mathem. Statistics *37* (1966), 5, 1284—1294.

[9] Veinott, A. F.: Discrete Dynamic Programming with Sensitive Discount Optimality Criteria. Annals Mathem. Statistics *40* (1969), 5, 1635—1660.

[10] Hardy, G. H.: Divergent Series. Oxford Univ. Press, Oxford 1949.

Ing. Karel Sladký, CSc.; Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation — Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 180 76 Praha 8. Czechoslovakia.