# ON THE ITERATIVE METHOD OF DYNAMIC PROGRAMMING ON A FINITE SPACE DISCRETE TIME MARKOV PROCESS[1]

By Barry W. Brown

*University of Chicago*

**1. Introduction and summary.** We consider a system with a finite number of states, 1, 2, $\cdots$, $S$. Periodically we observe the current state of the system and perform an action, $a$, from a finite set $A$ of possible actions. As a joint result of $s$, the current state, and $a$, the action performed, two things occur: (1) we receive an immediate return $r(s, a)$; and (2) the system moves to a new state $s'$ with probability $q(s' \mid s, a)$. (For several interesting and occasionally amusing concrete examples of this setup, the reader is referred to Howard's excellent book, [4].)

Let $F$ be the (finite) set of functions from $S$ to $A$. A policy $\pi$ is a sequence $(\cdots, f_n, \cdots, f_1)$ of members of $F$. Using $n$ steps of the policy $\pi$ means observing the system, and upon finding $s_0$, performing action $f_n(s_0)$, observing the next state $s_1$ and performing $f_{n-1}(s_1)$ and so on until after $n - 1$ of these steps have been completed one observes $s_{n-1}$ and performs action $f_1(s_{n-1})$ and then stops. The expected (total) return using $n$ steps of $\pi$ given that the initial state is $s_0$ is denoted $v_n(s_0, \pi)$.

A policy $\pi$ is optimal if $v_n(s, \pi) \geqq v_n(s; \pi')$ for any policy $\pi'$ and all $n$ and $s$. In other words, $\pi$ is optimal if the return using $n$ steps of $\pi$ cannot be exceeded by using $n$ steps of any other policy regardless of $n$ and the initial state of the system.

Obviously, $v_n(s, \pi)$ for optimal $\pi$ may be calculated by value iteration, that is, by the use of the recursion

$$v_{n+1}(s, \pi) = \max_a [r(s, a) + \textstyle\sum_{s'} q(s' \mid s, a) v_n(s', \pi)].$$

Similarly an optimal $\pi$ may be generated by letting $f_1(s)$ be any $a$ with $r(s, a) \geqq r(s, a')$ for all $a'$, and letting $f_{n+1}(s)$ be any $a$ for which the expression on the right side of the above equation attains its maximum.

This paper is concerned with the return of an optimal policy $\pi$. The principal results are as follows:

A policy is eventually stationary if for $m$, $n$ sufficiently large, $f_m = f_n$, that is, if the sequence which is the policy consists of one member of $F$ repeated infinitely often followed by a finite arbitrary sequence. In Section 3, an example is given where there is only one optimal policy and it has the form $(\cdots, f, g, f, g, f, g)$, an

Received 5 January 1965; revised 29 March 1965.

[1] Portions of this work were submitted as a partial fulfillment of requirements for the Ph.D. degree in mathematics at the University of California. During the period when these portions were written, the author held a National Science Foundation Fellowship. The investigation also was partially supported under USPHS Grant FR00013 from the Division of Research Facilities and Resources of the National Institutes of Health.

oscillating sequence of two members of $F$ which shows that there may not be an eventually stationary optimal policy. The author has not determined whether the example represents the worst possible behavior of optimal policies or whether there are cases where there is no periodic optimal policy.

For an optimal policy $\pi$, the gain of the policy, defined by $\lim_n n^{-1}v_n(s, \pi)$, exists; it will be denoted by $v^*(s, \pi)$. It is shown in Lemma 3.1 that if $\pi$ is optimal then there is an $N$ such that $\lim_n [v_{nN+r}(s, \pi) - (nN + r)v^*(s, \pi)]$ exists for any $r$, that is, asymptotically, the return oscillates periodically around the long range average return.

One of the major results of the paper is that there is a stationary policy $\sigma$, i.e., one for which $f_m = f_n$ for all $m, n$, which has the same gain as the optimal policy $\pi$, symbolically, $v^*(s, \sigma) = v^*(s, \pi)$ for all $s$. In fact a stronger result is true (Theorem 4.2), namely that there is some constant $C$ such that $v_n(s, \pi) - v_n(s, \sigma) < C$ for all $n$ and $s$.

Although it is not known whether all optimal policies are eventually periodic, it is shown (Corollary 4.9 of Theorem 4.8) that there is an eventually periodic policy whose return is an arbitrarily small amount less than that of the optimal policy. Formally, for any $\epsilon > 0$, there is a policy $\sigma$ with the property that for some $N$, $f_{n+N} = f_n$ for $n$ sufficiently large such that

$$v_n(s, \pi) - v_n(s, \sigma) < \epsilon$$

for all $n$ and any $s$.

The final section contains further results for the special case that $q(s' \mid s, a) > 0$ for all $s'$, $s$, $a$.

## 2. Preliminaries.

LEMMA 2.1. *Let $Q$ be any $S \times S$ Markov matrix.*

(a) *The sequence $(1/(n + 1)) \sum_0^n Q^i$ converges as $n \to \infty$ to a Markov matrix $Q^*$ with*

$$QQ^* = Q^*Q = Q^*Q^* = Q^*.$$

(b) *If $A$ is any $S \times S$ matrix with $A^n \to 0$ as $n \to \infty$, then $I - A$ is non-singular and $(I - A)^{-1} = \sum_0^\infty A^i$. In particular, if $\beta < 1$ then $[I - \beta(Q - Q^*)]$ is non-singular and*

$$\textstyle\sum_0^\infty \beta^i(Q - Q^*)^i = [I - \beta(Q - Q^*)]^{-1}.$$

(c) *The matrix $I - (Q - Q^*)$ is non-singular and*

$$\lim_{\beta \uparrow 1} [I - \beta(Q - Q^*)]^{-1} = [I - (Q - Q^*)]^{-1}.$$

The first two of these results may be found in [5], the third in [2].

With each $f \varepsilon F$ is associated an $S \times 1$ vector $R_f$ with the $s'$th coordinate of $R_f$ being $r(s, f(s))$ and an $S \times S$ Markov matrix $Q_f$ with the $(s, t)'$th position in the matrix equal to $q(t \mid s, f(s))$.

If $X$ and $Y$ are two $S \times 1$ vectors, we say that $X \leqq Y$ if each coordinate of $S$ is less than or equal to the corresponding coordinate of $Y$; $X < Y$ if $X \leqq Y$ and $X \neq Y$.

We will need some results of Blackwell [2] and Howard [4]. The summary presented here is taken from [2].

Suppose that we have the setup of the first paragraph of the introduction, but in addition we assume that a return $r$ obtained $n$ units in the future is worth only $\beta^n r$ with $\beta < 1$. A strategy $\pi$ is a sequence of members of $f$ of the form $\pi = (f_1, f_2, \cdots, f_n, \cdots)$. By using a strategy $\pi$ we mean that if $s$ is observed in the $n$'th period, then action $f_n(s)$ is performed. The probability that the system is in state $t$ in the $m$'th period given that it is initially in state $s$ is the $(s, t)$'th coordinate of $Q_{f_1}Q_{f_2} \cdots Q_{f_m}$. Hence, if $V_\beta(\pi)$ is the $S \times 1$ vector whose $s$'th coordinate is the total expected return given that $s$ is the initial state then

$$V_\beta(\pi) = \sum_0^\infty \beta^n Q_{f_1} \cdots f_n R_{f_{n+1}} .$$

A strategy is stationary if $f_i = f_0$ for all $i$ and some $f_0$. This strategy is denoted by $f_0^\infty$. A strategy $\pi$ is optimal if $V_\beta(\pi) \geqq V_\beta(\sigma)$ for all strategies $\sigma$ and all $\beta$ in an interval $\beta_0 < \beta < 1$.

Theorem 2.2. *There is an optimal stationary strategy* $g^\infty$.

Lemma 2.3. *For any* $f$,

$$V_\beta(f^\infty) = [(1/(1 - \beta)) - 1]Q_f^*R_f + [I - (Q_f - Q_f^*)]^{-1}R_f + \epsilon(\beta)$$

*where* $\epsilon(\beta)$ *is an* $S \times 1$ *vector*; $\epsilon(\beta) \rightarrow 0$ *as* $\beta \uparrow 1$. *We will denote* $[I - (Q_f - Q_f^*)]^{-1}R_f$ *by* $A_f$.

The proof of the above may be found in [2].

Definition. We define $L_\beta(f)(X) = R_f + \beta Q_f X$. In particular, if $(f, \pi)$ denotes the strategy which consists of $f$ followed by the sequence $\pi$ then $V_\beta(f, \pi) = L_\beta(f)(V_\beta(\pi))$. $L_1(f)$ is abbreviated $L_f$.

Lemma 2.4. $L_f(A_f) = Q_f^*R_f + A_f$.

Proof.

$$\begin{aligned} L_f(A_f) &= \lim_{\beta\uparrow 1} [R_f + \beta Q_f \textstyle\sum_0^\infty \beta^i (Q_f - Q_f^*)^i R_f] \\ &= \lim_{\beta\uparrow 1} (\textstyle\sum_0^\infty \beta^i (Q_f - Q_f^*)^i R_f + \beta Q_f^* R_f) = Q_f^*R_f + A_f . \end{aligned}$$

Lemma 2.5. *If* $g^\infty$ *is an optimal strategy then for any* $f \varepsilon F$ *and all* $s$, *either the* $s$'th *coordinate of* $Q_fQ_g^*R_g$ *is less than the same coordinate of* $Q_g^*R_g$ *or else the* $s$'th *coordinate of* $L_f(A_g)$ *is less than or equal the same coordinate of* $L_g(A_g)$.

Proof. Assume that the $s$'th coordinate of $Q_fQ_g^*R_g$ is greater than the $s$'th coordinate of $Q_g^*R_g$. Construct $h$ by $h(t) = g(t)$, $t \neq s$ and $h(s) = f(s)$. Then $L_\beta(h)V_\beta(g^\infty) = L_\beta(h)(A_g + \epsilon(\beta)) + [(1/(1 - \beta)) - 1]Q_hQ_g^*R_g > V_\beta(g^\infty)$ for $\beta$ near 1 which contradicts the optimality of $g^\infty$.

If the $s$'th coordinate of $Q_fQ_g^*R_g$ equals the $s$'th coordinate of $Q_g^*R_g$ but the $s$'th coordinate of $L_f(A_g)$ is greater than the $s$'th coordinate of $A_g$ the same contradiction is achieved.

## 3. Asymptotic return of a stationary policy and an example.

Definition. We define $V^n(\pi, X)$ where $\pi$ is a policy and $X$ an $s \times 1$ vector by

$$\begin{aligned} V^n(\pi, X) = L_{f_n}L_{f_{n-1}} \cdots L_{f_1}(X) &= R_{f_n} + Q_{f_n}R_{f_{n-1}} + Q_{f_n}Q_{f_{n-1}}R_{f_{n-2}} \\ &\quad + \cdots + Q_{f_n} \cdots Q_{f_2}R_{f_1} + Q_{f_n} \cdots Q_{f_1}X. \end{aligned}$$

Intuitively, the $s$'th coordinate of $V^n(\pi, X)$ is the expected total return given that $s$ is the initial state, that $n$ steps of $\pi$ were used, and that after the final action was performed, the state was again observed and if found to be $t$ an extra return equal to the $t$'th coordinate of $X$ was obtained. In particular, the $s$'th coordinate of $V^n(\pi, 0)$ is $v_n(s, \pi)$.

A policy $\pi$ is $X$ optimal if $V^n(\pi, X) \geqq V^n(\sigma, X)$ for any policy $\sigma$ and for all $n$.

A policy $\pi$ is optimal if it is 0 optimal.

EXAMPLE. There are two states of the system. In state 1 there are two actions: action 1 gives a return of 1 and the system stays in state 1 with probability 1/2; action 2 gives a return of 6/5 and the system stays in state 1 with probability 1/4. In state 2, one gets a return of 0 and moves to state 1 with probability 3/4.

Let $f(1) = 1$, $g(1) = 2$. Then

$$Q_f^*R_f = Q_g^*R_g = \begin{pmatrix} 3/5 \\ 3/5 \end{pmatrix}, \quad A_f = \begin{pmatrix} 23/25 \\ 3/25 \end{pmatrix}, \quad A_g = \begin{pmatrix} 1 \\ 1/5 \end{pmatrix}.$$

It follows from 2.3 that $g^\infty$ is the optimal stationary strategy.

Let us calculate the optimal policy. If $X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ define $T(X) = x_1 - x_2 - 4/5$. Then $L_f(X) \circ L_g(X)$ iff $T(X) \circ 0$ where $\circ$ is one of $>$, $=$, or $<$. Also $T(L_fX) = -1/4\ T(X)$, $T(L_gX) = -1/2\ T(X)$. Hence the optimal policy is $\pi = (\cdots, f, g, f, g, f, g)$ an alternating sequence of $f$'s and $g$'s ending with $g$.

In order to calculate the asymptotic gain of a periodic policy we need only be able to calculate that of a stationary policy. In the above example, to calculate $V^{2n}(\pi, 0)$, we can define an operator $L_h = L_fL_g$, then calculate $L_h^n(0)$.

LEMMA 3.1. *Denote $Q_f$ by $Q$ and $R_f$ by $R$. Then:*

(a) *If all of the ergodic sets of $Q$ are aperiodic then $V^n(f^\infty, X) = (n-1)Q^*R + A_f + Q^*X + \epsilon(n)$ where $\epsilon(n) \to 0$ as $n \to \infty$;*

(b) *If $N$ is the least common multiple of the periods of the ergodic sets of $Q$, let $Q_0 = Q^N$. Then $V^{nN+m}(f^\infty, X) = (n-1)NQ^*R + [I - (Q_0 - Q_0^*)]^{-1}[\sum_0^{N-1} Q^iR] + Q_0^*[\sum_0^{m-1} Q^iR] + Q_0^*[Q^mX] + \epsilon(n)$ where $\epsilon(n) \to 0$ as $n \to \infty$.*

PROOF. $V^n(f^\infty, X) = \sum_0^{n-1} Q^iR + Q^nX$.

*Case* (a). As $Q$ is aperiodic, $lim_n\, Q^n = Q^*$. $V^n(f^\infty, X) - nQ^*R = \sum_0^{n-1} (Q^i - Q^*)R + Q^nX = \sum_0^{n-1} (Q - Q^*)^iR - Q^*R + Q^nX$ as $Q^i - Q^* = (Q - Q^*)^i$ for $i > 0$ by (a) of 2.1. Thus by (b) of 2.1, $\lim_n V^n(f^\infty, X) - nQ^*R = A_g - Q^*R + Q^*X$.

*Case* b. Use the fact that $Q_0$ is aperiodic and $Q^* = Q_0^*(\sum_0^{N-1} Q^i)$.

In our example, we have

$$V^{2n}(\pi, 0) = (2n-1)\begin{pmatrix} 6/5 \\ 6/5 \end{pmatrix} + \begin{pmatrix} 58/35 \\ 6/7 \end{pmatrix} + \epsilon(n)$$

and

$$V^{2n+1}(\pi, 0) = L_g V^{2n}(\pi, 0) = \begin{pmatrix} 3/5 \\ 3/5 \end{pmatrix} + V^{2n}(\pi, 0) + \epsilon'(n)$$

and in particular

$$\lim_n [V^n(\pi, 0) - V^n(g^\infty, 0)] = \begin{pmatrix} 2/35 \\ 2/35 \end{pmatrix}.$$

**4. Return of an optimal policy.** Throughout this section $\pi$ will be an optimal policy, and $g^\infty$ will be an optimal strategy. $Q_g$, $R_g$, and $A_g$ will be abbreviated $Q$, $R$, and $A$. The policy $g^\infty$ is, of course, a policy whose every element is $g$.

LEMMA 4.1. *The policy $g^\infty$ is $mQ^*R + A$ optimal for some $m$.*

PROOF. $L_g(mQ^*R + A) - L_f(mQ^*R + A) = m(Q^*R - Q_fQ^*R) + L_g(A) - L_f(A)$. Hence by 2.5, for $m > m_0$, $L_f(mQ^*R + A) \leqq L_g(mQ^*R + A)$. But $V^n(g^\infty, mQ^*R + A) = (m + n)Q^*R + A$ by 2.4.

THEOREM 4.2. *$V^n(\pi, 0) - V^n(g^\infty, 0)$ is bounded uniformly in $n$.*

*Proof.* It is bounded below by 0. Suppose it unbounded above. Then for any $X$, $V^n(\pi, X) - V^n(g^\infty, X)$ is unbounded above as $V^n(\pi, X) - V^n(g^\infty, X) = V^n(\pi, 0) - V^n(g^\infty, 0) + (Q_{f_n} \cdots Q_{f_1} - Q^n)X$ and the last term is bounded uniformly in $n$. But for the $X = mQ^*R + A$ of 4.1, $V^n(\pi, X) - V^n(g^\infty, X) \leqq 0$.

COROLLARY 4.3. *$\lim n^{-1}V^n(\pi, 0) = Q^*R$ and $V^n(\pi, 0) - nQ^*R$ is bounded uniformly in $n$.*

PROOF. From 3.1, $V^n(g^\infty, 0) - nQ^*R$ is bounded uniformly in $n$.

COROLLARY 4.4. *If $f$ occurs infinitely often in $\pi$, then $Q_fQ^*R = Q^*R$.*

PROOF. $(1/(n + 1))V^{n+1}(\pi, 0) = (1/(n + 1))R_{f_{n+1}} + (n/(n + 1))\cdot Q_{f_{n+1}}[n^{-1}V^n(\pi, 0)]$. Let $n \to \infty$ through a sequence for which $f_{n_i+1} = f$ and use 4.3.

The remainder of this section is devoted to showing that $V^n(\pi, 0)$ is asymptotically periodic, that is, for some fixed $N$, $\lim_n [V^{nN+r}(\pi, 0) - (nN + r)Q^*R]$ exists for all $r$.

It will suffice to prove that $V^n(\pi_X, X)$ is asymptotically periodic where $\pi_X$ is $X$-optimal and $X$ is such that if $f$ occurs in $\pi_X$ then $Q_fQ^*R = Q^*R$. This is so because $V^{n+m}(\pi, 0) = V^n(\pi', V^m(\pi, 0))$ where $\pi' = (\cdots, f_{m+2}, f_{m+1})$. Let $X = V^m(\pi, 0)$ and choose $m$ large enough that those $f$ which occur only finitely often in $\pi$ occur only as $f_1, \cdots, f_m$ and not as $f_i$, $i > m$. The reduction is obtained by 4.4.

We can then replace each $R_f$ by $R_f - Q^*R$. In this way we need only show that $\lim_n V^{nN+n}(\pi_X, X)$ exists as $Q^*R = 0$.

Let $Y$ be a limit point of $\{V^n(\pi_X, X)\}$. (There must be one by 4.2.) It will then suffice to show that for some $N$, $V^N(\pi_Y, Y) = Y$ where $\pi_Y$ is $Y$ optimal.

LEMMA 4.5. *If $f_t$, $t \varepsilon T$, is a set of real valued functions with $|f_t(a) - f_t(b)| \leqq C$ for all $t \varepsilon T$ then $|\sup_t f_t(a) - \sup_t f_t(b)| \leqq C$.*

PROOF. Suppose $\sup_t f_t(a) - \sup_t f_t(b) = C + \epsilon$. Let $t_0$ be such that $f_{t_0}(a) > \sup_t f_t(a) - \epsilon/2$. Then

$$\sup_t f_t(a) - \sup_t f_t(b) < f_{t_0}(a) - f_{t_0}(b) + \epsilon/2 < C + \epsilon/2.$$

Interchange $a$ and $b$ to complete the proof.

NOTATION. $|X|_\infty$ will be the $L^\infty$ norm of $X$, that is $|X|_\infty = \sup_s |x_s|$ where $x$ is the $s'$th coordinate of $X$.

LEMMA 4.6. *$|V^n(\pi_U, U) - V^n(\pi_V, V)|_\infty \leqq |U - V|_\infty$ where $\pi_U$ is $U$ optimal and $\pi_V$ is $V$ optimal.*

PROOF. For any policy $\pi$, $|V^n(\pi, U) - V^n(\pi, V)|_\infty = |V^{n-1}(\pi', R_{f_1}) + Q_{f_n} \cdots Q_{f_1}U - V^{n-1}(\pi', R_{f_1}) - Q_{f_n} \cdots Q_{f_1}V|_\infty = |Q_{f_n} \cdots Q_{f_1}(U - V)|_\infty \leqq |U - V|_\infty$ where $\pi' = (\cdots, f_n, \cdots, f_3, f_2)$.

Apply 4.5 to each coordinate, where the supremum is taken over all policies $\pi$.

LEMMA 4.7. *Let $Y$ be a limit point of $V^n(\pi_X, X)$ where $\pi_X$ is $X$-optimal. Then for some $N$, $V^n(\pi_Y, Y) = Y$.*

PROOF. As $Y$ is a limit point of $V^n(\pi_X, X)$, it must be a limit point of $V^n(\pi_Y, Y)$. Consequently, either $V^N(\pi_Y, Y) = Y$ for some $N$ or else $V^m(\pi_Y, Y) \neq V^n(\pi_Y, Y)$ for any $m \neq n$. Suppose that the latter is the case.

Recall that $V^n(\pi_A, A) = A$ from 4.1 and 2.4. Let $Y_0$ be a limit point of $\{V^n(\pi_Y, Y)\}$ with $|Y_0 - A|_\infty$ a minimum (which must be attained as the set of limit points of $\{V^n(\pi_Y, Y)\}$ is a compact set).

We have

$$|V^n(\pi_{Y_0}, Y_0) - V^n(\pi_A, A)|_\infty = |V^n(\pi_{Y_0}, Y_0) - A|_\infty \leqq |Y_0 - A|_\infty$$

by 4.6. It follows that each limit point of $V^n(\pi_{Y_0}, Y_0)$ has the same $L_\infty$ distance from $A$. But since $Y$ is a limit point of $V^n(\pi_Y, Y)$, so each point of $V^n(\pi_Y, Y)$ is a limit point of $V^n(\pi_Y, Y)$. Hence all have the same $L_\infty$ distance from $A$.

Note that $A + K$, where each coordinate of $K$ is $k$, has the property that $Y^n(\pi_{A+K}, A + K) = A + K$. Hence by the same argument all $V^n(\pi_Y, Y)$ are the same $L_\infty$ distance from $A + K$.

Let an infinite number of $V^n(\pi_Y, Y)$ have

$$|[V^n(\pi_Y, Y) - A]_{s_1}| = |V^n(\pi_Y, Y) - A|_\infty$$

for some $s_1$ where $[X]_s$ denotes the $s'$th coordinate of $X$. Then there is a $k_1$ such that an infinite number of these have $[V^n(\pi_Y, Y)]_{s_1} = k_1$. All of these must have the same distance from $A - K_1$ where each coordinate of $K_1$ is $k_1$. If this distance is zero the proof is finished. If not, there is an $s_2 \neq s_1$ with $|[V^n(\pi_Y, Y) - (A - K_1)]_{s_2}| = |V^n(\pi_Y, Y) - (A - K)|_\infty$ for infinitely many of the selected $V^n(\pi_Y, Y)$. Thus infinitely many of these have $[V^n(\pi_Y, Y)]_{s_2} = k_2$ and $[V^n(\pi_Y, Y)]_{s_1} = k_1$. Continuing in this manner, after at most $S$ steps we have infinitely many $V^n(\pi_Y, Y)$ agreeing in all coordinates. But this contradicts the assumption that $V^m(\pi_Y, Y) \neq V^n(\pi_Y, Y)$ for $m \neq n$. We have thus proved

THEOREM 4.8. *If $\pi$ is an optimal policy then there is an $N$ such that $\lim_n [V^{nN+r}(\pi, 0) - (nN + r)Q^*R]$ exists for any $r$.*

COROLLARY 4.9. *Given $\epsilon > 0$, there is a policy $\pi' = (\cdots, f_n, \cdots, f_1)$ with $f_{m+N} = f_m$ for $m > m_0$ such that $|V^n(\pi, 0) - V^n(\pi', 0)|_\infty < \epsilon$ for optimal $\pi$, uniformly in $n$.*

PROOF. By 4.7, there is a periodic $Y$-optimal policy for any $Y$ which is a limit point of $V^n(\pi, 0)$. Lemma 4.6 then implies this result.

**5. Analysis of a special case.** In this section we analyse the case where each entry of $Q_f$ is positive. We denote by $q_0$ the minimum element of all the $Q_f$'s.

LEMMA 5.1. *If $L_f(A) < L_g(A)$ and $f$ occurs infinitely often in a policy $\pi$, then as $n \to \infty$ each coordinate of $V^n(\pi, A) - V^n(g^\infty, A)$ goes to $-\infty$.*

PROOF. $V^n(\pi, A) = nQ^*R + C_n + Q_{f_n}C_{n-1} + \cdots + Q_{f_n} \cdots Q_{f_2}C_1$ where $C_i = (L_{f_i} - L_g)A$. This may easily be shown inductively. The hypothesis is that an infinite number of $C_n$ are $< 0$. As $Q^*$ has identical rows, 2.5 implies that all $C_n \leqq 0$. Let the $s'$th coordinate of an infinite number of $C_n$ be less than $-\epsilon$. Then for such $C_n$, each coordinate of $Q_{f_m} \cdots Q_{f_{n+1}}C_n$ is less than $-q_0\epsilon$. The assertion follows.

THEOREM 5.2. *If $\pi$ is optimal and $f$ occurs infinitely often in $\pi$ then $L_f(A) = L_g(A)$ and $Q_f{}^*R_f = Q^*R$.*

PROOF. $L_f(A) = L_g(A)$ by 5.1 and 2.5. An easy calculation shows that if $Q_fQ^*R = Q^*R$ then $Q_f{}^*R_f - Q^*R = Q_f{}^*(L_f - L_g)A$. Finally, $Q_fQ^*R = Q^*R$ because $Q^*$ has identical rows.

LEMMA 5.3. *Let $[Q_{-i}, i = 1, 2, \cdots]$ be a sequence of Markov matrices with each coordinate of $Q_{-i} > \epsilon > 0$ for all $i$. Then $\lim_n \prod_{i=-n}^{-1} Q_i$ exists and is a Markov matrix with identical rows.*

The proof is found in [3].

THEOREM 5.4. *If $\pi$ is optimal, then $V^n(\pi, 0) - nQ^*R - A$ converges as $n \to \infty$ to a limit vector with identical coordinates.*

PROOF. Let $k$ be large enough that those members of $F$ which occur only finitely often in $\pi$ occur only in $f_1$ to $f_k$ and not in $f_i$ for $i > k$. Let $J = V^k(\pi, 0) - A$. Then, using 5.2, $V^{n+k}(\pi, 0) = V^n(\pi_k, V^k(\pi, 0)) = nQ^*R + A + Q_{f_n} \cdots Q_{f_{k+1}}J$ where $\pi_k = (\cdots, f_{k+2}, f_{k+1})$. By 5.3, the rightmost term converges as $n \to \infty$ to a vector with identical coordinates.

**6. Acknowledgment.** I wish to express my sincere gratitude to Dr. David Blackwell for his interest and guidance as my thesis advisor at the University of California.

## REFERENCES

[1] BELLMAN, RICHARD (1957). *Dynamic Programming.* Princeton Univ. Press.

[2] BLACKWELL, DAVID (1962). Discrete dynamic programming. *Ann. Math. Statist.* **33** 719–726.

[3] BLACKWELL, DAVID (1945). Finite non-homogeneous chains. *Ann. of Math.* **46** 594–599.

[4] HOWARD, RONALD A. (1960). *Dynamic Programming and Markov Processes.* Technology Press and Wiley, New York.

[5] KEMENY, J. G. and SNELL, J. L. (1960). *Finite Markov Chains.* Van Nostrand, New York.