SRC TR 87-112

Control of Markov Chains With Long-Run Average Cost Criterion II

by

Vivek S. Borkar

CONTROL OF MARKOV CHAINS WITH LONG-RUN AVERAGE COST CRITERION II

VIVEK S. BORKAR
Tata Institute of Fundamental Research
Bangalore Center, P.O. Box 1234, I.I. Sc. Campus
Bangalore 560012, INDIA

AND

Systems Research Center
University of Maryland
College Park, Maryland 20740, U.S.A.

Research supported by NSF Grant CDR-85-00108

ABSTRACT

The long-run average cost control problem for discrete time Markov chains on a countable state space is studied in a very general framework. Necessary and sufficient conditions for optimality in terms of the dynamic programming equations are given when an optimal stable stationary strategy is known to exist (e.g., for the situations studied in [5]). A characterization of the desired solution of the dynamic programming equations is given in a special case. Also included is a novel convex analytic argument for deducing the existence of an optimal stable stationary strategy when that of a randomized one is known.

Key Words: Markov chains, long-run average cost, optimal control, dynamic programming, stationary strategy.

1. Introduction

In [5], the long-run average cost control problem for a Markov chain on a countable state space was studied under a very general set-up. The two theoretical issues in this problem are: (i) establishing the existence of an optimal stable stationary strategy and (ii) characterizing the same via the dynamic programming equations. The main thrust of [5] was (i) whereas (ii) was only cursorily touched upon. The present paper has two objectives. The first is to provide a more elegant alternative for a part of the argument leading to (i) in [5]. This alternative approach unmasks the underlying convex analytic structure not apparent in the lengthier argument of [5] and is in the spirit of [6] where other cost criteria were considered in a similar light. The principal objective of this paper, however, is to give a detailed treatment of the dynamic programming equations, settling (ii) above. The class of cost functions considered here is much more general than that of [5], where the cost functions were assumed to be bounded.

Although this paper is a sequel to [5] in principle, it can be read independently.

The long-run average cost control problem for Markov chains dates back to [9] for the finite state space case and [7] for the countable state space. In most of its early development, the problem was treated as the '*vanishing discount limit*' of the discounted cost control problem. This classical approach is by now standard textbook material and the reader is referred to [2], [12] (among others) for a succinct treatment. The shortcoming of this approach is that it needs a strong uniform stability condition in one of its various garbs [8]. This condition fails in many applications of interest such as controlled queues, as evidenced in [11]. Motivated by this, [3] [4] developed an alternative approach for Markov chains exhibiting a '*nearest neighbor motion*'. The latter feature requires that each state have only finitely many neighbors and that the minimum path length from state i to any prescribed finite subset of the state space tend to infinity as i does. The approach was based on a characterization of the a.s. limit points of the empirical process of the joint state and control process. It was this approach that was carried over to a much more general set-up in [5]. The present work complements [5] in the sense already described.

The paper is organized as follows: The second section is devoted to a recapitulation of the notation introduced in [3] and used throughout [3] – [6]. This notation is not standard, but turns out to be extremely handy for the approach of [3] – [6] and the present paper. In section III, we briefly recall a part of the results of [5] and give the alternative existence argument alluded to above. Section IV states the principal assumptions under which the dynamic programming equations will be studied and discusses various ramifications thereof. Section V treats the necessary conditions for optimality in terms of the dynamic programming equations. The proofs here are very much along the lines of those of [4], section 5, except for the extra work needed to take care of a possibly unbounded cost function and the absence of 'nearest neighbor motion' hypothesis. We include the full details in order to make this account self-contained. The so-called 'value function' appearing in the dynamic programming equations is further studied in section VI. Section VII establishes sufficient conditions for optimality using the dynamic programming equations. Section VIII concludes with a discussion of the problem of characterizing the desired solution of the dynamic programming equations.

Note that we develop the dynamic programming formalism given the existence of an optimal stable stationary strategy by independent means, e.g., those of [4], [5]. This is the opposite of the conventional order of things.

2. Notation and Preliminaries

Let $X_n, n = 1, 2, \ldots$, be a controlled Markov chain on state space $S = [1, 2, \ldots]$ with transition matrix $P_u = [[p(i,j,u_i)]], i, j \varepsilon S$ indexed by the control vector $u = [u_1, u_2, \ldots]$. Here, $u_i \varepsilon D(i), i \varepsilon S$, for some prescribed compact metric spaces $D(i)$. The functions $p(i,j,\cdot)$ are assumed to be continuous. By replacing each $D(i)$ by $\Pi D(k)$ and $p(i,j,\cdot)$ by its composition with the projection $\Pi D(k) \rightarrow D(i)$, one may assume that all $D(i)$'s are replicas of the same compact metric space D. We do so and then let L denote the countable product of copies of D with the product topology.

For any Polish space $Y, M(Y)$ will denote the space of probability measures on Y with the topology of weak convergence and for $n = 1, 2, \ldots, \infty, Y^n$ will denote the n-times product of Y with itself.

A control strategy (CS) is a sequence $\{\xi_n\}$, $\xi_n = [\xi_n(1), \xi_n(2), \ldots]$ of L-valued random variables such that for $i \varepsilon S$, $n \geq 1$,

$$P(X_{n+1} = i / X_m, \xi_m, m \leq n) = p(X_n, i, \xi_n(X_n)). \tag{2.1}$$

We say that $\{X_n\}$ is governed by $\{\xi_n\}$ whenever (2.1) holds. If $\{\xi_n\}$ are identically distributed and ξ_n is independent of X_m, $m \leq n$; ξ_m, $m < n$, for each n, we call the control strategy a stationary randomized strategy (SRS). We call it a stationary strategy (SS) if in addition to the above, the law of $\xi_n, n \geq 1$, is assumed to be a Dirac measure. The motivation for this nomenclature is self-evident.

We assume throughout that $\{X_n\}$ has a single communicating class under any SRS. If $\{X_n\}$ is positive recurrent under an SRS, call the latter a stable SRS or SSRS. A stable SS (or SSS) is defined analogously.

Let $\{\xi_n\}$ be an SRS. Let $\Phi \varepsilon M(L)$ denote the common law of $\xi_n, n \geq 1$. As well shall be interested only in the law of the process $(X_n, \xi_n(X_n))$, $n \geq 1$, it suffices to consider Φ of the form $\Phi = \Pi \hat{\phi}_i, \hat{\Phi}_i \varepsilon M(D)$ for $i \varepsilon S$. We shall denote this SRS by $\gamma[\Phi]$ and the corresponding transition matrix by $P[\Phi] = [[\int p(i,j,u) \hat{\phi}_i(du)]]$. If the SRS is stable, it will have a unique invariant probability measure denote by $\pi[\Phi] = [\pi[\Phi](1), \pi[\Phi](2), \ldots] \varepsilon M(S)$. For $f : S \to R$ and measurable $g : S \times D \to R$, define

$$\begin{aligned} C_f[\Phi] &= \sum_i f(i)\pi[\Phi](i), \\ g_\Phi(i) &= \int g(i,u)\hat{\Phi}_i(du), \ i\varepsilon S, \\ C_g[\Phi] &= \sum_i g_\Phi(i)\pi[\Phi](i), \end{aligned}$$

whenever the quantity on the right is defined. If $\Phi = \delta_\xi$ (i.e., the Dirac measure at ξ) for some $\xi \varepsilon L, \gamma[\Phi]$ is an SS and will be denoted by $\gamma\{\xi\}$. Correspondingly, we replace $P[\Phi], \pi[\Phi], C_f[\Phi], C_g[\Phi]$ by $P\{\xi\} = P_\xi, \pi\{\xi\}, C_f\{\xi\}, C_g\{\xi\}$ resp.

Let $k : S \times D \to R^+$ be continuous. Define

$$\psi_n = \frac{1}{n}\sum_{m=1}^{n} k(X_m, \xi_m(X_m)) \tag{2.2}$$

$$\psi_\infty = \liminf_{n\to\infty} \psi_n \tag{2.3}$$

Our objective is to a.s. minimize ψ_∞ over all CS. If this is achieved for some CS, that CS will be said to be optimal.

Note that under an SSRS $\gamma[\Phi]$ or an SSS $\gamma\{\xi\}, \psi_n \to C_k[\Phi]$ a.s. ($\psi_n \to C_k\{\xi\}$ a.s., resp.) where $+\infty$ is a possible value for $C_k[\Phi], C_k\{\xi\}$. Our aim will be to show the existence of an optimal SSS and characterize the same. Thus it is natural to impose the condition that for at least one SSS $\gamma\{\xi\}, C_k\{\xi\} < \infty$. Let

$$\begin{aligned} \beta &= \inf_{SSRS} C_k[\Phi] \\ \alpha &= \inf_{SSS} C_k\{\xi\}. \end{aligned}$$

Then $\beta \leq \alpha$.

Finally, let $\tau(i) = \min\{n \geq 1 | X_n = i\} (= \infty \text{ if } X_n \neq i \text{ for all } n), i\varepsilon S$.

3. Existence of an Optimal SSS

Consider the following two sets of assumptions:

$$(A1) \qquad \liminf_{i\to\infty} \min_u k(i,u) \triangleq \eta > \beta \tag{3.1}$$

$$(A2) \qquad \sup_{\text{all } cs} E[\tau(1)^2/X_1 = 1] < \infty. \tag{3.2}$$

Remarks More directly verifiable conditions that imply (3.2) are given in [5], Section IX. These are either conditions on the graph of the chain or require the existence of a suitable 'Liapunov function'. See [5] for details.

In [5], it was proved that under (A1) or (A2) and for bounded k,

(1) $\psi_\infty \geq \beta$ a.s.

(2) There exists an SSRS $\gamma[\Phi]$ such that $C_k[\Phi] = \beta$

(3) $\beta = \alpha$

(4) There exists an SSS $\gamma\{\xi\}$ such that $C_k\{\xi\} = \alpha$.

In this section, we provide an alternative argument to deduce (3), (4) from (2). We proceed through a sequence of lemmas.

For an SSRS $\gamma[\Phi]$, define $\hat{\pi}[\Phi]\varepsilon M(S \times D)$ by $\int f d\hat{\pi}[\Phi] = C_f[\Phi]$ for all bounded continuous $f : S \times D \to R$. For an SSS $\gamma\{\xi\}$, define $\hat{\pi}\{\xi\}\varepsilon M(S \times D)$ analogously.

Lemma 3.1 The set $B = \{\hat{\pi}[\Phi] | \gamma[\Phi] \text{ an SSRS }\}$ is convex and closed in $M(S \times D)$.

Proof Let $\gamma[\Phi_1], \gamma[\Phi_2]$ be two SSRS with $\Phi_1 = \Pi_i \hat{\Phi}_{1i}, \Phi_2 = \Pi_i \hat{\Phi}_{2i}$. Let $0 \leq a \leq 1$ and define $\Phi = \Pi_i \hat{\Phi}_i$ by

$$\hat{\Phi}_i = \left(a\pi[\Phi_1](i)\hat{\Phi}_{1i} + (1-a)\pi[\Phi_2](i)\hat{\Phi}_{2i}\right) / \left(a\pi[\Phi](i) + (1-a)\pi[\Phi]_2(i)\right), \; i\varepsilon S.$$

From this definition and the fact that

$$\pi[\Phi_i]P[\Phi_i] = \pi[\Phi_i], \; i = 1, 2,$$

it is easily seen that

$$\left(a\pi[\Phi_1] + (1-a)\pi[\Phi_2]\right) P[\Phi] = \left(a\pi[\Phi_1] + (1-a)\pi[\Phi_2]\right).$$

Thus

$$\pi[\Phi] = a\pi[\Phi_1] + (1-a)\pi[\Phi_2].$$

Hence

$$\pi[\Phi](i)\hat{\Phi}_i = a\pi[\Phi_1](i)\hat{\Phi}_{1i} + (1-a)\pi[\Phi_2](i)\hat{\Phi}_{2i}, \; i\varepsilon S,$$

implying

$$\hat{\pi}[\Phi] = a\hat{\pi}[\Phi_1] + (1-a)\hat{\pi}[\Phi_2].$$

The convexity follows. Now let $\gamma[\Phi_n], n = 1,2,\ldots$, be SSRS such that $\hat{\pi}[\Phi_n] \to \hat{\pi}$ for some $\hat{\pi}\varepsilon M(S \times D)$. Let $\pi\varepsilon M(S)$, $\pi = [\pi(1), \pi(2), \ldots]$, be the image of $\hat{\pi}$ under the projection $S \times D \to S$. Then $\pi[\Phi_n] \to \pi$ in $M(S)$. Disintegrate $\hat{\pi}$ as $\hat{\pi}(\{i\} \times A) = \pi(i)\varphi_i(A)$, $i\varepsilon S, A$ a Borel subset of D, where $\varphi_i\varepsilon M(D)$ for each i. Define $\varphi\varepsilon L$ by $\varphi = \Pi_i\varphi_i$. Since $p(\cdot, j, \cdot)$, $j\varepsilon S$, are continuous,

$$\int p(\cdot, j, \cdot)d\hat{\pi}[\Phi_n] \to \int p(\cdot, j, \cdot)d\hat{\pi},\ j\varepsilon S.$$

Thus

$$\pi[\Phi_n]P[\Phi_n] \to \pi P[\varphi]$$

termwise. Since $\pi[\Phi_n] \to \pi$ and $\pi[\Phi_n]P[\Phi_n] = \pi[\Phi_n]$, for $n \geq 1$, we have $\pi P[\varphi] = \pi$, i.e., $\pi = \pi[\varphi]$. Hence $\hat{\pi} = \hat{\pi}[\varphi]$ and we are done. QED

Let $\gamma[\Phi], \Phi = \Pi_i\hat{\Phi}_i$ be an SSRS such that for some $i_o\varepsilon S$ and $0 < a < 1$, there exist φ_1, φ_2 in $M(D)$ such that

$$\int p(i_o, \cdot, u)\hat{\Phi}_{i_o}(du) = a\int p(i_o, \cdot, u)\varphi_1(du) + (1-a)\int p(i_o, \cdot, u)\varphi_2(du),$$
$$\int p(i_o, \cdot, u)\varphi_1(du) \neq \int p(i_o, \cdot, u)\varphi_2(du) \tag{3.3}$$

as vectors, the integrations being termwise. Without any loss of generality, we shall assume that $i_o = 1$.

<u>Lemma 3.2</u> $\hat{\pi}[\Phi]$ is not an extreme point of B.

<u>Proof</u> Define $\Phi_i\varepsilon M(L)$ by $\Phi_i = \varphi_i \times \Pi_{j=2}^{\infty}\hat{\Phi}_j$, $i = 1,2$. Let τ, τ_1, τ_2 denote the first return time to 1 under $\gamma[\Phi], \gamma[\Phi_1], \gamma[\Phi_2]$ resp. when the chain starts at 1. It is easily seen that

$$\begin{aligned}
E[\tau] &= 1 + \sum_{j\neq 1}\int p(1,j,u)\hat{\Phi}_1(du)E[\tau/X_1 = j] \\
&= a\left(1 + \sum_{j\neq 1}\int p(1,j,u)\varphi_1(du)E[\tau/X_1 = j]\right) \\
&\quad +(1-a)\left(1 + \sum_{j\neq 1}\int p(1,j,u)\varphi_2(du)E[\tau/X_1 = j]\right) \\
&= aE[\tau_1] + (1-a)E[\tau_2].
\end{aligned}$$

Since $E[\tau] < \infty$, $E[\tau_i] < \infty$ for $i = 1,2$, implying that $\gamma[\Phi_i], i = 1,2$, are SSRS. If $\pi[\Phi] = \pi[\Phi_1] = \pi[\Phi_2]$, the equation

$$\sum_k \pi[\Phi_i](k) \int p(k,j,u)\hat{\Phi}_{ik}(du) = \pi[\Phi_i](j), \quad i = 1,2,$$

contradicts (3.3) for some j. Hence any two of $\pi[\Phi], \pi[\Phi_1], \pi[\Phi_2]$ are distinct from each other. Let $b\varepsilon(0,1)$ be such that

$$a = b\pi[\Phi_1](1)/(b\pi[\Phi_1](1) + (1-b)\pi[\Phi_2](1)).$$

Argue as in the proof of the preceding lemma to conclude that

$$\hat{\pi}[\Phi] = b\hat{\pi}[\Phi_1] + (1-b)\hat{\pi}[\Phi_2].$$

The claim follows. QED

<u>Corollary 3.1</u> The extreme points of B are of the form $\hat{\pi}\{\xi\}$ where $\xi\varepsilon L$ satisfies:

(*) For each $i\varepsilon S, p(i,\cdot,\xi)$ is an extreme point of $\{p(i,\cdot,\xi')|\xi'\varepsilon L\} \subset M(S)$

<u>Theorem 3.1</u> If an optimal SSRS exists, an optimal SSS satisfying (*) exists. (In particular, $\beta = \alpha$.).

<u>Proof</u> Let $\gamma[\Phi]$ be an optimal SSS. Let $\bar{S} = SU\{\infty\}$ be the one point compactification of S. We may view B as a subset of $M(\bar{S} \times D)$ by identifying each element of $M(S \times D)$ with that element of $M(\bar{S} \times D)$ that coincides with it when restricted to $S \times D$ and has zero mass at $\{\infty\} \times D$. Let $\bar{B}$ denote the closure of B in $M(\bar{S} \times D)$. Viewing $\hat{\pi}[\Phi]$ as an element of $\bar{B}$, Choquet's theorem [10] implies that $\hat{\pi}[\Phi]$ is the barycenter of a probability measure ν supported on the set of extreme points of $\bar{B}$. Since $\hat{\pi}[\Phi]$ has no mass at $\{\infty\} \times D$, ν is a.s. supported on the set of extreme points of B itself. Letting E denote the latter set,

$$\int_E \left(\int k d\hat{\pi}\right) \nu(d\hat{\pi}) = C_k[\Phi].$$

Thus if there is no $\hat{\pi}\varepsilon E$ such that $\int k d\hat{\pi} = C_k[\Phi]$, there would necessarily exist a $\hat{\pi}\varepsilon E$ for which $\int k d\hat{\pi} < C_k[\Phi]$. By the preceding corollary, each $\hat{\pi}\varepsilon E$ is of the form $\hat{\pi}\{\xi\}$ for some SSS $\gamma\{\xi\}$ satisfying (*). Thus we have a contradiction to the optimality of $\gamma[\Phi]$ unless $C_k[\Phi] = C_k\{\xi\}$ for some SSS $\gamma\{\xi\}$ with $\hat{\pi}\{\xi\}\varepsilon E$. QED.

Remark As in [5], one can prove that (A1) or (A2) imply (1), (2) above. (Though k is assumed to be bounded in [5], this part of the arguments of [5] goes through without any difficulty for the more general k's considered here.) The above can then replace the arguments of [5] to deduce (3), (4) from (1), (2). This alternative approach is both simpler and says more.

4. Stability Under Local Perturbation

In later sections, we shall give necessary and sufficient conditions for an SSS $\gamma\{\xi\}$ to be optimal, using the 'dynamic programming' equations. Some of these were stated without detailed proofs for bouded k in [5]. We make the following two assumptions:

(1) There exists an optimal SSS. (This would be implied, e.g., by (A1) or (A2).)

(2) (Stability under local perturbation) If $\gamma\{\xi\}$ is an SSS satisfying $C_k\{\xi\} < \infty$, then for any $\xi' \varepsilon L$ such that $\xi'(i) \neq \xi(i)$ for at most one $i \varepsilon S, \gamma\{\xi'\}$ is an SSS and $C_k\{\xi'\} < \infty$.

In this section, we shall make a few remarks about these conditions.

(i) If k is bounded, (2) is implied by the condition: For any $\gamma\{\xi\}, \gamma\{\xi'\}$ is also an SSS whenever $\xi'(i) = \xi(i)$ for all but one $i \varepsilon S$. Conversely, if k is bouded away from zero from below, (2) implies the above.

(ii) If all SS are SSS and k is bounded, (2) trivially holds.

(iii) (2) holds whenever each state in S has only finitely many neighbors, i.e., for each $i \varepsilon S$, there is a finite set $R_i \subset S$ such that $p(i, j, \cdot) \equiv 0$ for $j \not\varepsilon R_i$. To see this, pick $i = 1$ for example. If $\gamma\{\xi\}$ is a SSS and $C_k\{\xi\} < \infty$,

$$\begin{aligned} E[\tau(1)/X_1 = 1] &= 1 + \sum_{j \varepsilon R_1 \backslash \{1\}} p(1, j, \xi(i)) E[\tau(1)/X_1 = j] < \infty \\ E[\sum_{m=1}^{\tau(1)-1} k(X_m, \xi(X_m))/X_1 = 1] &= k(1, \xi(1)) + \sum_{j \varepsilon R_1 \backslash \{1\}} p(1, j, \xi(i)) \\ \times E[\sum_{m=1}^{\tau(1)-1} k(X_m, \xi(X_m))/X_1 = j] &= C_k\{\xi\} E[\tau(1) - 1/X_1 = 1] < \infty \end{aligned}$$

Hence under $\gamma\{\xi\}$ (and therefore under $\gamma\{\xi'\}$),

$$a_j \triangleq E[\tau(1)/X_1 = j] < \infty \ \forall \, j \varepsilon R_1$$

$$b_j \triangleq E[\sum_{m=1}^{\tau(1)-1} k(X_m, \xi_m(X_m))/X_1 = j] < \infty \ \forall\, j \varepsilon R_1 \setminus \{1\}$$

Thus under $\gamma\{\xi'\}$,

$$E[\tau(1)/X_1 = 1] = 1 + \sum_{j\varepsilon R_1\setminus\{1\}} p(1, j, \xi'(1))a_j < \infty$$

$$E[\sum_{m=1}^{\tau(1)-1} k(X_m, \xi'_m(X_m)] = k(1, \xi'(1)) + \sum_{j\varepsilon R_1\setminus\{1\}} p(1, j, \xi'(1))b_j < \infty$$

In particular, this situation covers controlled queueing networks.

(iv) The following example describes a situation where (2) fails: Relabel S as $\{a_{00}, a_{10}, a_{11}, a_{20}, a_{21}, a_{22}, a_{30}, a_{31}, a_{32}, a_{33}, a_{40}, \ldots\}$. Let $D = [1.5, 3]$. Let $p(i, j, u) = 1 \quad \forall u \varepsilon D$, $i = a_{mn}, j = a_{m(n+1)}, m = 1, 2, \ldots,\ n = 0, 1, \ldots, m-1$, and for $i = a_{mm}, j = a_{00}, m = 1, 2, \ldots$. Let $f(\alpha) = \Sigma_n n^{-\alpha}$ for $\alpha \varepsilon D$ and $p(a_{00}, a_{m0}, \alpha) = f(\alpha)^{-1} m^{-\alpha}, m = 1, 2, \ldots$.

Let $\{X_n\}$ be a Markov chain governed by the SS which picks the control α whenever the chain is in a_{00}. (The transition probabilities for all transitions except those out of a_{00} are control-independent.) Letting $\tau = \inf\{n > 1 | X_n = a_{00}\}$, we have

$$E[\tau/X_1 = a_{00}] = f(\alpha)^{-1} \sum_{m=1}^{\infty} (m+2)m^{-\alpha}$$

which is finite for $\alpha\varepsilon(2, 3]$ and ∞ for $\alpha\varepsilon[1.5, 2]$.

As an immediate consequence of these assumptions, we have the following:

Lemma 4.1 Let $\gamma\{\xi\}$ be an SSS for which $C_k\{\xi\} < \infty$. Then for any $i\varepsilon S, u\varepsilon D$,

$$\sum_{j\varepsilon S} p(i, j, u) E_\xi[\sum_{n=1}^{\tau(1)} k(X_m, \xi(X_m))/X_1 = j] < \infty, \tag{4.1}$$

$$\sum_{j\varepsilon S} p(i, j, u) E_\xi[\tau(1)/X_1 = j] < \infty, \tag{4.2}$$

where $E_\xi[\ ]$ denotes the expectation under $\gamma\{\xi\}$.

Proof Note that

$$\infty > E_\xi[\sum_{n=1}^{\tau(1)} k(X_n, \xi(X_n))/X_1 = 1] \geq a E_\xi[\sum_{n=1}^{\tau(1)} k(X_n, \xi(X_n))/X_1 = j]$$

where

$$a = P(\{X_n, n \geq 2\} \text{ hits } j \text{ before hitting } 1/X_1 = 1) > 0.$$

Thus

$$E_\xi[\sum_{n=1}^{\tau(1)} k(X_n, \xi(X_n))/X_1 = j] < \infty \quad \forall j \varepsilon S. \tag{4.3}$$

Similarly,

$$E_\xi[\tau(1)/X_1 = j] < \infty \quad \forall j \varepsilon S. \tag{4.4}$$

Let $\{\xi_n'\}$ denote a CS such that $\xi_n' = \xi$ for $n \geq 2$ and $\xi_1'(i) = u$ for some fixed $i \varepsilon S$, $u \varepsilon D$. Let $\{X_n'\}\{X_n\}$ be the chains governed by $\{\xi_n'\}, \gamma\{\xi\}$ resp. with $X_1' = X_1 = i$ Let $\tau'(i) = \inf\{n > 1 | X_n' = i\}$. Then

$$k(i,u) + \sum_{j \varepsilon S} p(i,j,u) E_\xi \left[\sum_{n=1}^{\tau(1)} k(X_n, \xi(X_n))/X_1 = j \right]$$

$$\begin{aligned} &= E\left[\sum_{n=1}^{\tau'(1)} k(X_n', \xi_n'(X_n')) \right] \\ &= E\left[(\sum_{n=1}^{\tau'(1)} k(X_n', \xi_n'(X_n'))) I\{\tau'(1) < \tau'(i)\} \right] \\ &\quad + E\left[(\sum_{n=1}^{\tau'(1)} k(X_n', \xi_n'(X_n'))) I\{\tau'(1) > \tau'(i)\} \right] \end{aligned}$$

Defining $\varphi \varepsilon L$ by $\varphi(j) = \xi(j)$ for $i \neq j$ and $\varphi(i) = u$, the above is

$$\leq E_\varphi \left[\sum_{n=1}^{\tau(1)} k(X_n, \varphi(X_n))/X_1 = i \right] + E_\varphi \left[\sum_{n=1}^{\tau(i)} k(X_n, \varphi(X_n))/X_1 = i \right]$$

$$+ E_\xi \left[\sum_{n=1}^{\tau(1)} k(X_n, \xi(X_n))/X_1 = i \right] < \infty$$

by virtue of (4.3). (4.1) follows. (4.2) follows from (4.4) by analogous arguments. QED.

5. Necessary Conditions for Optimality

This section establish necessary conditions for the optimality of an SSS in terms of the dynamic programming equations (Theorem 5.1 below).

Let $\gamma\{\xi\}$ be an SSS. Define $V\{\xi\} = [V\{\xi\}(1), V\{\xi\}(2), \ldots]^T$ by

$$V\{\xi\}(i) = E_\xi \left[\sum_{n=1}^{\tau(1)-1} (k(X_n, \xi(X_n)) - C_k\{\xi\})/X_1 = i \right], \quad i \varepsilon S.$$

This is well-defined by virtue of (4.3), (4.4). By Lemma 4.1,

$$\sum_{j \varepsilon S} p(i,j,u) V\{\xi\}(j)$$

is also well-defined for $u \varepsilon D$. Let $1_c = [1,1,\ldots]^T, U =$ the infinite identity matrix $[[\delta_{ij}]]$ and $Q_\xi = [k(1,\xi(1)), k(2,\xi(2)),\ldots]^T$. The following lemma is recalled from [4].

Lemma 5.1 $V\{\xi\}(1) = 0$ and

$$C_k\{\xi\} 1_c = (P\{\xi\} - U) V\{\xi\} + Q_\xi \tag{5.1}$$

Proof The first claim follows from the fact that

$$\begin{aligned} C_k\{\xi\} &= \sum_{i \varepsilon S} \pi\{\xi\}(i) k(i,\xi(i)) \\ &= E_\xi \left[\sum_{n=1}^{\tau(1)-1} k(X_n, \xi(X_n)) / X_1 = 1 \right] / E[\tau(1) - 1/X_1 = 1] \end{aligned}$$

Since $V\{\xi\}(1) = 0$, one has

$$\begin{aligned} V\{\xi\}(i) = k(i,\xi(i)) - C_k\{\xi\} + E_\xi &\left[\left(\sum_{n=2}^{\tau(1)-1} (k(X_n, \xi(X_n)) - C_k\{\xi\}) \right) I\{\tau(1) > 2\} / X_1 = i \right] \\ &= k(i,\xi(i)) - C_k\{\xi\} + E[V\{\xi\}(X_2) I\{\tau(1) > 2\} / X_1 = i] \\ &= k(i,\xi(i)) - C_k\{\xi\} + E[V\{\xi\}(X_2)/X_1 = i] \\ &= k(i,\xi(i)) - C_k\{\xi\} + \sum_{j \varepsilon S} p(i,j,\xi(i)) V\{\xi\}(j) \end{aligned}$$

for $i \varepsilon S$. (5.1) follows. QED

Let $A \subset S$ be a finite set containing 1 and $\xi' \varepsilon L$ such that $\xi'(i) = \xi(i)$ for $i \not\varepsilon A$. Let $A_n, n = 1,2,\ldots$ be an increasing family of finite subsets of S containing A and increasing to S. Define $\sigma_m = \min\{n \geq 1 | X_n \not\varepsilon A_m\}, m = 1,2,\ldots$ and $\sigma = \min\{n \geq 1 | X_n \varepsilon A\}$.

Observe that by the assumptions of the previous section, $\gamma\{\xi'\}$ is an SSS.

Lemma 5.2 $\lim_{n\to\infty} E_{\xi'}[V\{\xi\}(X_{\sigma_n}) I\{\tau(1) \geq \sigma_n\} / X_1 = 1] = 0$

Proof For $i \not\varepsilon A$,

$$\begin{aligned} V\{\xi\}(i) &= E_\xi \left[\sum_{n=1}^{\sigma-1} (k(X_n, \xi(X_n)) - C_k\{\xi\}) / X_1 = i \right] \\ &\quad + E_\xi \left[\sum_{n=\sigma}^{\tau(1)-1} (k(X_n, \xi(X_n)) - C_k\{\xi\}) / X_1 = i \right], \end{aligned}$$

where we use the fact that $V\{\xi\}(1) = 0$. The first term on the right remains unchanged if $E_\xi[\ ]$ is replaced by $E_{\xi'}[\ ]$ and $k(X_n, \xi(X_n))$ by $k(X_n, \xi'(X_n))$. The second term is bounded in absolute value by

$$K = \max_{i\varepsilon A} E_\xi \left[\sum_{n=1}^{\tau(1)-1} (k(X_n, \xi(X_n)) + C_k\{\xi\})/X_1 = i \right].$$

Let $c = \max(C_k\{\xi\}, C_k\{\xi'\})$. Then for $i \not\varepsilon A$,

$$\begin{aligned} |V\{\xi\}(i)| &\leq E_{\xi'} \left[\sum_{n=1}^{\sigma-1} (k(X_n, \xi'(X_n)) + c)/X_1 = i \right] + K. \\ &\leq E_{\xi'} \left[\sum_{n=1}^{\tau(1)-1} (k(X_n, \xi'(X_n)) + c)/X_1 = i \right] + K. \end{aligned}$$

Hence

$$\begin{aligned} & |E_{\xi'}[V\{\xi\}(X_{\sigma_n}) I\{\tau(1) \geq \sigma_n\}/X_1 = 1]| \\ \leq \quad & E_{\xi'} \left[\left(\sum_{m=\sigma_n}^{\tau(1)-1} (k(X_m, \xi'(X_m)) + c) \right) I\{\tau(1) \geq \sigma_n\}/X_1 = 1 \right] \\ & + K E_\xi[I\{\tau(1) \geq \sigma_n\}/X_1 = 1] \\ \longrightarrow \quad & 0 \text{ as } n \to \infty. \qquad\qquad QED. \end{aligned}$$

Theorem 5.1 If $\gamma\{\xi\}$ is an optimal SSS,

$$\beta 1_c = \min_\varphi((P\{\varphi\} - U)V\{\xi\} + Q_\varphi) \tag{5.2}$$

Remark By (5.1), it follows that the minimum in (5.2) is attained at $\varphi = \xi$.

Proof Suppose not. Then there exist $i\varepsilon S, u\varepsilon D$ and $\Delta > 0$ such that if $\varphi \varepsilon L$ is defined by $\varphi(j) = \xi(j)$ for $j \neq i, \varphi(i) = u$, then

$$\beta 1_c = (P\{\varphi\} - U)V\{\xi\} + Q_\varphi + J, \tag{5.3}$$

$J = [0, 0, \ldots, 0, \Delta, 0, \ldots, 0]$ with Δ in the i-th place. Let $\{X_n\}$ be the chain governed by $\gamma\{\varphi\}$ with $X_1 = i$. We may take $i = 1$ by relabelling S if necessary. By our assumptions of the preceding section, $\gamma\{\varphi\}$ is an SSS with $C_k\{\varphi\} < \infty$. Set $A = \{1\}$ and $\{A_n\}$ as above. By (5.3),

$$\beta = E_\varphi[V\{\xi\}(X_{m+1})/X_m] - V\{\xi\}(X_m) + k(X_m, \varphi(X_m)) + \Delta I\{X_m = 1\}.$$

Thus for $n \geq 1$,

$$\beta(\tau(1) \wedge \sigma_n - 1) = \sum_{m=1}^{\tau(1)\wedge\sigma_n - 1} (E_\varphi[V\{\xi\}(X_{m+1})/X_m] - V\{\xi\}(X_m)) + \sum_{m=1}^{\tau(1)\wedge\sigma_n - 1} k(X_m, \varphi(X_m)) + \Delta \sum_{m=1}^{\tau(1)\wedge\sigma_n - 1} I\{X_m = 1\} \tag{5.4}$$

Since $V\{\xi\}(X_{\tau(1)}) = V\{\xi\}(1) = 0$, we have

$$\begin{aligned}
& E\left[\sum_{m=1}^{\tau(1)\wedge\sigma_n - 1} (E_\varphi[V\{\xi\}(X_{m+1})/X_m] - V\{\xi\}(X_m))\right] \\
= \; & -E\left[\sum_{m=1}^{\tau(1)\wedge\sigma_n} (V\{\xi\}(X_m) - E_\varphi[V\{\xi\}(X_m)/X_{m-1}])\right] \\
& + E[V\{\xi\}(X_{\sigma_n}) I\{\tau(1) \geq \sigma_n\}] \\
= \; & E[V\{\xi\}(X_{\sigma_n}) I\{\tau(1) \geq \sigma_n\}] \\
& \text{(by the optional sampling theorem)} \\
\longrightarrow \; & 0 \text{ as } n \to \infty,
\end{aligned}$$

by Lemma 5.2. Taking expectations in (5.4), letting $n \to \infty$ and then dividing by $E_\varphi[\tau(1)]$, we get

$$\beta = C_k\{\varphi\} + \Delta\pi\{\varphi\}(1) > C_k\{\varphi\},$$

contradicting the definition of β. The claim follows. QED.

The function $i \to V\{\xi\}(i)$ corresponding to an optimal SSS $\gamma\{\xi\}$ is called the value function and (5.2) the dynamic programming equations.

6. The Value Function

The definition of the value function in the preceding section depends upon our choice of the specific optimal SSS $\gamma\{\xi\}$ and the state '1'. In this section, we eliminate this dependence.

Lemma 6.1 Suppose $W = [W(1), W(2), \ldots]^T$ satisfies

$$\beta 1_c = \inf_\varphi((P\{\varphi\} - U)W + Q_\varphi)$$

and

$$\sup_i |W(i) - V\{\xi\}(i)| < \infty$$

for some optimal SSS $\gamma\{\xi\}$. Then $W = V\{\xi\}+$ constant $\times 1_c$. In particular, it $W(1) = 0$, then $W = V\{\xi\}$.

Proof Since

$$\beta 1_c = (P\{\xi\} - U)V\{\xi\} + Q_\xi \leq (P\{\xi\} - U)W + Q_\xi,$$

we have

$$(P\{\xi\} - U)(W - V\{\xi\}) \geq 0.$$

It follows that under $\gamma\{\xi\}, W(X_n) - V\{\xi\}(X_n)$, $n \geq 1$, is a bounded submartingale w.r.t.- the natural filtration of $\{X_n\}$ and hence converges. Since $\{X_n\}$ visits each $i \varepsilon S$ infinitely often, this is possible only if $W(X_n) - V\{\xi\}(X_n)$, $n \geq 1$, is a.s. a constant sequence. QED

Lemma 6.2 Let $\gamma\{\xi\}, V\{\xi\}$ be as above and for some $i \varepsilon S$, define $V'\{\xi\} = [V'\{\xi\}(1), V'\{\xi\}(2), \ldots]^T$ by

$$V'\{\xi\}(j) = E_\xi[\sum_{m=1}^{\tau(i)-1} (k(X_m, \xi(X_m)) - \beta)/X_1 = j]$$

Then

$$V'\{\xi\} = V\{\xi\} + \text{ constant } \times 1_c.$$

Remark Note that (5.2) remains unchanged if we change $V\{\xi\}$ by a constant multiple of 1_c.

Proof For any $j \varepsilon S$,

$$\begin{aligned} |V\{\xi\}(j) - V'\{\xi\}(j)| &\leq E\left[\sum_{m=\tau(1)\wedge\tau(i)}^{\tau(1)\vee\tau(i)} (k(X_m, \xi(X_m)) + \beta)/X_1 = j\right] \\ &\leq E\left[\sum_{m=1}^{\tau(1)} (k(X_m, \xi(X_m)) + \beta)/X_1 = i\right] \\ &\quad + E\left[\sum_{m=1}^{\tau(i)} (k(X_m, \xi(X_m)) + \beta)/X_1 = 1\right]. \end{aligned} \tag{6.1}$$

Since the choice of state '1' in the preceding section was arbitrary, it is clear that (5.2) also holds with $V'\{\xi\}$ replacing $V\{\xi\}$. The claim now follows from (6.1) and the preceding lemma. QED.

Lemma 6.3 For $i\varepsilon S, \gamma\{\xi\}, V\{\xi\}$ as above,

$$V\{\xi\}(i) = \min E_\varphi\left[\sum_{m=1}^{\tau(1)-1}(k(X_m,\varphi(X_m)) - \beta)/X_1 = i\right]$$

where the minimum is over all SSS $\gamma\{\varphi\}$.

Proof For $i = 1$, the claim follows from the fact that for any SSS $\gamma\{\varphi\}$,

$$E_\varphi\left[\sum_{m=1}^{\tau(1)-1}(k(X_m,\varphi(X_m)) - \beta)/X_1 = 1\right]$$

$$= E_\varphi[\tau(1) - 1](C_k\{\varphi\} - \beta) \geq 0 = V\{\xi\}(1).$$

Take $i \neq 1$. Suppose the claim is false. Then for some SSS $\gamma\{\varphi\}$,

$$E_\varphi\left[\sum_{m=1}^{\tau(1)-1}(k(X_m,\varphi(X_m)) - \beta)/X_1 = i\right] < V\{\xi\}(i) \tag{6.2}$$

Consider the chain $\{X_n\}$ with $X_1 = 1$ governed by a CS $\{\xi_n\}$ such that between each successive returns to state 1, $\xi_n = \xi$ till $\{X_n\}$ hits i (if it does) and $= \varphi$ then on till it returns to 1. From (6.2), it follows that

$$E\left[\sum_{m=1}^{\tau(1)-1}(k(X_m,\xi_m(X_m)) - \beta)\right] < V\{\xi\}(1) = 0$$

under $\{\xi_n\}$. Letting $\{\tau_n\}$ denote the successive return times to 1, it is not hard to see that

$$\sum_{m=\tau_i}^{\tau_{i+1}-1}(k(X_m,\xi_m(X_m)) - \beta), \quad i \geq 1,$$

are i.i.d. Thus by the strong law of large numbers,

$$\left[\sum_{m=1}^{\tau_n}(k(X_m,\xi_m(X_m)) - \beta)\right]/\tau_n \xrightarrow{a.s.}$$

$$E\left[\sum_{m=1}^{\tau(1)-1}(k(X_m,\xi_m(X_m)) - \beta)/X_1 = 1\right]/E[\tau(1) - 1] < 0.$$

This contradicts the optimality of $\gamma\{\xi\}$, proving the claim. QED.

Corollary 6.1 $V\{\xi\}$ above does not depend on our choice of a specific optimal SSS $\gamma\{\xi\}$.

7. Sufficient Conditions for Optimality

In this section, we shall develop suficient conditions for optimality in terms of the dynamic programming equations. Let $\gamma\{\xi_o\}$ be an optimal SSS and $V\{\xi_o\}$ the value function. The traditional form of the sufficient conditions is as follows.

Theorem 7.1 Suppose an SSS $\gamma\{\xi\}$ satisfies $C_k\{\xi\} < \infty$ and

$$\beta 1_c = (P\{\xi\} - U)V\{\xi_o\} + Q_\xi. \tag{7.1}$$

Then $\gamma\{\xi\}$ is optimal.

Proof Argue as in the proof of Theorem 5.1 to conclude that

$$\begin{aligned}\beta E_\xi[\tau(1) \wedge \sigma_n - 1/X_1 = 1] &= E_\xi\left[\sum_{m=1}^{\tau(1)\wedge\sigma_n - 1} k(X_m, \xi(X_m))/X_1 = 1\right] \\ &\quad + E_\xi[V\{\xi_o\}(X_{\sigma_n}) I\{\tau(1) > \sigma_n\}/X_1 = 1]\end{aligned}$$

By Lemma 6.3, the last term is dominated by

$$\begin{aligned}&E_\xi\left[E_\xi\left[\sum_{m=0}^{\tau(1)-\sigma_n-1} (k(X_{\sigma_n+m}, \xi(X_{\sigma_n+m})) - \beta)\right]/X_{\sigma_n}\right] I\{\tau(1) > \sigma_n\}/X_1 = 1] \\ &= E_\xi\left[\left(\sum_{m=0}^{\tau(1)-\sigma_n-1} (k(X_{\sigma_n+m}, \xi(X_{\sigma_n+m})) - \beta)\right) I\{\tau(1) > \sigma_n\}/X_1 = 1\right] \\ &\longrightarrow 0 \text{ as } n \to \infty.\end{aligned}$$

Thus

$$\beta \geq E_\xi\left[\sum_{m=1}^{\tau(1)-1} k(X_m, \xi(X_m))/X_1 = 1\right] / E_\xi[\tau(1) - 1/X_1 = 1] = C_k\{\xi\}.$$

Since $\beta \leq C_k\{\xi\}$ in any case, $\beta = C_k\{\xi\}$ and the claim follows. QED.

We shall consider another variant of this. Call an SSS $\gamma\{\xi\}$ locally optimal if $C_k\{\xi\} \leq C_k\{\xi'\}$ for all ξ' for which $\xi(i) \neq \xi'(i)$ for at most finitely many i.

Theorem 7.2 Suppose an SSS $\gamma\{\xi\}$ satisfies $C_k\{\xi\} < \infty$ and

$$C_k\{\xi\} 1_c = \min_\varphi((P\{\varphi\} - U)V\{\varphi\} + Q_\varphi). \tag{7.2}$$

Then $\gamma\{\xi\}$ is locally optimal.

Proof Let $\xi' \varepsilon L$ be such that $\xi'(i) \neq \xi(i)$ for at most finitely many i. Then $\gamma\{\xi'\}$ is an SSS by our hypothesis (2) of section IV. Let $\{X_n\}$ be a chain governed by $\gamma\{\xi'\}$ with $X_1 = 1$. By (7.2),

$$C_k\{\xi\} \leq E_{\xi'}[V\{\xi\}(X_{n+1})/X_n] - V\{\xi\}(X_n) + k(X_n, \xi'(X_n))$$

for $n \geq 1$. As in the proof of Theorem 5.1, we can prove that for $n \geq 1$,

$$\begin{aligned} C_k\{\xi\}E_{\xi'}[\tau(1) \wedge \sigma_n - 1] \;\leq\; & E_{\xi'}\left[\sum_{m=1}^{\tau(1)\wedge\sigma_n - 1} k(X_m, \xi'_m(X_m))\right] \\ & + E_{\xi'}[V\{\xi\}(X_{\sigma_n})I\{\tau(1) \geq \sigma_n\}] \end{aligned}$$

The last term on the right tends to zero as $n \to \infty$ by Lemma 5.2. Thus letting $n \to \infty$ in the above and then dividing through by $E_{\xi'}[\tau(1) - 1]$, we get

$$C_k\{\xi\} \leq C_k\{\xi'\}. \qquad \text{QED}$$

Corollary 7.1 Suppose all locally optimal SSS are optimal. Then an SSS $\gamma\{\xi\}$ is optimal if and only if (7.2) holds.

Corollary 7.2 Suppose all SS are SSS and $\{\pi\{\xi\} | \xi \varepsilon L\}$ is tight in M(S). In addition, suppose that k is bounded. Then an SSS $\gamma\{\xi\}$ is optimal if and only if (7.2) holds.

Proof Let $\gamma\{\xi_o\}$ be an optimal SSS and $\gamma\{\xi\}$ a locally optimal SSS. Define $\xi^n \varepsilon L$ by: $\xi^n(i) = \xi_o(i)$ for $i \leq n$, $= \xi(i)$ for $i > n$. Then $P\{\xi^n\} \to P\{\xi_o\}$ termwise. Let $\pi\{\xi^n\} \to \pi \varepsilon M(S)$ along a subsequence. By Scheffe's theorem, this convergence is also in total variation. Thus letting $n \to \infty$ along this subsequence in the equation $\pi\{\xi^n\}P\{\xi^n\} = \pi\{\xi^n\}$, we get $\pi P\{\xi_o\} = \pi$, i.e., $\pi = \pi\{\xi\}$. Thus $\hat{\pi}\{\xi^n\} \to \hat{\pi}\{\xi\}$ and hence

$$C_k\{\xi^n\} \to C_k\{\xi_o\} \leq C_k\{\xi\}.$$

But $C_k\{\xi\} \leq C_k\{\xi^n\}$ by local optimality. Thus $C_k\{\xi\} = C_k\{\xi_o\}$ and $\gamma\{\xi\}$ is optimal. The claim follows from the prededing corollary. QED.

It is tempting to conjecture that local optimality always implies optimality. Note also that whenever this is the case, (7.2) is a much better sufficient condition for optimality than (7.1), because all the quantities involved depend only on the SSS $\gamma\{\xi\}$ under scrutiny and no prior knowledge of β or $V\{\xi_o\}$ is needed.

8. Characterizing the Solution of the Dynamic Programming Equations

By a solution of the dynamic programming equations, we mean a pair (c, W), $c \varepsilon R^+$, $W = [W(1), W(2), \ldots]^T$ an infinite column vector, such that

$$c1_c = \min_{\xi}((P\{\xi\} - U)W + Q_\xi). \tag{8.1}$$

Clearly, $(\beta, V\{\xi_o\})$ in the foregoing is one solution. Note that if (c, W) is a solution, so is $(c, W + \Delta 1_c)$ for any $\Delta \varepsilon R$.

In this section, we shall give a characterization that isolates the distinguished solution $(\beta, V\{\xi_o\})$ from among the solution set for the special case when (A1) holds.

Lemma 8.1 Under (A1), $V\{\xi_o\}(i), i\varepsilon S$, is bounded from below.

Proof By (A1), $A = \{i\varepsilon S | k(i,u) < \beta$ for some $u \varepsilon D(i)\}$ is a finite set. Let $\sigma = \min\{n \geq 1 | X_n \varepsilon A\}$. Then for $i \varepsilon S$,

$$\begin{aligned} V\{\xi_o\}(i) &= E_{\xi_o}\left[\sum_{m=1}^{\tau(1)-1} (k(X_m, \xi_o(X_m)) - \beta)/X_1 = 1\right] \\ &\geq E_{\xi_o}\left[\sum_{m=\sigma}^{\tau(1)-1} (k(X_m, \xi_o(X_m)) - \beta) I\{\tau(1) > \sigma\}/X_1 = i\right] \\ &\geq -\beta \sum_{j\varepsilon A} E_{\xi_o}[\tau(1)/X_1 = j] \qquad \text{QED} \end{aligned}$$

Let $G = \{f : S \to R | f(1) = 0 \text{ and } \inf_i f(i) > -\infty\}$.

Lemma 8.2 Let $(c, W) \varepsilon R^+ \times G$ be a solution of (8.1). Then $c \geq \beta$.

Proof Let $\epsilon > 0$. Then there exists an SS $\gamma\{\xi\}$ such that

$$(\epsilon + c)1_c \geq (P\{\xi\} - U)W + Q_\xi. \tag{8.2}$$

Let $\{X_n\}$ be a chain governed by $\gamma\{\xi\}$ with $X_1 = 1$. Summing up $X_1, X_2, \ldots, X_n$-th rows of (8.2), one has

$$(c + \epsilon)n \;\geq\; E_\xi[W(X_{n+1})/X_n] - \sum_{m=2}^{n} (W(X_m) - E_\xi[W(X_m)/X_{m-1}])$$

$$+\sum_{m=1}^{n} k(X_m, \xi(X_m)). \tag{8.3}$$

Let $\mathcal{F}_n = \sigma(X_i, i \leq n)$. Taking successive conditional expectations with respect to $\mathcal{F}_{n-1}, \mathcal{F}_{n-2}, \ldots, \mathcal{F}_1$ in (8.3), one gets

$$\begin{aligned}(c+\epsilon)n &\geq E_\xi[W(X_{n+1})/X_1] + E_\xi\left[\sum_{m=1}^{n} k(X_m, \xi(X_m))/X_1\right]. \\ &\geq K + E_\xi\left[\sum_{m=1}^{n} k(X_m, \xi(X_m))/X_1\right],\end{aligned}$$

where $K > -\infty$ is a lower bound on $W(i), i \varepsilon S$. Divide by n and let $n \to \infty$. If $\gamma\{\xi\}$ is not an SS, it is not hard to deduce from (A1) that

$$\liminf_{n\to\infty} \frac{1}{n} E_\xi\left[\sum_{m=1}^{n} k(X_m, \xi(X_m))/X_1 = 1\right] \geq \eta.$$

Thus

$$c + \epsilon \geq \eta \geq \beta.$$

If $\gamma\{\xi\}$ is an SSS,

$$\liminf_{n\to\infty} \frac{1}{n} E_\xi\left[\sum_{m=1}^{n} k(X_m, \xi(X_m))/X_1 = 1\right] \geq C_k\{\xi\} \geq \beta.$$

Since $\epsilon > 0$ was arbitrary, the claim follows. QED.

Lemma 8.3 If $(\beta, W), W \varepsilon G$, is a solution of (8.1), then $W \geq V\{\xi_o\}$ termwise.

Proof Let $0 < \epsilon < \eta - \beta$. Then there exists an SS $\gamma\{\xi\}$ such that

$$(\beta + \epsilon)1_c \geq (P\{\xi\} - U)W + Q_\xi. \tag{8.4}$$

If $\gamma\{\xi\}$ is not an SSS, one may argue as in the proof of the preceding lemma to conclude that $\beta + \epsilon \geq \eta$, a contradiction. Hence $\gamma\{\xi\}$ is an SSS. Let $\{\xi_n\}$ be a chain governed by $\gamma\{\xi\}$ with $X_1 = i$ for some $i \varepsilon S$. For $\{\sigma_n\}$ as before, one can deduce from (8.4) that

$$\begin{aligned}W(i) \geq & \sum_{m=1}^{\tau(1)\wedge\sigma_n - 1} (k(X_m, \xi(X_m)) - (\beta + \epsilon)) \\ & - \sum_{m=2}^{\tau(1)\wedge\sigma_n} (W(X_m) - E_\xi[W(X_m)/X_{m-1}]) \\ & + W(X_{\tau(1)\wedge\sigma_n}).\end{aligned} \tag{8.5}$$

By Fatou's lemma,

$$\liminf_{n\to\infty} E_\xi[W(X_{\tau(1)\wedge\sigma_n})/X_1 = i] = \liminf_{n\to\infty} E_\xi[W(X_{\sigma_n})I\{\tau(1) > \sigma_n\}/X_1 = i]$$

$$\geq 0$$

Thus taking expectations in (8.5) and letting $n \to \infty$,

$$W(i) \geq E_\xi \left[\sum_{m=1}^{\tau(1)-1} (k(X_m, \xi(X_m)) - (\beta + \epsilon))/X_1 = i \right]$$

Since ϵ can be made arbitrarily close to zero,

$$\begin{aligned} W(i) &\geq E_\xi \left[\sum_{m=1}^{\tau(1)-1} (k(X_m, \xi(X_m)) - \beta)/X_1 = i \right] \\ &\geq V\{\xi_o\}(i) \end{aligned}$$

where the last inequality follows from Lemma 6.3. QED

Lemma 8.4 For W as above, $W = V\{\xi_o\}$.

Proof Let $K > 0$ be a finite number such that $W(i) \geq -K$ for $i\varepsilon S$. Then for each $i\varepsilon S$,

$$\sum_{j\varepsilon S} p(i,j,u)W(j) + k(i,u) = \sum_{j\varepsilon S} p(i,j,u)(W(j) + K) + k(i,u) - K.$$

The first term on the right is a monotone increasing limit of continuous functions in the variable u and hence is lower semicontinuous in u. Since $k(i,\cdot)$ is continuous, the left hand side above is lower semicontinuous in u and hence attains a minimum at some $u_i\varepsilon D$. Let $\xi = [u_1, u_2, \ldots]\varepsilon L$.

Then

$$\beta 1_c = (P\{\xi\} - U)W + Q_\xi.$$

By the arguments used in the proof of Lemma 8.2, $\gamma\{\xi\}$ is an SSS. Since

$$\beta 1_c \leq (P\{\xi\} - U)V\{\xi_o\} + Q_\xi,$$

we have

$$(P\{\xi\} - U)(W - V\{\xi_o\}) \leq 0.$$

Letting $\{X_n\}$ be a chain governed by $\gamma\{\xi\}$ with $X_1 = 1$, this and Lemma 8.3 imply that $V = W - V\{\xi_o\}$ satisfies

$$V(X_n) \geq o = V(X_1).$$

$$E_\xi[V(X_{n+1})/X_n] \leq V(X_n),$$

$n = 1, 2, \ldots.$ This is possible only if $V(X_n) = o$ a.s. for each n. Since $\gamma\{\xi\}$ is an SSS, $X_n = i$ infinitely often a.s. for each $i \varepsilon S$. Hence $V(i) = 0$ for $i \varepsilon S$. The claim follows. QED.

The following theorem summarizes the above.

Theorem 8.1 Among all solutions (c, W) of (8.1) in $R^+ \times G, (\beta, V\{\xi_o\})$, is the unique solution corresponding to the least value of c.

References

[1] P. Billingsley, *Convergence of Probability Measures*, Wiley, 1968.

[2] D.P. Bertsekas, *Dynamic Programming and Stochastic Control*, Academic Press, 1976.

[3] V.S. Borkar, *Controlled Markov Chains and Stochastic Networks*, SIAM J. Control and Opt. 21 (1983), 652-666.

[4] V.S. Borkar, *On Minimum Cost per Unit Time Control of Markov Chains*, SIAM J. Control and Opt. 22 (1984), 965-978.

[5] V.S. Borkar, *Control of Markov Chains with Long-run Average Cost Criterion*, Proc. of the I.M.A. Workshop on Stoch. Diff. Systems with Appl. in Electrical/Computer Engg., Control Theory and Op. Research (W. Fleming and P.L. Lions, Eds.), Springer, 1987.

[6] V.S. Borkar, *A Fresh Look at Markov Decision Processes*, Preprint No. TR-87-26, Systems Research Center, University of Maryland, College Park, 1987.

[7] C. Derman, *Denumerable State Markovian Decision Processes - Average Cost Criterion*, Annals of Math. Stat. 37 (1966), 1545-1554.

[8] A. Federgruen, A. Hordijk, H.C. Tijms, *A Note on Simultaneous Recurrence Conditions on a Set of Denumerable Stochastic Matrices*, J. Appl. Prob. 15 (1978), 842-847.

[9] R. Howard, *Dynamic programming and Markov decision Processes*, M.I.T. Press, 1960.

[10] R. Phelps, *Lectures on Choquet's Theorem*, Van Nostrand, 1966.

[11] Z. Rosberg, P. Varaiya, J. Walrand, *Optimal Control of Service in Tandem Queues*, IEEE Trans. on Auto. Control, AC-27 (1982), 600-609.

[12] S. Ross, *Introduction to Stochastic Dynamic Programming*, Academic Press, 1984.