Copyright © 1977, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

INTERLIBRARY LOAN DEPARTMENT
(PHOTODUPLICATION SECTION)
THE GENERAL LIBRARY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720

OPTIMAL AND SUBOPTIMAL STATIONARY CONTROLS FOR MARKOV CHAINS

by

Pravin Varaiya

Memorandum No. UCB/ERL 77/17

24 March 1977

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

OPTIMAL AND SUBOPTIMAL STATIONARY CONTROLS FOR MARKOV CHAINS

Pravin Varaiya

Department of Electrical Engineering and Computer Sciences
and the Electronics Research Laboratory
University of California, Berkeley, California 94720

ABSTRACT

The problem studied is that of controlling a Markov chain so as to minimize the long run expected cost per unit time. Three results are obtained. First, a necessary and sufficient condition for optimality is given. The second gives for any strategy u, an easily computable bound $B(u) \geq J(u) - J^*$, where J^* is the minimum cost. The third result consists of an algorithm which, starting with any strategy, successively generates alternative strategies so that the bound $B(u)$ decreases monotonically to zero.

Research sponsored in part by the National Science Foundation Grant ENG76-16816. The author is grateful to J.-P. Forestier, G. Salut, I. Rhodes and J. Walrand for helpful discussions.

1. Introduction

We consider a finite state Markov chain x_t, $t = 0,1,\ldots$ whose transition probability at time t depends upon the control u_t chosen at t. Thus $p(x_{t+1}|x_t) = p(x_{t+1}|x_t,u_t)$. x_t is observed at each t and u_t may depend upon it. Hence we are concerned with feedback controls or strategies $u_t = u(t,x_t)$. Such a strategy incurs a cost $k(x_t,u_t)$ at time t so that over the long run the expected cost per unit time is

$$J(u) = \lim_{T\to\infty} \frac{1}{T+1} E \sum_{0}^{T} k(x_t,u(t,x_t)). \tag{1.1}$$

To guarantee that this limit is meaningful we assume that whenever u is stationary i.e. $u(t,x_t)$ is independent of t, then the Markov chain has a single ergodic class in the sense of Doob [1,p.181]. Among all (stationary) strategies we are interested in those which achieve the minimum cost J as well as those which achieve a low cost even when they are not optimal.

This mathematical problem has been extensively studied and the available results are presented in several texts including those by Howard [2], Ross [4], Kushner [3], and Bertsekas [5].

Three new results are presented here. The first states a necessary and sufficient condition for the optimality of a strategy. The second gives, for <u>any</u> strategy u, an easily computable bound $B(u) \geq J(u) - J^*$, where J^* is the minimum cost. It is our opinion that this result will be of value in the practical situation when optimal strategies are too complicated to discover or to implement and when "good" strategies can be proposed on the basis of previous experience or simplified models. The third result is an algorithm which, starting with any strategy, successively generates in an easily computable manner alternative strategies u such that the bound B(u) decreases strictly monotonically to zero.

Some of the proofs are involved. In order to maintain continuity these are collected in the Appendix.

2. Problem Formulation

The state space of the Markov chain x_t is $\{1,\ldots,s\}$. If $x_t = i$, then any control $u_t \in U(i)$ may be used. U(i) is a prespecified compact set. A (stationary) strategy is any element $u = (u(1),\ldots,u(s)) \in U = U(1)\times\ldots\times U(s)$. If $x_t = i$ and u(i) is used then

$$p_{ij}(u(i)) = \text{Prob}\{x_{t+1} = j \mid x_t = i\}$$

where the $p_{ij} : U(i) \rightarrow R$ are continuous functions such that

$$p_{ij} \geq 0, \quad \sum_j p_{ij} = 1.$$

For $u \in U$, P(u) denotes the s x s transition probability matrix $\{p_{ij}(u(i))\}$. Note that the ith row of P(u) depends only on u(i). The following assumption is in force throughout.

Ergodicity Assumption For each u the Markov chain x_t has a single ergodic class.

An equivalent assumption is that for each u there is a unique probability (row) vector $\pi(u) = (\pi_1(u),\ldots,\pi_s(u))$ such that

$$\pi(u) = \pi(u)P(u). \qquad (2.1)$$

A proof of this assertion may be found in [1,p.181]. To prevent misunderstanding we note that this assumption is strictly weaker than that of the "single ergodic class assumption" in [3,p.150] and of the "bounded mean recurrence time" condition of [4, Theorem 6.19] and [5, Proposition 3, p. 337]. In particular we permit the single ergodic class to depend on u. Also note that U is not required to be finite as in [4,5].

For each i $k(i,\cdot) : U(i) \to R$ is a continuous function giving the cost. If x_t is the Markov chain corresponding to u in U, then the average cost per unit time is given by (1.1). Now under our assumption it is known [1,pp.175-181] that

$$\lim_{T\to\infty} \frac{1}{T+1} \sum_{t=0}^{T} [P(u)]^t = \pi(u)\underset{\sim}{1} ,$$

where $\underset{\sim}{1} = (1,\ldots,1)'$. Substituting this into (1.1) shows that J(u) does not depend on the initial state and is given more simply as

$$J(u) = \pi(u)k(u) \tag{2.2}$$

where $k(u) = (k(1,u(1)),\ldots,k(s,u(s)))'$.

u is <u>optimal</u> if $J(u) = J^* = \inf\{J(v) | v \in U\}$. Since P(u) is assumed continuous, it follows from (2.1) that $\pi(u)$ is continuous. Since k(u) is also assumed continuous, so is J(u). Hence, by compactness of U, an optimal strategy u* exists. It is shown in Corollary 3.1 below that u* is then optimal with respect to all time-varying strategies also.

Our problem is to investigate strategies whose cost is close to J*.

3. <u>Optimality Conditions</u>

Recall the notation $\underset{\sim}{1} = (1,\ldots,1)'$. It is convenient to introduce

$$Q(u) = P(u) - I$$

where I is the identity matrix. Then Q(u) has rank s-1, $Q(u)\underset{\sim}{1} = 0$ and $\pi(u)$ is the unique solution of

$$\pi(u)Q(u) = 0, \ \pi(u)\underset{\sim}{1} = 1 \tag{3.1}$$

The next result is known.

Lemma 3.1 For u in U consider the s linear equations in the 1+s variables $\gamma \in R$, $c \in R^s$,

$$\gamma \underset{\sim}{1} = Q(u)c + k(u). \tag{3.2}$$

(i) If (γ, c) is a solution, then $\gamma = J(u)$. (ii) If (γ, c) is a solution, then so is $(\gamma, c+\delta\underset{\sim}{1})$ for every δ. (iii) A solution always exist.

Proof (i)follows by multiplying (3.2) on the left by $\pi(u)$, and (ii) by substitution. Finally note that $\pi(u)\ [J(u)\underset{\sim}{1}-k(u)] = 0$ so that $J(u)\underset{\sim}{1} - k(u)$ is orthogonal to the null space of $[Q(u)]'$, hence it is in the range of $Q(u)$.

□

Note that one may evaluate $J(u)$ by first solving (3.1) for $\pi(u)$ and then substituting into (2.2) or by solving (3.2) directly. In either case one has to solve s linear equations.

Let $Q_i(u)$ be the ith row of $Q(u)$. It depends only on $u(i)$. For any c let

$$H(c,u) = Q(u)c + k(u).$$

Then $H_i(c,u) = H_i(c,u(i)) = Q_i(u(i))c + k(i,u(i))$. Let $h(c)$ be given by

$$h_i(c) = \min\{H_i(c,v)\,|\,v \in U(i)\}. \tag{3.3}$$

The function H plays the role of the Hamiltonian and c the role of the dual variable. This is evident in the results below which give minimum principles. In the proofs repeated use will be made of the fact that $\pi(u)\ H(c,u) = \pi(u)\ [Q(u)c + k(u)] = \pi(u)\ k(u) = J(u)$, for all c,u. We first give a sufficiency condition.

Theorem 3.1 Let $u \in U$. Suppose there exist γ, c such that

$$\gamma \underset{\sim}{1} = h(c) = \min\{H(c,v) = |v \in U\}, \tag{3.4}$$

$$\gamma = H_i(c,u) \text{ whenever } \pi_i(u) > 0. \tag{3.5}$$

Then u is optimal and $J(u) = \gamma$. In particular, if $v \in U$ satisfies $H(c,v) = h(c)$, then v is optimal.

Proof From (3.5)

$$\pi(u)\gamma\underset{\sim}{1} = \pi(u)H(c,u) = \pi(u)k(u).$$

Let $v \in U$. From (3.4)

$$\pi(v)\gamma\underset{\sim}{1} \leq \pi(v)H(c,v) = J(v).$$

Hence $\gamma = J(u) \leq J(v)$. The last assertion is immediate. ◻

The converse of this result appears to need a much more difficult proof.

Theorem 3.2 Let u be optimal. Then there exists γ, c such that (3.4), (3.5) hold.

Proof See the Appendix. ◻

Any c for which there exists γ satisfying (3.4) is called an optimal dual variable. Evidently then $\gamma = J^*$. Let $\Pi := \{\pi | \pi \geq 0, \pi\underset{\sim}{1} = 1\}$. Then finding the optimal strategy is equivalent to solving the following nonlinear programming problem:

$$\text{Min}\{\pi k(u) | \pi Q(u) = 0, \pi \in \Pi, u \in U\}.$$

Note that the problem is not convex, hence a duality theorem appears unlikely. Nevertheless consider the dual problem

$$\underset{c}{\text{Max}}\{\text{Min}\{L(c,\pi,u) | \pi \in \Pi, u \in U\}\},$$

where the Lagrangian L is given by

$$L(c,\pi,u) = \pi k(u) + \pi Q(u)c = \pi(u)H(c,u).$$

Theorem 3.3 There exist c^*,π^*,u^* such that for all c, $\pi \in \Pi$, $u \in U$

$$L(c,\pi^*,u^*) \leq L(c^*,\pi^*,u^*) = J^* \leq L(c^*,\pi,u).$$

Moreover in this case c^* is an optimal dual variable and u^* an optimal strategy.

Proof Follows readily from Theorems 3.1, 3.2 and the fact that an optimal strategy exists. ⌑

There is a very important special case for which (3.4), (3.5) can be strengthed and which furthermore is easy to prove. We say that the strong erodicity assumption holds if for every u $\pi(u)$ is strictly positive.

Theorem 3.4 Under the strong ergodicity assumption u is optimal if and only if there exist γ, c such that

$$\gamma \underset{\sim}{1} = h(c) = H(c,u). \tag{3.6}$$

Moreover $\gamma = J(u)$.

Proof Sufficiency follows from Theorem 3.1 so that only the necessity need be shown. Suppose u is optimal and let γ, c solve (3.2),

$$\gamma \underset{\sim}{1} = Q(u)c + k(u) = H(c,u).$$

Let $v \in U$ be such that $H(c,v) = h(c) \leq \gamma \underset{\sim}{1}$. Multiplying this on the left by $\pi(v)$ gives

$$J(v) = \pi(v)H(c,v) = \pi(v)h(c) \leq \gamma\pi(v)\underset{\sim}{1} = \gamma.$$

Since $\gamma = J(u)$ by Lemma 3.1, we must have equality above. Hence $\pi(v)[h(c) - \gamma \underset{\sim}{1}] = 0$ and since $\pi(v)$ is strictly positive by assumption, this implies $h(c) = \gamma \underset{\sim}{1}$. ⌑

It is interesting to observe that under the stronger assumption the optimal dual variable is essentially unique.

Lemma 3.2 Under the strong ergodicity assumption if c, ξ are optimal dual variables then $c - \xi = \delta \underset{\sim}{1}$ for some δ.

Proof Let u be optimal. From (3.6)

$$H(c,u) - H(\xi,u) = Q(u)(c-\xi) = 0.$$

Since $\underset{\sim}{1}$ spans the null space of $Q(u)$ the result follows. ¤

As a corollary of Theorem 3.2 we show following [3,p.159] that an optimal stationary strategy is optimal with respect to arbitrary feedback controls. Let γ, c satisfy (3.2). We know that $\gamma = J^*$. Let $X_t = (x_0,\ldots,x_t)$ be the observations made up to t and let $u_t = u_t(X_t)$ be any fixed feedback control. Let

$$p(x_t = i|X_{t-1}) = \text{Prob}\{x_t = i|X_{t-1}\}, \quad i = 1,\ldots,s$$

and let $p(x_t|X_{t-1})$ be the row vector with these as components. Similarly define $p(x_{t+1}|X_{t-1})$. Evidently

$$p(x_{t+1}|X_{t-1}) = p(x_t|X_{t-1})P(u_t(x_t,X_{t-1})). \tag{3.7}$$

Now by (3.4)

$$[P(u_t(x_t,X_{t-1}))-I]c + k(u_t(x_t,X_{t-1})) \geq \gamma\underset{\sim}{1}.$$

Premultiplication by $p(x_t|X_{t-1})$ and using (3.7) gives

$$p(x_{t+1}|X_{t-1})c - p(x_t|X_{t-1})c + p(x_t|X_{t-1})k(u_t(X_t)) \geq \gamma .$$

Adding the inequalities for $t = 0,\ldots,T$ gives

$$\frac{1}{T+1} \sum_{t=0}^{T} p(x_t|X_{t-1})k(u_t(X_t)) \geq \gamma + \frac{1}{T+1} [p(x_0) - p(x_{T+1}|X_T)]c.$$

Taking expectations and the limit infimum proves the next result.

Corollary 3.1 Let $u_t(X_t)$ be any feedback control. Then

$$\liminf_{T\to\infty} \frac{1}{T+1} E \sum_0^T k(x_t, u_t(X_t)) \geq J^*.$$

4. Bounds

Recall the definition of $H(c,u)$ and $h(c)$. Also remember that $\pi(u)H(c,u) = J(u)$. Define

$$\underline{h}(c) = \min_i h_i(c) = \min_i \min_U H_i(c,u),$$

$$\bar{h}(c) = \max_i h_i(c) = \max_i \min_U H_i(c,u).$$

We begin with an elementary, but very useful, result.

Theorem 4.1 Let c be arbitrary, and u a minimizer of $H(c,\cdot)$ i.e., $H(c,u) = h(c)$. Then

$$\underline{h}(c) \leq J(u) \leq \bar{h}(c),$$

$$\underline{h}(c) \leq J^* \leq \bar{h}(c).$$

Also, u is optimal if $\underline{h}(c) = \bar{h}(c)$.

Proof Evidently, $\underline{h}(c) \leq \pi(u)h(c) = J(u) \leq \bar{h}(c)$. Next, let w be arbitrary. Then again $\underline{h}(c) \leq \pi(w)H(w,c) = J(w)$. Hence $\underline{h}(c) \leq J^*$. ¤

Consider this naive algorithm: Step 1, select c arbitrarily; Step 2, find a minimizer u of $H(c,\cdot)$. Then without computing $J(u)$ we have the bound

$$0 \leq J(u) - J^* \leq \bar{h}(c) - \underline{h}(c).$$

Of course if $J(u)$ is computed one has the better bound

$$0 \leq J(u) - J^* \leq J(u) - \underline{h}(c).$$

The only aspect of this algorithm which recommends itself is the fact that Step 2 involves s pointwise or decoupled minimizations.

A more sophisticated use of the result above is the following. Suppose a strategy u is proposed on the basis of experience or working with a simplified model. Then we find γ, c so that

$$\gamma \underset{\sim}{1} = Q(u)c + k(u).$$

Next we calculate $\underline{h}(c)$ to get the following bound. The proof follows from Theorems 3.1 and 4.1.

Lemma 4.1 (i) $J(u) = \gamma \geq J^* \geq \underline{h}(c)$. (ii) Let $\hat{U}(i) = \{v(i) \in U(i) | H_i(c,v(i)) \leq \gamma\}$. Then u is an optimal strategy with respect to $\hat{U} = \hat{U}(1) x \ldots x \hat{U}(s)$.

An application of this Lemma is given in Varaiya, Schwiezer and Hartwick [10].

5. An algorithm

Several computational algorithms are available for finding an optimal $u \in U$. The well-known "Interation in policy space" algorithm of Howard is known to generate, under some additional assumptions, a sequence of strategies $u_n \in U$ such that $J(u_n)$ converges monotonically to J* (see [3,p.154] or [5,p.349]). However, at each iteration this algorithm requires a solution of the equation $\gamma_n \underset{\sim}{1} = Q(u_n)c_n + k(u_n)$ for (γ_n, c_n), and this may be prohibitive for large s unless Q has some special structure as for example in the problem treated by Larson [6]. White [7] shows that a modification of Bellman's method of "successive approximation" converges under somewhat restrictive assumptions. However, the computational burden is considerably less than for Howard's algorithm since at each iteration only a pointwise minimization (similar to evaluation of h(c)) need be carried out. Furthermore, at each iteration a bound is available, and the bound converges monotonically (see [5,p.347].) Finally, if U is finite then it is possible to find an optimum by solving a linear programming problem [4,p.152]. The number of variables is approximately sxN where N is the cardinality of the largest U(i), so that this approach is impractical unless some special structure obtains as in Kushner and Chen [8].

The algorithm proposed here bears a family resemblance to White's algorithm in that successive dual variables are generated. However our motivation comes from the duality result of Theorem 3.3. The algorithm can be viewed essentially as a "dual method" and we search for an optimal dual variable. We begin by determining those directions along which changes in c lead to a reduction in $\bar{h}(c) - \underline{h}(c)$.

Note that $\sum_j Q_{ij}(u) = 0$, $Q_{ii}(u) \leq 0$ and $Q_{ij} \geq 0$ for $j \neq i$.

<u>Lemma 5.1</u> For c, θ, u

$$H_i(c+\theta,u) \geq H_i(c,u) \text{ if } \theta_i \leq \theta_j \text{ for all } j,$$

$$H_i(c+\theta,u) \leq H_i(c,u) \text{ if } \theta_i \geq \theta_j \text{ for all } j.$$

<u>Proof</u> $H_i(c+\theta,u) - H_i(c,u) = \sum_j Q_{ij}(u)\theta_j = \sum_j Q_{ij}(u)[\theta_j - \theta_i] \geq (\leq) 0$ if $\theta_i \leq (\geq) \theta_j$.

⌑

For any c let $\underline{S}(c) = \{i \mid h_i(c) = \underline{h}(c)\}$, $\bar{S}(c) = \{i \mid h_i(c) = \bar{h}(c)\}$. Let

$$\Theta(c) = \{\theta \mid \theta_i = \min_\ell \theta_\ell < \theta_j < \max_\ell \theta_\ell = \theta_k,\ i \in \underline{S}(c),\ k \in \bar{S}(c),\ j \notin \underline{S}(c) \cup \bar{S}(c)\}.$$

<u>Lemma 5.2</u> Suppose c is not an optimal dual variable so that $\underline{h}(c) < \bar{h}(c)$ i.e., $\underline{S}(c) \neq S$, $\bar{S}(c) \neq S$, where S is the state space. Let $\theta \in \Theta(c)$. Then

$$h_i(c+\theta) \geq h_i(c), \; i \in \underline{S}(c), \tag{5.1}$$

$$h_i(c+\theta) \leq h_i(c), \; i \in \bar{S}(c). \tag{5.2}$$

Furthermore there exists i for which at least one of these inequalities is strict.

Proof By Lemma 5.1, $H_i(c+\theta,u) \geq H_i(c,u)$ for $i \in \underline{S}(c)$. Since the inequality is preserved when minimizing over U we get (5.1); (5.2) follows in a similar way. To prove the final assertion suppose we have equality in (5.1),

$$h_i(c+\theta) = h_i(c), \; i \in \underline{S}(c).$$

Let v,w be such that

$$H(c+\theta,w) = h(c+\theta) \text{ and } H(c,v) = h(c).$$

Then for $i \in \underline{S}(c)$

$$H_i(c,v) = H_i(c+\theta,w) \leq H_i(c,w);$$

but by Lemma 5.1,

$$H_i(c+\theta,w) \geq H_i(c,w).$$

Hence

$$H_i(c+\theta,w) - H_i(c,w) = \sum_j Q_{ij}(w)\theta_j = 0, \; i \in \underline{S}(c).$$

Since $\theta_i < \theta_j$ for $i \in \underline{S}(c)$ this implies

$$Q_{ij}(w) = P_{ij}(w) = 0, \; i \in \underline{S}(c), \; j \notin \underline{S}(c). \tag{5.3}$$

Now suppose we have equality in (5.2). Then a similar argument shows that

$$Q_{kj}(v) = P_{kj}(v) = 0, \; k \in \bar{S}(c), \; j \notin \bar{S}(c) \tag{5.4}$$

Finally consider any $u \in U$ for which $u(i) = w(i)$, $i \in \underline{S}(c)$ and $u(k) = v(k)$, $k \in \bar{S}(c)$. Then (5.3), (5.4) hold for $\{P_{ij}(u)\}$. Hence there are at least two ergodic classes for u so that the ergodicity assumption is violated.

$\square$

From the proof we see that in the strongly ergodic case the lemma can be strengthened.

Corollary 5.1 Under the strong ergodicity assumption both inequalities (5.1), (5.2) above are strict, for some i.

The lemma suggests that to search for an optimal dual variable we may change the proposed vector $c(t)$ to $c(t) + \theta\Delta t$ where $\theta \in \Theta(c(t))$. It is desirable to make the choice of θ continuous in c to avoid "jamming". (see Zangwill [9] for a discussion of jamming). Consider the function $\theta(c)$ where

$$\theta_i(c) = h_i(c) - \underline{h}(c).$$

Notice that $\theta(c) \in \Theta(c)$. The proof of the next result is in the Appendix.

Theorem 5.1 Consider the differential equation

$$\frac{dc}{dt} = f(c) = \theta(c) - s^{-1}[\theta(c)'\underset{\sim}{1}]\underset{\sim}{1}. \tag{5.5}$$

(i) For every initial condition c_0 there is a unique solution $c(t,c_0)$ of (5.5) defined for all $t \geq 0$ with $c(0,c_0) = c_0$.

(ii) $c(t,c_0)$ converges to the set of all optimal dual variables c^* for which $(c^*)'\underset{\sim}{1} = c_0'\underset{\sim}{1}$. In particular, if the strong ergodicity assumption holds then c^* is unique.

(iii) $\bar{h}(c(t,c_0))$ and $\underline{h}(c(t,c_0))$ converge monotonically to J^*, and $\bar{h}(c(t,c_0)) - \underline{h}(c(t,c_0))$ decreases strictly monotonically to zero. If the strong ergodicity assumption holds then $\bar{h}$, $\underline{h}$ are strictly monotonic

Now suppose u^0 is any initially proposed strategy. Then to improve upon u^0 we should start the algorithm (5.5) at c_0 where $J(u^0)\underset{\sim}{1} = H(c_0,u^0)$.

Appendix: Proofs of Theorem 3.2, 5.1.

We first prove Theorem 5.1 via a sequence of lemmas.

Lemma A1 $h_i(c) = \min\{H_i(c,u) | u \in U\}$ is a uniformly Lipschitz function.

Proof Let $q_i = \max\{|Q_i(u)| \;|u \in U\}$ where Q_i is the ith row of Q and $|\cdot|$ is the Euclidean norm. Let c^r, u^r, $r = 1,2$ be such that

$$H_i(c^r,u^r) = Q_i(u^r)c^r + k(i,u^r(i)) = h_i(c^r).$$

Then

$$\begin{aligned} h_i(c^1) &\leq Q_i(u^2)c^1 + k(i,u^2(i)) \\ &= Q_i(u^2)(c^1-c^2) + Q_i(u^2)c^2 + k(i,u^2(i)) \\ &\leq q_i|c^1-c^2| + h_i(c^2). \end{aligned}$$

Similarly $h_i(c^2) \leq q_i|c^1-c^2| + h_i(c^1)$ and so $|h_i(c^1) - h_i(c^2)| \leq q_i|c^1-c^2|$.

¤

Corollary A1 The function f(c) in (5.1) is uniformly Lipschitz, and so the solution $c(t,c_0)$ is defined for all c_0 and $t \geq 0$.

Lemma A2 $c(t)'\underset{\sim}{1} \equiv c_0'\underset{\sim}{1}$ where $c(t) = c(t,c_0)$.

Proof $\frac{d}{dt} c(t)'\underset{\sim}{1} \equiv \theta(c)'\underset{\sim}{1} - \theta(c)'\underset{\sim}{1} \equiv 0$. ¤

Lemma A3 Let c_0 be such that $\underline{h}(c_0) < \bar{h}(c_0)$. Then

$\bar{h}(c(t,c_0))$ is non-increasing, $\underline{h}(c(t,c_0))$ is non-decreasing in t; (A1)

$\bar{h}(c(t,c_0)) - \underline{h}(c(t,c_0)) < h(c_0) - \underline{h}(c_0)$ for $t > 0$. (A2)

<u>Proof</u> To prove (A1) let $\varepsilon > 0$ and for each integer N define the function $c^N(t,c_0)$, $0 \leq t \leq \varepsilon$ by linear interpolation between the values $c^N(\frac{n}{N}\varepsilon,c_0)$, $n = 0,\ldots,N$ where

$$c^N(0,c_0) = c_0$$

$$c^N(\frac{n+1}{N}\varepsilon,c_0) = c^N(\frac{n}{N}\varepsilon,c_0) + \frac{1}{N} f(c^N(\frac{n}{N}\varepsilon,c_0)),\ n \geq 0.$$

Since f is Lipschitz by Corollary A1

$$c(t,c_0) = \lim_N c^N(t,c_0),$$

and since $\underline{h}(c)$ is continuous by Lemma A1,

$$\underline{h}(c(t,c_0)) = \lim_N \underline{h}(c^N(t,c_0)). \qquad \text{(A3)}$$

Now for each n, $f(c^N(\frac{n}{N}\varepsilon,c_0)) \in \Theta(c^N(\frac{n}{N}\varepsilon,c_0))$ and so by Lemma 5.2, for ε small, we must have

$$\underline{h}(c^N(\frac{n+1}{N}\varepsilon,c_0)) \geq \underline{h}(c^N(\frac{n}{N}\varepsilon,c_0))$$

which together with (A3) implies that $\underline{h}(c(t,c_0))$ is non-decreasing. The other assertion in (A1) follows in a similar manner.

We now prove (A2). Let $t > 0$. Then because of (A1) it is enough to show that there exists $0 < \varepsilon < t$ at which

$$\bar{h}(c(\varepsilon,c_0)) - \underline{h}(c(\varepsilon,c_0)) < \bar{h}(c_0) - \underline{h}(c_0). \qquad \text{(A4)}$$

Let

$$\underline{S}(\varepsilon) = \{i | h_i(c(\varepsilon,c_0)) = \underline{h}(c_0)\},\ \bar{S}(\varepsilon) = \{i | h_i(c(\varepsilon,c_0)) = \bar{h}(c_0)\}$$

Then to prove (A4) it is equivalent to show that either $\underline{S}(\varepsilon)$ or $\bar{S}(\varepsilon)$ is empty.

Now, by the continuity of h_i there is $0 < \delta < t$ so that for $0 < \varepsilon < \delta$

$$\underline{h}(c_0) < h_i(c(\varepsilon,c_0)) < \bar{h}(c_0),\ i \notin \underline{S}(0) \cup \bar{S}(0). \tag{A5}$$

On the other hand for $\tau > 0$

$$c(\tau,c_0) = c_0 + \tau f(c_0) + o(\tau).$$

Since $f(c_0) \in \Theta(c_0)$, therefore by Lemma 5.2,

$$h_i(c_0+\varepsilon f(c_0)) \geq \underline{h}(c_0),\ i \in \underline{S}(0),$$

$$h_i(c_0+\varepsilon f(c_0)) \leq \bar{h}(c_0),\ i \in \bar{S}(0),$$

and there is an i for which one of these inequalities is strict.

Suppose this is $i_0 \in \underline{S}(0)$ so that

$$h_{i_0}(c_0+\varepsilon_f(c_0)) > \underline{h}(c_0). \tag{A6}$$

We claim that (A6) implies that there is ε_1 so that

$$h_{i_0}(c(\varepsilon,c_0)) > \underline{h}(c_0),\quad 0 < \varepsilon \leq \varepsilon_1 \tag{A7}$$

To see this note that $h_{i_0}(c_0+\varepsilon f(c_0))$ is concave in ε[1]; hence (A6) implies that

$$\frac{\partial}{\partial\varepsilon} h_{i_0}(c_0+\varepsilon f(c_0))\Big|_{\varepsilon=0+} = \eta > 0$$

Hence

$$h_{i_0}(c(\varepsilon,c_0)) = \underline{h}(c_0) + \eta\varepsilon + o(\varepsilon)$$

from which (A7) follows. From (A5) and (A7) we obtain

$$\underline{S}(\varepsilon_1) \subset \underline{S}(0) - \{i_0\}. \tag{A8}$$

[1] Since, for fixed u, $H_i(c,u)$ is affine in c, and $h_i(c) = \min\{H_i(c,u) | u \in U\}$, therefore $h_i(c)$ is concave in c. Hence $h_i(c+\varepsilon\theta)$ is concave in ε.

On the other hand, if the strict inequality held for $i_0 \in \bar{S}(0)$ then instead of (A8) we would have shown

$$\bar{S}(\varepsilon_1) \subset \bar{S}(0) - \{i_0\}. \tag{A9}$$

Now, if either $\underline{S}(\varepsilon_1)$ or $\bar{S}(\varepsilon_1)$ is empty we are done. Otherwise both are non-empty and we repeat the argument starting with the initial condition $c_1 = c(\varepsilon_1, c_0)$ and we find ε_2 with $\varepsilon_1 < \varepsilon_2 < t$ and $i_1 \in \underline{S}(\varepsilon_1) \cup \bar{S}(\varepsilon_1)$ such that either

$$\underline{S}(\varepsilon_2) \subset \underline{S}(\varepsilon_1) - \{i_1\}$$

or $\quad \bar{S}(\varepsilon_2) \subset \underline{S}(\varepsilon_1) - \{i_1\}.$

If either $\underline{S}(\varepsilon_2)$ or $\bar{S}(\varepsilon_2)$ is empty we are done. Otherwise we repeat the argument with the initial condition $c_2 = c(\varepsilon_2, c_0)$. Since at each step k either $\underline{S}(\varepsilon_k)$ or $\bar{S}(\varepsilon_k)$ is reduced by at least one element we must arrive at a step at which one of these is empty.

¤

<u>Corollary A2</u> If the strong ergodicty assumption holds then in (A1) we have strict monotonicity.

<u>Proof</u> In the proof of Lemma A3 we can now use Corollary 5.1 in place of Lemma 5.2 so that we would have both $\underline{S}(\varepsilon_k)$ and $\bar{S}(\varepsilon_k)$ empty at some step k.

¤

<u>Lemma A4</u> Let $c(t) = c(t, c_0)$ be the solution of (5.5). Then there is $M < \infty$ such that $|c(t)| < M$ for $t \geq 0$.

<u>Proof</u> If the assertion is false, then there exists $t_n \to \infty$ such that $|c(t_n)| \to \infty$. Taking subsequences if necessary we can assume that there is a sequence $\rho_n \to \infty$ and $\theta \neq 0$ so that

$$\lim_n \rho_n^{-1} c(t_n) = \theta \tag{A10}$$

Let $u^n \in U$ be such that

$$H(c(t_n),u^n) = h(c(t_n)).$$

By Lemma A3 the sequence $h(c(t_n))$ is bounded. Hence, taking subsequences if necessary, we may assume that

$$\lim_n H(c(t_n),u^n) = h$$

for some vector h, i.e.

$$\lim_n [Q(u^n)c(t_n) + k(u^n)] = h.$$

Since U is compact we may assume that there is $u \in U$ such that u^n converges to u and so

$$\lim_n Q(u^n)c(t_n) = h- k(u).$$

Multiplying both sides by ρ_n^{-1} and using (A10) this implies that

$$Q(u)\theta = 0,$$

and so $\theta = \delta\underset{\sim}{1}$ for some δ. But by Lemma A2

$$\lim \rho_n^{-1} c(t_n)'\underset{\sim}{1} = \lim \rho_n^{-1} c_0'\underset{\sim}{1} = 0$$

which implies $\delta = 0$ and contradicts $\theta \neq 0$. ◻

<u>Proof of Theorem 5.1</u> (i) follows from Corollary A1. By Lemma A4 the trajectory $t \rightarrow c(t) = c(t,c_0)$ is bounded. Let c* be any limit point. We claim that c* is an optimal dual variable. Let $t_n \rightarrow \infty$ such that

$$\lim c(t_n) = c^*. \tag{A12}$$

Let $B(c) = \bar{h}(c) - \underline{h}(c) \geq 0$. Then by Lemma A3, $B(c(t))$ decreases strictly monotonically. By Theorem 3.1, c* is optimal if

$$B(c^*) = \lim B(c(t_n)) = 0. \tag{A13}$$

Suppose in contradiction that $B(c^*) > 0$. Then by Lemma A3, B must decrease along the trajectory strating at c^*, i.e.,

$$B(c(t,c^*)) < B(c^*), \quad t > 0. \tag{A14}$$

Because f is Lipschitz the solution of (5.5) varies continuously with initial conditions. Hence from (A12)

$$c(t,c^*) = \lim_{n\to\infty} c(t,c(t_n)) = \lim_{n\to\infty} c(t+t_n,c_0).$$

Since B is continuous, this implies

$$B(c(t,c^*)) = \lim_{n\to\infty} B(c(t+t_n,c_0)) = B(c^*)$$

which contradicts (A14). Then (A13) must hold proving the first half of (ii) and (iii). If the strong ergodicity assumption holds then c^* is unique by Lemma 3.2, and the strict monotonicity follows from Corollary A2.

¤

The proof of Theorem 3.2 also requires some preliminary results. Let u^0 be optimal. By Lemma 3.1 there are c_0 and $\gamma = J(u^0) = J^*$ so that

$$\gamma\underset{\sim}{1} = Q(u^0)c_0 + k(u^0) = H(c_0,u^0). \tag{A15}$$

<u>Lemma A5</u> Let v be such that $H(c_0,v) \leq \gamma\underset{\sim}{1}$. If $\pi_i(v) > 0$ then $H_i(c_0,v) = \gamma$.

<u>Proof</u> Premultiplying $H(c_0,v) \leq \gamma\underset{\sim}{1}$ by $\pi(v)$ gives

$$J(v) = \pi(v)H(c_0,v) \leq \pi(v)\gamma\underset{\sim}{1} = \gamma = J^*$$

so that we must have equaltiy. Hence

$$\pi(v)[H(c_0,v) - \gamma\underset{\sim}{1}] = 0$$

from which the result follows.

¤

Lemma A6 If $\pi_i(u^0) > 0$ then $H_i(c_0,u^0) = h_i(c_0) = \min\{H_i(c_0,v) | v \in U\}$.

Proof Let $J = \{i | H_i(c_0,u^0) < h_i(c_0)\}$, and let $v \in U$ be such that

$$H_i(c_0,v) < H_i(c_0,u^0) = \gamma, \; i \in J.$$

Define u by

$$u(i) = \begin{cases} v(i) & , i \in J \\ u^0(i) & , i \notin J. \end{cases}$$

Then

$$H_i(c_0,u) = \begin{cases} H_i(c_0,u) < \gamma & , i \in J \\ H_i(c_0,u^0) = \gamma & , i \notin J. \end{cases}$$

By Lemma A5 $\pi_i(u) = 0$, $i \in J$. Hence

$$P_{ij}(u) = P_{ij}(u^0) = 0, \; i \notin J, \; j \in J. \tag{A16}$$

But then $\pi_i(u^0) = 0$, $i \in J$. ¤

Let $c(t) = c(t,c_0)$ be the solution of (5.5) starting at c_0. From (A15) it follows that

$$\bar{h}(c(0)) = \gamma, \; \underline{h}(c(0)) \leq \gamma .$$

Since, by Theorem 5.1, $\bar{h}(c(t)) \geq J^* = \gamma \geq \underline{h}(c(t))$, it follows that

$$\bar{h}(c(t)) \equiv \gamma .$$

Now let $c_I(t)$ be the subvector of $c(t)$ with components in $I = \{i | i \notin J\}$.

Lemma A7 $c_I(t) - c_I(0) = \delta(t)\underset{\sim}{1}$ for some function $\delta(t)$.

Proof Suppose the assertion holds for some $t \geq 0$. Then from (A16) we can see that

$$H_i(c(t),u^0) = H_i(c(0),u^0) = \gamma, \; i \in I.$$

From (A17) and (5.5) it then follows that

$$\frac{d}{dt} c_I(t) = \xi(t)\underset{\sim}{1}$$

for some number $\xi(t)$. The result follows upon integrating this equation.

¤

Proof the Theorem 3.2 From (A16) and Lemma A7 it follows that

$$H_i(c(t),u^0) = \bar{h}(c(t)) \equiv \gamma,\ i \in I. \tag{A18}$$

Let c* be a limit point of the trajectory c(t). Then c* is an optimal dual variable by Theorem 5.1, hence it satisfies (3.4). Also from (A18) it follows that

$$H_i(c^*,u^0) = \gamma,\ i \in I.$$

Since $\pi_i(u^0) > 0$ implies $i \in I$ by Lemma A6, therefore (3.5) holds.

¤

References

[1] J. L. Doob, Stochastic Processes, New York: John Wiley and Sons, 1953.

[2] R. A. Howard, Dynamic Programming and Markov Processes, Cambridge: MIT Press, 1960.

[3] H. Kushner, Introduction to Stochastic Control, New York: Holt, Rinehart and Winston, 1971.

[4] S. M. Ross, Applied Probability Models with Optimization Applications, San Francisco: Holden-Day, 1970.

[5] D. P. Bertsekas, Dynamic Programming and Stochastic Control, New York: Academic Press, 1976.

[6] R. C. Larson, "A Hypercube Queueing Model for Facility Location and Redistricting in Urban Emergency Sciences," Computers and Operations Research 1, 67-95, 1974.

[7] D. J. White, "Dynamic Programming, Markov Chains, and the Method of Successive Approximations," J. Mathematical Analysis and Applications, 6, 373-376, 1963.

[8] H. J. Kushner and C.-H. Chen, "Decomposition of systems governed by Markov chains," IEEE Trans. Automatic Control AC-19 (5), 501-507, 1974.

[9] W. I. Zangwill, Nonlinear Programming: A Unified Approach, New York: Prentice-Hall, 1969.

[10] P. Varaiya, U. Schweizer and J. Hartwick, "A class of Markovian problems related to the districting problem for urban services," to appear in Ricerche di Automatica, 1977.