

Controlled Markov chains with constraints

VIVEK S BORKAR

Department of Electrical Engineering, Indian Institute of Science, Bangalore
560 012, India

Abstract. We consider the ergodic control of a Markov chain on a countable state space with a compact action space in presence of finitely many (say, m) ergodic constraints. Under a condition on the cost functions that penalizes instability, the existence of an optimal stable stationary strategy randomized at a maximum of m states is established using convex analytic arguments.

Keywords. Controlled Markov chains; ergodic control; control under constraints; optimal strategy; stationary strategy.

1. Introduction

Ross (1989) studied ergodic (or 'long-run average cost') control of Markov chains with finite state and action spaces when finitely many (say, m) other ergodic costs are required to satisfy prescribed bounds. Using linear programming arguments, he proved the existence of an optimal strategy randomized at a maximum of m states. [See Hordijk & Kallenberg (1984), Beutler & Ross (1985) and Altman & Shwartz (1990) for related work.] This result is extended in Borkar (1991) to countable state space and compact action space using convex analytic arguments under the hypothesis that the chain be positive recurrent under all stationary randomized strategies and the corresponding invariant probability measures tight. In this work, we drop the latter condition, replace it by a condition on the costs that discourages instability and recover the same result.

This class of problems is motivated by the following considerations. Controlled Markov chains are a popular paradigm for dynamic decision-making under uncertainty. Important application areas include control of queueing networks used to model computer and communication networks, flexible manufacturing systems etc., not to mention economic applications. In many of these situations, the problem calls for a simultaneous consideration of more than one optimization criterion. There are several ways to approach such multiobjective problems. One is to reduce it to a single objective problem by coalescing the objectives into one via a 'utility function'. However, the choice of such a function is not always obvious nor are the desirable ones necessarily the most amenable to analysis. An alternative approach, in tune with standard engineering practice, is to optimize one objective function (presumably the most important) while keeping the rest within reasonable bounds. This leads to the constrained control problem we discuss here.

A remark on the cost structure: In situations where money or effort contributed now is more expensive than that contributed later (due to interest, inflation etc.), one 'discounts' the future leading to the 'discounted cost control problems'. This is typical in economic applications. There are, however, situations where one plans for and anticipates a near-equilibrium (or 'steady state') behaviour for long stretches of time with no reason to favour short-term optimality over long-term optimality and with set-up costs, transients etc. either negligible or separately taken care of. The prime example of such is the case of computer and communication networks. Then it makes sense to consider the 'long-run average' or 'ergodic' cost as we do here whereby one averages the 'running cost' function with respect to the equilibrium distribution.

As for the assumptions used in Borkar (1991) and here, the former assumes a blanket stability, thereby ruling out a priori any possibility that the system may become unstable (e.g. the queue size blows up). This is often unreasonable in the applications mentioned above. The present set of assumptions allows instability but imposes a cost structure which penalizes instability and thereby ensures that the optimal system will also be stable.

The next section gives a precise statement of the problem. The main result is proved in §4 following some preliminaries in §3.

2. Notation and problem description

Let $X_n, n \geq 0$, be a controlled Markov chain on state space $S = \{1, 2, \dots\}$ with transition matrix $P_u = [[p(i, j, u_i)]]$, $i, j \in S$, indexed by the control vector $u = [u_1, u_2, \dots]$. Here $u_i \in D(i)$ for $i \in S$, $D(i)$ being a prescribed compact metric space. As argued in Borkar (1989, p. 642), we may assume that the $D(i)$'s are identical copies of a fixed compact metric space D . Let $L = D^\infty$. The maps $p(i, j, \cdot)$ are assumed to be continuous.

For any Polish space Y , let $P(Y)$ = the space of probability measures on Y with Prohorov topology (Billingsley 1968). If Y is countable, say $\{1, 2, \dots\}$, write $\nu \in P(Y)$ as a row vector $[\nu(\{1\}), \nu(\{2\}), \dots]$ or simply $[\nu(1), \nu(2), \dots]$.

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

$$P(X_{n+1} = i / X_m, Z_m, m \leq n) = p(X_n, i, Z_n(X_n)). \quad (1)$$

If $\{Z_n\}$ are independent and identically distributed (i.i.d.) with a common law $\Phi \in P(L)$, call it a stationary randomized strategy (SRS for short), denoted $\gamma[\Phi]$. As argued in (Borkar 1989, p. 642), we may take Φ to be a product measure $\Phi = \prod \hat{\Phi}_i$ for $\hat{\Phi}_i \in P(D)$, $i \in S$. Under $\gamma[\Phi]$, $\{X_n\}$ is a Markov chain with a stationary transition matrix

$$P[\Phi] = \left[\left[\int p(i, j, u) \hat{\Phi}_i(du) \right] \right].$$

If Φ is a Dirac measure at some $\xi \in L$, call it a stationary strategy (SS for short), denoted $\gamma\{\xi\}$. The corresponding transition matrix is $P\{\xi\} = P_\xi$.

We assume throughout that S is a single communicating class under all SRS. Say that an SRS $\gamma[\Phi]$ (respectively an SS $\gamma\{\xi\}$) is a stable SRS or SSRS (respectively a stable SS or SSS) if the corresponding Markov chain is positive recurrent and thus has a unique invariant probability measure $\pi[\Phi]$ (respectively $\pi\{\xi\}$). Define the corres-

ponding 'ergodic occupation measure' $\hat{\pi}[\Phi] \in P(S \times D)$ by

$$\int f d\hat{\pi}[\Phi] = \sum \pi[\Phi](i) \int f(i, u) \hat{\Phi}_i(du), f \in C_b(S \times D).$$

($\hat{\pi}\{\xi\}$ is defined analogously.) Let $G = \{\hat{\pi}[\Phi] | \gamma[\Phi] \text{ an SSRS}\}$.

Let $k_i: S \times D \rightarrow R^+, 0 \leq i \leq m$, be continuous and $\alpha_i > 0, 1 \leq i \leq m$, be prescribed. Let $H \subset G$ be the set on which

$$\int k_i d\hat{\pi}[\Phi] \leq \alpha_i, \quad 1 \leq i \leq m. \tag{2}$$

We assume that H is nonempty and furthermore requires $\{k_i\}$ to satisfy

$$\liminf_{j \rightarrow \infty} \inf_u k_i(j, u) > \alpha_i, \quad 0 \leq i \leq m, \tag{3}$$

where

$$\alpha_0 = \inf_H \int k_0 d\hat{\pi}[\Phi],$$

is assumed to be finite. (3) is satisfied in particular for $\{k_i\}$ of the form $k_i(j, u) = f_i(j)$ where $f_i(j)$ increases to ∞ as $j \rightarrow \infty$.

We also make the following assumption, called 'stability under local perturbation' in Borkar (1989): If $\gamma[\Phi]$ is an SSRS and product measures $\Phi', \Phi \in P(L)$ differ in at most finitely many components, then $\gamma[\Phi']$ is an SSRS. As argued in Borkar (1989, p. 645), this is satisfied in particular when each $i \in S$ has only finitely many neighbours, i.e., $p(i, j, \cdot) = 0$ for all but a finitely many j .

Our problem is to minimize over H the cost

$$\int k_0 d\hat{\pi}[\Phi]. \tag{4}$$

3. Preliminaries

This section establishes some technical lemmas, the first being recalled from Borkar (1989). Let $P_0 \subset L$ be the set of product probability measures on L .

Lemma 3.1. (Borkar 1989) G is closed convex and its extreme points correspond to SSS.

Proof. Let $\gamma[\Phi_1], \gamma[\Phi_2]$ be two SSRS with $\Phi_j = \Pi \hat{\Phi}_{j_i}, j = 1, 2$. Let $0 \leq a \leq 1$ and define $\Phi = \Pi \hat{\Phi}_{i \in P_0(L)}$ by

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

From this definition and the fact that

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

it is easily seen that

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

Thus, $\pi[\Phi] = a\pi[\Phi_1] + (1 - a)\pi[\Phi_2]$. Then a straightforward computation leads to

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

Convexity follows. Now let $\gamma[\Phi_n], n = 1, 2, \dots$, be a sequence of SSRS such that $\hat{\pi}[\Phi_n] \rightarrow \nu$ for some $\nu \in P(S \times D)$. Let $\pi \in P(S)$ be the image of ν under the projection $S \times D \rightarrow S$. Then $\pi[\Phi_n] \rightarrow \pi$ in $P(S)$. Disintegrate ν (Schwartz 1961) as $\nu(\{i\} \times B) = \pi(\{i\})\hat{\Phi}_i(B)$ for $i \in S$ and B a Borel subset of D , with the regular conditional law $\hat{\Phi}_i \in P(D)$ for each i . Let $\Phi = \Pi\hat{\Phi}_i$. Since $p(\cdot, j, \cdot), j \in S$, are continuous,

$$\int p(\cdot, j, \cdot) d\hat{\pi}[\Phi_n] \rightarrow \int p(\cdot, j, \cdot) d\nu, \quad j \in S.$$

Also, $\pi[\Phi_n] \rightarrow \pi$ in $P(S)$ implies by Scheffe's theorem (Billingsley 1968) that

$$\pi[\Phi_n] \rightarrow \pi \text{ in total variation.} \tag{6}$$

Putting the two together

$$\pi[\Phi_n]P[\Phi_n] \rightarrow \pi P[\Phi], \tag{7}$$

termwise. From (5)–(7), we have $\pi P[\Phi] = \pi$, i.e., $\pi = \pi[\Phi]$. Hence $\nu = \hat{\pi}[\Phi]$ and G is closed. Next, let $\gamma[\Phi], \Phi = \Pi\hat{\Phi}_i$, be an SSRS such that for some $i_0 \in S$ and $0 < a < 1$, there exist ϕ_1, ϕ_2 in $P(D)$ such that

$$\int p(i_0, \dots, u) \hat{\Phi}_{i_0}(du) = a \int p(i_0, \dots, u) \phi_1(du) + (1 - a) \int p(i_0, \dots, u) \phi_2(du), \tag{8}$$

$$\int p(i_0, \dots, u) \phi_1(du) \neq \int p(i_0, \dots, u) \phi_2(du), \tag{9}$$

as vectors, the integrations being termwise. Without any loss of generality, let $i_0 = 1$. Define $\Phi_i \in P_0(L)$ by $\Phi_i = \phi_i \times \prod_{j=2}^{\infty} \hat{\Phi}_j, i = 1, 2$.

From the assumption of stability under local perturbations, it follows that $\gamma[\Phi_i], i = 1, 2$, are SSRS. If $\pi[\Phi] = \pi[\Phi_1] = \pi[\Phi_2]$, the equation

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

contradicts (9) for some j . Hence any two of $\pi[\Phi], \pi[\Phi_1], \pi[\Phi_2]$ are distinct from each other. Let $b \in (0, 1)$ be such that

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

Argue as above to conclude that

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

Thus $\hat{\pi}[\Phi]$ cannot be an extreme point of G , proving the second claim.

QED

COROLLARY 3.1

H is closed convex.

Proof. Convexity follows easily from the above. Closedness follows on observing that (2) is preserved under sequential limits in G . QED

Lemma 3.2. Any element of H is the barycentre of a probability measure supported on the extreme points of H .

Proof. Let $\bar{S} = S \cup \{\infty\}$ denote the one point compactification of S . View $P(S \times D)$ as a subset of $P(\bar{S} \times D)$ by identifying each $\mu \in P(S \times D)$ with the unique $\bar{\mu} \in P(\bar{S} \times D)$ that restricts to μ on $S \times D$. Let $\bar{H}(G)$ be the closure of $H(G)$ in $P(\bar{S} \times D)$. Then \bar{H} is compact. Let H_e, \bar{H}_e be the sets of extreme points of H, \bar{H} respectively. Suppose $v \in H_e \setminus \bar{H}_e$. Then v is a convex combination of two distinct elements of \bar{H} at least one of which must lie in $\bar{H} \setminus H$ and thus assign a strictly positive mass to $\{\infty\} \times D$. This is possible only if $v(\{\infty\} \times D) > 0$ which is false. Thus $H_e \subset \bar{H}_e$. By Choquet's theorem (Phelps 1966) each $\mu \in H$ is the barycentre of a probability measure η on \bar{H}_e . If $\eta(\bar{H}_e \setminus H_e) > 0$, then we must have $\mu(\{\infty\} \times D) > 0$, a contradiction. Thus $\eta(H_e) = 1$ and the claim follows. QED

Lemma 3.3. Each $v \in \bar{H}$ is of the form

$$v(A) = \delta v'(A \cap (S \times D)) + (1 - \delta) v''(A \cup (\{\infty\} \times D)) \tag{10}$$

for A Borel in $\bar{S} \times D$ where $\delta \in [0, 1]$, $v' \in G$ and $v'' \in P(\{\infty\} \times D)$.

Proof. That (10) holds for some $v' \in P(S \times D)$ is clear. Without loss of generality, let $\delta > 0$. If $\delta = 1$, $v \in H$ and the claim is immediate. If not, there exist $\{\hat{\pi}[\Phi_n]\}$ in H such that

$$\hat{\pi}[\Phi_n] \rightarrow v \text{ in } P(\bar{S} \times D).$$

For $j \in S$,

$$\int p(\cdot, j, \cdot) d\hat{\pi}[\Phi_n] = \hat{\pi}[\Phi_n](\{j\} \times D).$$

Since $\{j\} \times D$ is both open and closed in $\bar{S} \times D$,

$$\hat{\pi}[\Phi_n](\{j\} \times D) \rightarrow v(\{j\} \times D) = \delta v'(\{j\} \times D). \tag{11}$$

Disintegrate v' as

$$v'(\{i\} \times du) = \pi(\{i\}) \phi_i(du)$$

for $\pi \in P(S), i \rightarrow \phi_i: S \rightarrow P(D)$. Set $\Phi = \Pi_i \phi_i$. For $N \geq 1$,

$$\sum_{i=1}^N \pi[\Phi_n](i) \int p(i, j, u) \hat{\Phi}_{ni}(du) \rightarrow \delta \sum_{i=1}^N \pi(\{i\}) \int p(i, j, u) \phi_i(du).$$

Thus

$$\liminf_{n \rightarrow \infty} \int p(\cdot, j, \cdot) d\hat{\pi}[\Phi_n] \geq \delta \int p(\cdot, j, \cdot) dv'. \tag{12}$$

Equations (11), (12) together imply

$$\int p(\cdot, j, \cdot) d\nu' \leq \nu'(\{j\} \times D), \quad j \in S.$$

Both sides sum up to 1 when summed over j . Thus the equality must hold for each j . Thus

$$\sum_i \pi(\{i\}) \int p(i, j, u) \phi_i(du) = \pi(\{j\}), \quad j \in S.$$

Thus $\nu' = \hat{\pi}[\Phi]$, completing the proof. QED

Theorem 3.1 Expression (4) attains its minimum at an extreme point of H .

Proof. Let $\hat{\pi}[\Phi_n]$, $n \geq 1$, be a sequence in H such that

$$\int k_0 d\hat{\pi}[\Phi_n] \downarrow \alpha_0.$$

In view of the foregoing, we may drop to a subsequence if necessary and suppose that $\hat{\pi}[\Phi_n] \rightarrow \nu$ in \bar{H} for a ν as in (10) with $\nu' = \hat{\pi}[\Phi]$ for some $\gamma[\Phi]$. Fix j , $1 \leq j \leq m$. Pick $\varepsilon_j > 0, m_j \leq 1$ such that

$$\inf_u k_j(i, u) \geq \alpha_j + \varepsilon_j, \text{ for } i \geq m_j.$$

For $n \geq 1$, set

$$k_{jn}(i, u) = k_j(i, u)I\{i \leq m_j + n\} + (\alpha_j + \varepsilon_j)I\{i > m_j + n\}.$$

Then

$$\begin{aligned} \alpha_j &\geq \liminf_{l \rightarrow \infty} \int k_j d\hat{\pi}[\Phi_l] \\ &\geq \lim_{l \rightarrow \infty} \int k_{jn} d\hat{\pi}[\Phi_l] \\ &= \delta \int k_{jn} d\hat{\pi}[\Phi] + (1 - \delta)(\alpha_j + \varepsilon_j). \end{aligned}$$

Letting $n \rightarrow \infty$ on the right,

$$\alpha_j \geq \delta \int k_j d\hat{\pi}[\Phi] + (1 - \delta)(\alpha_j + \varepsilon_j).$$

This is possible only if $\delta > 0$ and

$$\int k_j d\hat{\pi}[\Phi] \leq \alpha_j.$$

Since $j, 1 \leq j \leq m$, was arbitrary, $\hat{\pi}[\Phi] \in H$. Now let $\varepsilon > 0$ and $m_0 \geq 1$ be such that

$$\inf_u k(i, u) \geq \alpha_0 + \varepsilon, \text{ for } i \geq m_0.$$

Argue as above to conclude

$$\alpha_0 \geq \delta \int k_0 d\hat{\pi}[\Phi] + (1 - \delta)(\alpha_0 + \varepsilon)$$

and therefore

$$\int k_0 d\hat{\pi}[\Phi] \leq \alpha_0.$$

From the definition of α_0 , equality holds. Thus (4) attains its minimum on H at $\hat{\pi}[\Phi]$. By lemma 3.2, $\hat{\pi}[\Phi]$ is the barycentre of a probability measure η on H_e . Thus

$$\alpha_0 = \int d\eta(\rho) \left(\int k_0 d\rho \right).$$

Since $\int k_0 d\rho \geq \alpha_0$ for $\rho \in H_e$, we must have $\int k_0 d\rho = \alpha_0$ for η -a.s. ρ , proving the claim. QED

4. Proof of the main results

Our main result is the following:

Theorem 4.1 *The constrained problem above has an optimal SSRS $\gamma[\Phi]$, $\Phi = \Pi \hat{\Phi}_i$, satisfying: For all but at most m values of i , $\hat{\Phi}_i$ is a Dirac measure. For the remaining i , it is a convex combination of $r_i \geq 2$ Dirac measures where the integers $\{r_i\}$ satisfy $\sum_i r_i \leq m$.*

The proof proceeds in several steps. It helps at times to view G, H as subsets of the topological vector space of signed measures on $S \times D$. Let $m = 1$ for the time being. Let $\hat{\pi}[\Phi] \in H_e$ be such that $\int k_0 d\hat{\pi}[\Phi] = \alpha_0$. If $\hat{\pi}[\Phi]$ is an extreme point of G , $\hat{\pi}[\Phi] = \hat{\pi}\{\xi\}$ for some $\xi \in L$ by lemma 3.1 and we are done. Suppose not. Then it is a convex combination of two distinct elements ν_1, ν_2 of G . (Note that $\int k_1 d\nu_i < \infty$ for each i because $\int k_1 d\hat{\pi}[\Phi] < \infty$). If $\nu_1, \nu_2 \in H$, $\hat{\pi}[\Phi]$ cannot be in H_e . If $\nu_1, \nu_2 \in G \setminus H$, $\hat{\pi}[\Phi]$ cannot be in H because $G \setminus H$ is convex (recall $m = 1$). Thus exactly one of the ν_i 's, say ν_2 , is in $G \setminus H$. Then $\int k_1 d\nu_2 > \alpha_1$ and therefore $\int k_1 d\nu_1 < \alpha_1$. The value of $\int k_1 d\nu$ increases continuously from one strictly less than α_1 to one strictly exceeding α_1 as ν moves from ν_1 to ν_2 along the line segment joining the two. This segment must intersect the hyperplane $\{\nu \mid \int k_1 d\nu = \alpha_1\}$ at $\hat{\pi}[\Phi]$ because $\hat{\pi}[\Phi] \notin H_e$ otherwise. Now one can use a result of Dubins (1962) (see also Witsenhausen 1980, p. 265) to conclude that $\hat{\pi}[\Phi]$ is a convex combination of two distinct extreme points of \bar{G} . (Recall $H_e \subset \bar{H}_e$). If either of the latter assigns a strictly positive mass to $\{\infty\} \times D$, so would $\hat{\pi}[\Phi]$ itself. Thus both must be extreme points of G itself and hence are of the form $\hat{\pi}\{\xi_i\}$ for some $\xi_i \in L, i = 1, 2$. Let Z denote the line segment joining the two inclusive of the end points. Then $Z \subset P(S \times D)$ is compact.

The rest of the proof mimics Borkar (1991) closely with a few essential differences.

Lemma 4.1. *If $\hat{\pi}[\Phi]$ is a convex combination of $\hat{\pi}[\Phi_i], i = 1, 2$, in G , the latter must lie on Z .*

Proof. If not, the intersection of the rectangle formed by $\hat{\pi}[\Phi_i], \hat{\pi}\{\xi_i\}, i = 1, 2$, with the hyperplane $\{v \mid \int k_1 dv = \alpha_1\}$ is a line segment in H containing $\hat{\pi}[\Phi]$ in its relative interior. This contradicts the fact $\hat{\pi}[\Phi] \in H_e$, proving the claim. QED

For $i = 1, 2$, let $\xi_i = [\xi_i(1), \xi_i(2), \dots]$ and define $\xi'_n, n \geq 0$ by $\xi'_0 = \xi_1, \xi'_n = [\xi_2(1), \dots, \xi_2(n), \xi_1(n+1), \xi_1(n+2), \dots]$ for $n \geq 1$.

Lemma 4.2. $\hat{\pi}\{\xi'_n\} \in Z$ for $n \geq 0$.

Proof. We shall prove a more general result: For $\xi \in L$ such that $\xi(i) =$ either $\xi_1(i)$ or $\xi_2(i)$ for each $i, \gamma\{\xi\}$ is an SSS with $\hat{\pi}\{\xi\} \in Z$. Suppose that $p(1, \dots, \xi_1(1)) \neq p(1, \dots, \xi_2(1))$. [If not, take the least j for which $p(j, \dots, \xi_1(j)) \neq p(j, \dots, \xi_2(j))$.] Let ϕ_i denote the Dirac measures at $\xi_i(1), i = 1, 2$. From the proof of lemma 3.1, it is clear that for $\Phi_i = \phi_i \times \prod_{j=2}^\infty \hat{\Phi}_j, i = 1, 2, \hat{\pi}[\Phi_i], i = 1, 2$ are distinct and $\hat{\pi}[\Phi]$ is distinct from either and is a convex combination of the two. (By our assumption of stability under local perturbation, $\gamma[\Phi_i], i = 1, 2$, are SSRS.) By lemma 4.1, $\hat{\pi}[\Phi_i], i = 1, 2$, lie on Z . Repeat the argument with Φ_1 or Φ_2 in place of Φ and state 2 in place of state 1. It follows that for $\Phi' \in P(L)$ of the type: $\Phi' = \psi_1 \times \psi_2 \times \prod_{j=3}^\infty \hat{\Phi}_j$, where ψ_i are Dirac measures at either $\xi_1(i)$ or $\xi_2(i)$ for $i = 1, 2, \hat{\pi}[\Phi'] \in Z$. Iterating the argument, we have: for any $\Phi' = \prod \hat{\Phi}_i$, such that for some $n \geq 1, \hat{\Phi}'_i = \hat{\Phi}_i$ for $i > n$ and the Dirac measure at either $\xi_1(i)$ or $\xi_2(i)$ for $i \leq n$, we have $\hat{\pi}[\Phi'] \in Z$. Let Q denote the set of such Φ' . Then there exists a sequence $\{\Phi'_n\}$ in Q such that $\Phi'_n \rightarrow$ the Dirac measure at ξ as $n \rightarrow \infty$. Since Z is tight, we may drop to a subsequence if necessary and suppose that $\pi[\Phi'_n] \rightarrow \pi$ in $P(S)$ (and hence in total variation) for some π . Letting $n \rightarrow \infty$ in

$$\int p(\cdot, j, \cdot) d\hat{\pi}[\Phi'_n] = \hat{\pi}[\Phi'_n](\{j\} \times D), \quad j \in S,$$

we have

$$\sum \pi(i) p(i, j, \xi(i)) = \pi(j), \quad j \in S.$$

Thus $\pi = \pi\{\xi\}$. It is now easy to check that $\hat{\pi}[\Phi'_n] \rightarrow \hat{\pi}\{\xi\}$. Since Z is closed, the claim follows. QED

Consider Z as a union of two closed line segments Z_1 and Z_2, Z_1 (respectively Z_2) being the line segment joining $\hat{\pi}\{\xi_1\}$ (respectively $\hat{\pi}\{\xi_2\}$) with $\hat{\pi}[\Phi]$. By lemma 4.2, $\hat{\pi}\{\xi'_0\} \in Z_1$ and as $n \rightarrow \infty, \hat{\pi}\{\xi'_n\}$ moves from Z_1 to Z_2 . Then either $\hat{\pi}\{\xi'_n\} = \hat{\pi}[\Phi]$ for some n (and we are done) or there is an n such that $\hat{\pi}\{\xi'_{n-1}\} \in Z_1, \hat{\pi}\{\xi'_n\} \in Z_2$. But ξ'_{n-1}, ξ'_n differ only in their n th component and $\hat{\pi}[\Phi]$ is a convex combination of the two. Argue as in lemma 3.1 to conclude that $\hat{\pi}[\Phi] = \hat{\pi}[\hat{\Phi}]$ for $\hat{\Phi} = \prod \hat{\Phi}_j$ satisfying: $\hat{\Phi}_i$ is a Dirac measure at $\xi_1(i)$ (respectively $\xi_2(i)$) for $i > n$ (respectively $i < n$) and a convex combination of Dirac measures at $\xi_1(n), \xi_2(n)$ for $i = n$. This proves theorem 4.1 for $m = 1$. The general case follows by induction, replacing G at the n th induction step ($n < m$) by $G \cap \{\hat{\pi}[\Phi] \mid \int k_j d\hat{\pi}[\Phi] \leq \alpha_j, 1 \leq j \leq n\}$.

The following then follows immediately from standard Lagrange multiplier theory (Luenberger 1967, pp. 216-219).

Theorem 4.2 Suppose that (2) is a strict inequality for all $j, 1 \leq j \leq m$, and some $\hat{\pi}[\Phi'] \in H$. Then there exist numbers $\lambda_1, \dots, \lambda_m \in R^+$ such that the map

$$v \rightarrow \int k_0 dv + \sum_{i=1}^m \lambda_i \left(\alpha_i - \int k_i dv \right)$$

attains its minimum on G at the $\hat{\pi}[\Phi]$ of theorem 4.1. Moreover, the following holds for $v \in G$ and $\mu_1, \dots, \mu_m \in R^+$:

$$\begin{aligned} \int k_0 dv + \sum_{i=1}^m \lambda_i \left(\alpha_i - \int k_i dv \right) &\geq \int k_0 d\hat{\pi}[\Phi] + \sum_{i=1}^m \lambda_i \left(\alpha_i - \int k_i d\hat{\pi}[\Phi] \right) \\ &\geq \int k_0 d\hat{\pi}[\Phi] + \sum_{i=1}^m \mu_i \left(\alpha_i - \int k_i d\hat{\pi}[\Phi] \right). \end{aligned}$$

References

- Altman E, Shwartz A 1990 Sensitivity of constrained Markov decision processes, EE Pub. No. 741, Dept. of Electrical Eng., Technion, Haifa, Israel
- Beutler F J, Ross K W 1985 Optimal policies for controlled Markov chains with a constraint, *J. Math. Anal. Appl.* 112: 236–252
- Billingsley P 1968 *Convergence of probability measures* (New York: Wiley)
- Borkar V S 1989 Control of Markov chains with long-run average cost criterion: the dynamic programming equations. *SIAM J. Control Optim.* 27: 642–657
- Borkar V S 1991 Topics in controlled Markov chains, *Pitman research notes in mathematics* (Harlow: Longman) Chap. 7
- Dubins L 1962 On extreme points of convex sets, *J. Math. Anal. Appl.* 5: 237–244
- Hordijk A, Kallenberg L C M 1984 Constrained undiscounted stochastic dynamic programming. *Math. Oper. Res.* 9: 276–289
- Luenberger D 1967 *Optimization by vector space methods* (New York: Wiley)
- Phelps R 1966 *Lectures on Choquet's theorem* (New York: Van Nostrand)
- Ross K W 1989 Randomized and past-dependent policies for Markov decision processes with multiple constraints. *Oper. Res.* 37: 474–477
- Schwartz L 1961 *Disintegration of measures* (Bombay: Tata Institute of Fundamental Research)
- Witsenhausen 1980 Some aspects of convexity useful in information theory. *IEEE Trans. Inf. Theory* IT-26: 265–271