MATHEMATICS OF OPERATIONS RESEARCH
Vol. 17, No. 4, November 1992
Printed in U.S.A.

OPTIMAL STATIONARY POLICIES IN GENERAL STATE SPACE MARKOV DECISION CHAINS WITH FINITE ACTION SETS*

ROBERT K. RITT AND LINN I. SENNOTT

The result of Sennott [9] on the existence of optimal stationary policies in countable state Markov decision chains with finite action sets is generalized to arbitrary state space Markov decision chains. The assumption of finite action sets occurring in a global countable action space allows a particularly simple theoretical structure for the general state space Markov decision chain. Two examples illustrate the results. Example 1 is a system of parallel queues with stochastic work requirements, a movable server with controllable service rate, and a reject option. Example 2 is a system of parallel queues with stochastic controllable inputs, a movable server with fixed service rates, and a reject option.

1. Introduction. In this paper we consider general state space Markov decision chains (MDCs) with finite action sets and (possibly) unbounded costs. The result of Sennott [9] on the existence of optimal stationary policies in countable state space MDCs is generalized to this case.

Previous treatments of the general state space MDC (for example, Bertsekas and Shreve [2], Ross [6], and Hernandez-Lerma [4]) have assumed that the costs or rewards are bounded. In addition, the state space is generally assumed to be a Borel space, i.e., a Borel subset of a complete separable metric space, and the action space in a given state is assumed compact. There are difficult theoretical issues in dealing with compact action spaces, and in conjunction with the assumption of unbounded costs, these become formidable.

In this paper, we present a particularly simple theoretical structure for the general state space MDC with unbounded costs, under the assumption of finite action sets residing in a global countable action space. In the examples, we show that this framework allows treatment of interesting problems in the control of queueing systems.

In §2 we outline our theoretical structure, and in §3 the existence of discounted optimal stationary policies is proved. In §4 we prove the existence of expected average cost optimal stationary policies, and in §5 the examples are presented.

2. A simplified structure for general state space Markov decision chains. Let the state space X be an arbitrary measure space with σ-field $\mathscr{S}$. We require that singleton subsets of X be measurable. Let A denote the global action set, assumed to be nonempty and countable. Later, a restriction will be placed on the cost function insuring the existence of only finitely many feasible actions in each state. The assumption of a global countable action set allows a simplified structure for the set of sample paths under the process.

*Received June 1, 1990; revised August 20, 1991.
AMS 1980 subject classification. Primary: 90C47. Secondary: 90B22.
IAOR 1973 subject classification. Main: Programming: Markov Decision.
OR/MS Index 1978 subject classification. Primary: 119 Dynamic Programming/Markov/Infinite State. Secondary: 697 Queues/Networks.
Key words. General state space Markov decision chains, average cost optimal stationary policies, control of queueing systems.

0364-765X/92/1704/0901/$01.25

Given a state $x \in X$ and action $a \in A$, the next state is chosen according to a stochastic kernel $P(B|x, a)$, where $B \in \mathcal{S}$. Thus: (i) for every x and a, $P(\cdot|x, a)$ is a probability measure on $\mathcal{S}$, and (ii) for every $B \in \mathcal{S}$, $P(B|\cdot)$: $XA \to [0, 1]$ is measurable with respect to the σ-field $\mathcal{S} \times 2^A$.

Let μ be a probability measure (called the initial measure) on $(X, \mathcal{S})$. Usually, μ will be the point probability measure concentrated at the known initial state, but for now μ is completely general. Informally, a policy δ is a rule for choosing actions, given the history of the process. It may be randomized. The process operates under δ as follows: The initial state x_0 is chosen according to μ. Then the initial action a_0 is chosen according to the rule δ. Then the next state x_1 is chosen according to $P(\cdot|x_0, a_0)$. Then a_1 is chosen according to δ, and this process is continued indefinitely.

To rigorously define a policy δ, let $X_0, X_1, X_2, \ldots$ be copies of X and $A_0, A_1, A_2, \ldots$ be copies of A. Define the n-stage history H_n inductively by $H_0 = X_0$, and given that H_n is defined, $H_{n+1} = H_n A_n X_{n+1}$. Thus H_n represents the information available for decision making at stage n. Let $H = H_0 \cup H_1 \cup H_2 \cup \cdots$ and let $\mathcal{P}$ be the set of probability distributions on A.

DEFINITION 1. A policy is a map δ: $H \to \mathcal{P}$ such that for a fixed n and $a \in A$, $(\delta|H_n)(a)$ is measurable on H_n with respect to the product σ-field $\mathcal{S} \times 2^A \times \cdots \times \mathcal{S}$.

To insure that the process operates in a well-defined manner, Theorem 2.7.2 of Ash [1] may be applied to define a probability measure on the set of sample paths $\Omega = X_0 A_0 X_1 A_1 \ldots$ under the policy δ. Equip Ω with the σ-field $\mathcal{S} \times 2^A \times \mathcal{S} \times 2^A \times \cdots$ with initial measure μ on $(X_0, \mathcal{S})$. Given $(x_0, a_0, \ldots, x_n, a_n)$ we need a measure on $\mathcal{S}$. That measure is $P(\cdot|x_n, a_n)$. Given $(x_0, a_0, \ldots, x_n)$ we need a measure on 2^A. That measure is induced by the probability distribution $\delta(x_0, a_0, \ldots, x_n)$, where for notational convenience the extra set of parentheses has been omitted. The hypotheses of Theorem 2.7.2 [1] are met, and hence there is a probability measure on the set of sample paths of the process.

This theoretical structure is basically that of Hernandez-Lerma [4]; however, we do not assume that the state space is a Borel space. In addition, because the cost function will be allowed to be extended real-valued, infinite costs can be used to render certain actions infeasible in certain states. This further simplifies the structure of the set of sample paths as presented in [4].

The important special classes of randomized stationary policies and stationary policies may be defined as in [4]. In particular, a stationary policy is defined by a measurable map f: $X \to A$ such that whenever $x_n = x$, the policy chooses action $f(x)$. This policy will be identified with the map f.

ASSUMPTION 1. *The cost function is a measurable map* C: $XA \to [0, +\infty]$ *such that, for each* $x \in X$, $\{a|C(x, a) < +\infty\}$ *is finite.*

If an action a is infeasible in x, set $C(x, a) = +\infty$. As discussed above, this allows the theoretical structure to be simplified. In a slight abuse of notation, the cost function evaluated at the point x under a stationary policy f will be denoted $C(x, f)$.

3. The existence of optimal stationary discounted policies. For a discount factor α, where $0 < \alpha < 1$, we now define the total discounted cost under a policy δ. For $n \geqslant 0$, define C_n: $\Omega \to [0, +\infty]$ by $C_n(x_0, a_0, \ldots) = C(x_n, a_n)$. These functions are measurable on Ω, and hence the function $\Sigma_n \alpha^n C_n$ is measurable on Ω. Conditioned on $x_0 = x$, we denote its integral (expectation) by $V_\alpha(\delta, x)$. This quantity may be $+\infty$ for some policies. Let $V_\alpha(x) = \inf_\delta V_\alpha(\delta, x)$, where the infimum is taken over all policies.

ASSUMPTION 2. *$V_\alpha(x) < \infty$ for each x and $0 < \alpha < 1$.*

Given an initial state x, by Assumption 2, there must be at least one action a such that $C(x, a) < +\infty$. Hence the finite set postulated in Assumption 1 is nonempty.

In proving the existence of optimal discounted stationary policies, it is usual to appeal to a "measurable selection" theorem that guarantees the existence of a (measurable) stationary policy realizing the minimum in the discount optimality equation. Under the assumptions in this paper, measurable selectors may be easily constructed as follows.

LEMMA 1. *Assume a family of measurable functions h_a: $X \to [0, +\infty]$, for $a \in A$, such that $\{a|h_a(x) < +\infty\}$ is finite, for $x \in X$. Let $h = \min_a h_a$. Order the elements of A as $a_0 < a_1 < a_2 < \cdots$, and define f: $X \to A$ by $f(x) =$ smallest a such that $h_a(x) = h(x)$. Then f is measurable.*

PROOF. The function h is measurable [1]. We must show that $f^{-1}(\{a_i\})$ is measurable. But

$$f^{-1}(\{a_i\}) = \bigcap_{j=0}^{i-1} \{x|h_{a_i}(x) < h_{a_j}(x)\} \cap \bigcap_{j \geqslant i+1} \{x|h_{a_i}(x) \leqslant h_{a_j}(x)\},$$

which is measurable. If for some x, $h(x) = +\infty$, the result also holds.

Before proving the existence of a discounted optimal stationary policy, we require two further lemmas.

LEMMA 2. (i) *If h is measurable on X, then for any a, $\int h(y)P(dy|x, a)$ is measurable on X.*

(ii) *Let Y be a copy of X. If h is measurable on XY, then for any a, $\int h(x, y)P(dy|x, a)$ is measurable on X.*

(iii) *If h is measurable on X, and if f: $X \to A$ is a stationary policy, then $\int h(y)P(dy|x, f)$ is measurable on X.*

PROOF. Part (i) is a special case of (ii). It is sufficient to prove (ii) for simple functions [1]. Let

$$h = \sum_{j=1}^{J} r_j I_{B_j} I_{D_j},$$

where B_j and D_j are measurable. Then

$$\int h(x, y)P(dy|x, a) = \sum_{j=1}^{J} r_j I_{B_j}(x) P(D_j|x, a),$$

which is measurable on X.

By an argument involving simple functions, to prove (iii), it is sufficient to prove that $P(B|x, f(x))$ is measurable on X, for each measurable B. Let g: $X \to XA$ be the measurable function defined by $g(x) = (x, f(x))$. Then $P(B|x, f(x)) = P(B|g(x))$, a composition of measurable functions and hence measurable.

We will need the concept of a "shifted policy".

DEFINITION 2. For a policy δ and fixed (x, a), the shifted policy $\delta(x, a)$ is defined by $\delta(x, a)(x_0, a_0, \ldots, x_n) = \delta(x, a, x_0, a_0, \ldots, x_n)$.

Informally, the decision made by policy $\delta(x, a)$ at stage n is the same as the decision made by δ at stage $n + 1$, given that the initial state and initial decision were (x, a).

LEMMA 3. *For each policy δ, the value function $V_\alpha(\delta, x)$ is measurable on X. Moreover, if Y is a copy of X, then for a fixed a, $V_\alpha(\delta(x, a), y)$ is measurable on XY.*

PROOF. To prove the first statement, it is sufficient to prove that $E_\delta(C_k|x_0 = x)$ is measurable on X for $k \geqslant 0$. Now

$$E_\delta(C_k|x_0 = x) = \sum_a \delta(x)(a) \int_{X_1} \sum_{a_1} \delta(x, a, x_1)(a_1) P(dx_1|x, a)$$

$$\cdots \int_{X_{k-1}} \sum_{a_{k-1}} \delta(x, a, x_1, \ldots, x_{k-1})(a_{k-1}) P(dx_{k-1}|x_{k-2}, a_{k-2})$$

$$\int_{X_k} \sum_{a_k} C(x_k, a_k) \delta(x, a, x_1, \ldots, x_k)(a_k) P(dx_k|x_{k-1}, a_{k-1}).$$

Working from right to left, using Definition 1 and Lemma 2, we may show that this is measurable in x.

We may similarly write out $E_{\delta(x,a)}(C_k|x_0 = y)$ and show that this is measurable on XY.

The theorem giving an α-discount optimal stationary policy may now be proved.

THEOREM 1. *Let Assumptions 1 and 2 hold. The (finite) value function V_α is the minimal nonnegative measurable solution of the optimality equation*

$$(1) \qquad V_\alpha(x) = \min_a \left\{ C(x, a) + \alpha \int V_\alpha(y) P(dy|x, a) \right\}, \qquad x \in X.$$

If for each x, we choose $f(x)$ to be the smallest action that realizes the minimum on the right of (1), then f is an α-discount optimal policy.

PROOF. Fix $0 < \alpha < 1$ and inductively define v_n: $X \to [0, +\infty]$ as follows. Let $v_0 \equiv 0$, and given that v_{n-1} has been defined, let

$$(2) \qquad v_n(x) = \min_a \left\{ C(x, a) + \alpha \int v_{n-1}(y) P(dy|x, a) \right\}, \qquad x \in X.$$

The following will be proven: (a) Each v_n is measurable. (b) $v_n(x)$ is increasing in n, for $x \in X$. (c) $v_n \leqslant V_\alpha(\delta, \cdot)$ for $n \geqslant 0$ and all policies δ.

Part (a) follows by induction on n, using Lemma 2(i) and the fact that the infimum of a countable number of measurable functions is measurable. Part (b) follows easily by induction on n. To prove (c), let $V_n(\delta, x)$ be the expected discounted cost if the process is operated under δ for n steps, and then terminated with a terminal cost of 0. We prove: (*)$v_n \leqslant V_n(\delta, \cdot)$ for all policies δ and $n \geqslant 0$. Since the costs are nonnegative, this will prove (c). Clearly $v_1(x) \leqslant \sum_a C(x, a)\delta(x)(a) = V_1(\delta, x)$. Assume (*) is true for $n - 1$. It may be seen that

$$(3) \qquad V_n(\delta, x) = \sum_a \delta(x)(a) \left\{ C(x, a) + \alpha \int V_{n-1}(\delta(x, a), y) P(dy|x, a) \right\},$$

where $\delta(x, a)$ is the shifted policy of Definition 2. By the proof of Lemma 3, it follows that $V_n(\delta(x, a), y)$ is measurable on XY and hence by Lemma 2(ii), the integration in (3) produces a measurable function on X. For (x, a) *fixed*, $\delta(x, a)$ is a policy. Hence

by the induction hypothesis, $v_{n-1}(y) \leqslant V_{n-1}(\delta(x, a), y)$. Thus for each a,

$$v_n(x) \leqslant C(x, a) + \alpha \int V_{n-1}(\delta(x, a), y) P(dy|x, a)$$

and the result follows from (3).

Since the v_n form an increasing sequence of measurable functions that is bounded above (by (c) and Assumption 2), $\lim_n v_n \equiv u$ exists and is measurable and finite. Since the minimization on the right of (2) is over a finite set, the monotone convergence theorem applied to (2) proves that u satisfies

$$u(x) = \min_a \left\{ C(x, a) + \alpha \int u(y) P(dy|x, a) \right\}, \qquad x \in X. \tag{4}$$

By Lemma 1, there exists a stationary policy f realizing the minimum on the right of (4). Hence (4) may be written

$$u(x) = C(x, f) + \alpha \int u(y) P(dy|x, f), \qquad x \in X. \tag{5}$$

Note that, by Lemma 2(iii), the integral in (5) produces a measurable function on X. Iterating (5) yields $u(x) \geqslant V_n(f, x)$, and thus $u \geqslant V_\alpha(f, \cdot)$. But by (c), this proves that $u = V_\alpha$ and that f is α-discount optimal.

Let w be a nonnegative measurable solution of (1) and let e be a stationary policy realizing the minimum on the right of (1) for w. Applying the argument just given yields $w \geqslant V_\alpha(e, \cdot) \geqslant V_\alpha$, and hence V_α is the minimum solution.

4. Average cost optimal stationary policies. Given a policy δ and initial state x, define the long-run expected average cost ("average cost") under δ as

$$J(\delta, x) = \limsup_{n \to \infty} \frac{E_\delta(\sum_{t=0}^n C(x_t, a_t)|x_0 = x)}{n + 1}.$$

Our goal is to find a constant J and a stationary policy f such that $J = J(f, x) \leqslant J(\delta, x)$ for every policy δ and initial state x. Such a policy f is said to be (expected) average cost optimal with (expected) average cost J.

For each discount factor α, let s_α be a nonnegative real number and define $h_\alpha(x) = V_\alpha(x) - s_\alpha$. (In the examples, $s_\alpha = V_\alpha(0)$, but, for the treatment of later examples, it will be useful to have the additional flexibility of an arbitrary s_α.) We utilize the assumptions of [9] as slightly modified by Cavazos-Cadena [3].

ASSUMPTION 3. *There exists a measurable function* m: $X \to [0, +\infty)$ *such that* $h_\alpha(x) \leqslant m(x)$, *for* $0 < \alpha < 1$. *In addition, there exist* $x_0 \in X$ *and* $a_0 \in A$ *such that* $C(x_0, a_0) + \int m(y) P(dy|x_0, a_0) < \infty$.

ASSUMPTION 4. *There exists a nonnegative number* N *such that* $-N \leqslant h_\alpha(x)$, *for* $x \in X$ *and* $0 < \alpha < 1$.

In the case of compact action spaces, and with additional assumptions on the state space, theorems similar to the following were proved by Hernandez-Lerma and Lasserre [5] and Schal [7]. Our theorem 2 and those just mentioned were obtained independently.

THEOREM 2. *Assume that Assumptions* 1 *through* 4 *hold. Then there exist a constant* J *and measurable function* h, *with* $-N \leqslant h(x) \leqslant m(x)$, *such that*

$$J + h(x) \geqslant \min_a \left\{ C(x,a) + \int h(y) P(dy|x,a) \right\}, \qquad x \in X. \tag{6}$$

If for each x, $f(x)$ *is chosen to be the smallest action realizing the minimum on the right of* (6), *then* f *is expected average cost optimal with expected average cost* J. *Moreover*,

$$\lim_{\alpha \uparrow 1} (1-\alpha) V_\alpha(x) = J \quad \text{for } x \in X.$$

PROOF. Equation (1) may be written

$$(1-\alpha)s_\alpha + h_\alpha(x) = \min_a \left\{ C(x,a) + \alpha \int h_\alpha(y) P(dy|x,a) \right\}, \qquad x \in X. \tag{7}$$

From (7) and Assumptions 3 and 4, it follows that

$$0 \leqslant (1-\alpha)s_\alpha \leqslant C(x_0,a_0) + \int m(y) P(dy|x_0,a_0) + N. \tag{8}$$

Because the right side of (8) is finite, there exist a sequence of discount factors $\alpha_n \uparrow 1$ and a constant J such that $\lim_{n\to\infty}(1-\alpha_n)s_{\alpha_n} = J$. We claim that, for every x,

$$\lim_{n\to\infty} (1-\alpha_n) V_{\alpha_n}(x) = J. \tag{9}$$

To show (9), observe that $0 \leqslant |h_\alpha(x)| \leqslant \max\{m(x), N\}$, hence $\lim_{\alpha\uparrow 1}(1-\alpha)h_\alpha(x) = 0$. Thus

$$\lim_{n\to\infty} (1-\alpha_n) V_{\alpha_n}(x) = \lim_{n\to\infty} (1-\alpha_n) h_{\alpha_n}(x) + \lim_{n\to\infty} (1-\alpha_n) s_{\alpha_n} = J.$$

Now let

$$h = \liminf_{n\to\infty} h_{\alpha_n}.$$

Then h is measurable [1] and $-N \leqslant h(x) \leqslant m(x)$ for $x \in X$. Consider (7) with $\alpha = \alpha_n$. Take the limit infimum of both sides. Since the limit infimum of a minimum of finitely many sequences is equal to the minimum of the limit infimums, the lim inf may be passed through the minimization. Then apply Fatou's Lemma to obtain (6).

By Lemma 1, a stationary policy f that realizes the right side of (6) exists. Lemma A1 of [9], which carries over directly to the general state space case, yields $J(f,x) \leqslant J$ for every x. But by Lemma A2 of [9], we obtain $J \leqslant J(\delta, x)$, for every policy δ, and hence f is expected average cost optimal with expected average cost J. It can be shown that $\lim_{\alpha\uparrow 1}(1-\alpha)V_\alpha(x) = J$ as in [9].

Note that this proof is somewhat simpler than the countable state space proof given in [9].

5. Examples. In the examples, distribution functions on R will be induced by Lebesgue-Stieltjes measures ([1, §1.4]).

EXAMPLE 1. Parallel queues with stochastic work requirements and a single movable server. Consider a system of J parallel queues. In each time slot, a

nonnegative amount of work is chosen for each queue, with the work requirement for queue j chosen from distribution F_j. The state space X is the nonnegative orthant of R^J with state $x = (x_1, \ldots, x_J)$, where x_j is the amount of work current in queue j. At the beginning of each slot, the server must decide whether to admit work to each of the queues. Admitted work is not available for service until the beginning of the next slot. In addition, the server must decide to either remain idle or to serve one of the (nonempty) queues. If queue k contains work, the server may choose to serve it at rate $\beta(k) \in B_k$, a nonempty finite set of nonzero service rates.

Let e be a J-tuple such that $e_j = 1$ (respectively, 0) if new work is admitted (respectively, not admitted) to queue j. Let i represent the decision to remain idle. In state 0, the possible decisions are (e, i), and if there is current work in the system, the possible decisions are (e, i) or $(e, \beta(k))$, where this denotes serving the kth queue (if nonempty) at rate $\beta(k) \in B_k$. The global action set is given by $A = \{0,1\}^J \times (\cup_j B_j \cup \{i\})$.

Observe that when $J = 1$, this model reduces to a single-server queue with stochastic work requirement, controllable service rates, and the option to reject new work and/or to idle the server.

To write the discount optimality equations requires some further notation. Let $y = (y_1, \ldots, y_J)$ be a vector of new work. The notation ey means coordinatewise multiplication; thus $(ey)_j$ is 0 if $e_j = 0$ and y_j otherwise. Let $E(e) = \{j | e_j = 1\}$. If $x_k > 0$, then $x(\beta(k))$ is a vector with $(x(\beta(k)))_j = x_j$ $(j \neq k)$ and $(x(\beta(k)))_k = (x_k - \beta(k))^+$. The discount optimality equations are

$$V_\alpha(0) = \min_e \left\{ C(0,(e,i)) + \alpha \int_0^\infty \cdots \int_0^\infty V_\alpha(ey) \prod_{j \in E(e)} dF_j(y_j) \right\},$$

$$V_\alpha(x) = \min_e \left\{ C(x,(e,i)) + \alpha \int_0^\infty \cdots \int_0^\infty V_\alpha(x + ey) \prod_{j \in E(e)} dF_j(y_j) \right\}$$

$$\wedge \min_e \min_k \min_{\beta(k)} \left\{ C(x,(e,\beta(k))) \right.$$

$$\left. + \alpha \int_0^\infty \cdots \int_0^\infty V_\alpha(x(\beta(k)) + ey) \prod_{j \in E(e)} dF_j(y_j) \right\}, \tag{10}$$

where the minimization is taken over all k such that $x_k > 0$. If $E(e) = \varnothing$, the integration is understood to produce the single value given in the integrand.

PROPOSITION 1. *Assume*:

(1) *The cost function C is measurable on XA (with infeasible actions assigned cost $+\infty$).*

(2) *For each vector e and state x, the cost $C(x,(e,i))$ is increasing in each coordinate of x, with the other coordinates held fixed.*

(3) *Fix e and service rate $\beta(k)$ and assume $x_k > 0$. Then $C(x,(e,\beta(k)))$ is increasing in each coordinate of x, with the others held fixed. Moreover, if x^* is a vector identical to x but with $x_k^* = 0$, then $C(x,(e,\beta(k))) \geq C(x^*,(e,i))$.*

Then there exists an expected average cost optimal stationary policy.

PROOF. To verify Assumption 2, let e^* be the vector with all 0 coordinates. Let f be the policy that always chooses e^* and that serves the current work in the queues in some fixed order and for some fixed service rates (for example, it may specify serving

the first queue until its work reaches 0, then the second, etc., skipping any queues that are empty). Beginning in state x, the process will clearly reach state 0 in finitely many steps with finite (undiscounted) cost, say c_{x0}. From (10) it follows that $V_\alpha(0) \leqslant C(0,(e^*,i))/(1-\alpha) < \infty$ and hence, by the above reasoning, $V_\alpha(x) < \infty$ for $x \in X$.

Now let $s_\alpha = V_\alpha(0)$. Using the policy f and the reasoning in Proposition 1 of [9], it follows that $h_\alpha(x) \leqslant c_{x0}$ and Assumption 3 holds. It remains to verify Assumption 4. Consider the n-step minimal expected costs defined in (2). Using the form of (10), it may be shown by induction on n that $v_n(x)$ is increasing in each coordinate of x, with the other coordinates held fixed. Since $v_n \uparrow V_\alpha$, this also holds for V_α. From this, it easily follows that $V_\alpha(x) \geqslant V_\alpha(0)$. Hence Assumption 4 holds with $N = 0$ and the proof is complete.

As an example of this cost structure, let $H(x)$ be a measurable holding cost, increasing in each coordinate of x, with the others held fixed, and satisfying $H(0) = 0$. Assume $R(e)$ is the cost of the decision vector e. Further, assume no cost for idling the server and a cost of $Q(\beta(k))$ of serving at rate $\beta(k)$. Then the cost structure given by $C(x,(e,i)) = H(x) + R(e)$ and $C(x,(e,\beta(k))) = H(x) + R(e) + Q(\beta(k))$, if $x_k > 0$, satisfies the conditions of Proposition 1.

EXAMPLE 2. Parallel queues with stochastic controllable inputs and a single movable server. This is similar to Example 1, except that the inputs, rather than the service rates, are controllable. Consider a system of J parallel queues. The state space X is the nonnegative orthant of R^J with state $x = (x_1, \ldots, x_J)$, where x_j is the amount of work currently in queue j. In each slot, the server must decide to either remain idle or to serve one of the (nonempty) queues. Let i represent the decision to remain idle. If it chooses to serve queue k, it serves at fixed rate $\beta(k)$.

In addition, the server must decide whether or not to admit work to each of the queues, and if work is admitted, which distribution will be used. If work is admitted to queue j, it is chosen from nonnegative distribution $F_{a(j)}$, where $a(j) \in A_j$, a nonempty finite set of parameters. Let e be a J-tuple such that $e_j = 0$ if new work is not admitted to queue j and $e_j = a(j)$ if new work is admitted and chosen from distribution $F_{a(j)}$. Let $E(e) = \{j | \text{work is admitted to queue } j\}$.

In state 0, the possible decisions are (e,i), and if there is current work in the system, the possible decisions are (e,i) or (e,k), where this denotes serving the kth queue (if nonempty) at fixed rate $\beta(k)$. The action set is given by $A = \prod_j(\{0\} \cup A(j)) \times (J \cup \{i\})$.

Let $y(e)$ be a vector of new work. That is, $(y(e))_j = 0$ if $e_j = 0$ and $(y(e))_j = y_j$ chosen from $F_{a(j)}$ if $e_j = a(j)$. The vector $x(k)$ is defined analogously to the vector $x(\beta(k))$ in Example 1. Note that since there is a single service rate available for queue k, we may suppress the β. The discount optimality equations are

$$V_\alpha(0) = \min_e \left\{ C(0,(e,i)) + \alpha \int_0^\infty \cdots \int_0^\infty V_\alpha(y(e)) \prod_{j \in E(e)} dF_{a(j)}(y_j) \right\},$$

$$V_\alpha(x) = \left(\min_e \left\{ C(x,(e,i)) + \alpha \int_0^\infty \cdots \int_0^\infty V_\alpha(x + y(e)) \prod_{j \in E(e)} dF_{a(j)}(y_j) \right\} \right)$$

$$\wedge \left(\min_e \min_k \left\{ C(x,(e,k)) + \alpha \int_0^\infty \cdots \int_0^\infty V_\alpha(x(k) + y(e)) \prod_{j \in E(j)} dF_{a(j)}(y_j) \right\} \right), \tag{11}$$

where the minimization is over all k for which $x_k > 0$.

PROPOSITION 2. *Assume*:

(1) *The cost function C is measurable on XA (with infeasible actions assigned cost $+\infty$).*

(2) *For each vector e and state x, the cost $C(x,(e,i))$ is increasing in each coordinate of x, with the other coordinates held fixed.*

(3) *Fix e and k and assume $x_k > 0$. Then $C(x,(e,k))$ is increasing in each coordinate of x, with the others held fixed. Moreover, if x^* is a vector identical to x but with $x_k^* = 0$, then $C(x,(e,k)) \geqslant C(x^*,(e,i))$.*

Then there exists an expected average cost optimal stationary policy.

PROOF. Minor changes in the proof of Proposition 1 yield the result.

As an application of this structure, assume for the moment that $J = 1$. Consider a finite collection of tasks of varying complexity. At each time unit, the controller has the option of either selecting a new task or of not selecting a task. If a new task is selected, it brings a certain amount of work into the system; this work requirement is drawn from a specified distribution (which may be concentrated on a fixed value). In addition, the controller has the option of remaining idle or of serving the accumulated work at a fixed rate.

Acknowledgements. The authors would like to thank an anonymous referee for pointing out an error in an earlier version of the examples. We would also like to thank the area editor, Richard Serfozo, for his help.

References

[1] Ash, R. B. (1972). *Real Analysis and Probability*. Academic Press, New York.

[2] Bertsekas, D. P. and Shreve, S. E. (1978). *Stochastic Optimal Control: The Discrete Time Case*. Academic Press, New York.

[3] Cavazos-Cadena, R. (1991). Recent Results on Conditions for the Existence of Average Optimal Stationary Policies, *Ann. Oper. Res.* **28** 3–28.

[4] Hernandez-Lerma, O. (1989). *Adaptive Markov Control Processes*. Springer-Verlag, New York.

[5] _____ and Lasserre, J. B. (1990). Average Cost Optimal Policies for Markov Control Processes with Borel State Space and Unbounded Costs. *System Control Lett.* **15** 349–356.

[6] Ross, S. M. (1968). Arbitrary State Markovian Decision Processes. *Ann. Math. Statist.* **39** 2118–2122.

[7] Schal, M. (forthcoming). Average Optimality in Dynamic Programming with General State Space. *Math. Oper. Res.*

[8] Sennott, L. I. (1986). A New Condition for the Existence of Optimal Stationary Policies in Average Cost Markov Decision Processes. *Oper. Res. Lett.* **5** 17–23.

[9] _____ (1989). Average Cost Optimal Stationary Policies in Infinite State Markov Decision Processes with Unbounded Costs. *Oper. Res.* **37** 626–633.

DEPARTMENT OF MATHEMATICS, ILLINOIS STATE UNIVERSITY, NORMAL, ILLINOIS 61761