Optimal Control of Two Interacting Service Stations

BRUCE HAJEK, MEMBER, IEEE

Abstract—**Optimal controls described by switching curves in the two-dimensional state space are shown to exist for the optimal control of a Markov network with two service stations and linear cost. The controls govern routing and service priorities. Finite horizon and long run average cost problems are considered and value iteration is a key tool.**

Nonconvex value functions are shown to exist for slightly more general networks. Nonconvex value functions are also shown to arise for a simple single station control problem in which the instantaneous cost is convex but not monotone. Nevertheless, optimality of threshold policies is established for the single station problem. The proof is based on a novel use of stochastic coupling and policy iteration.

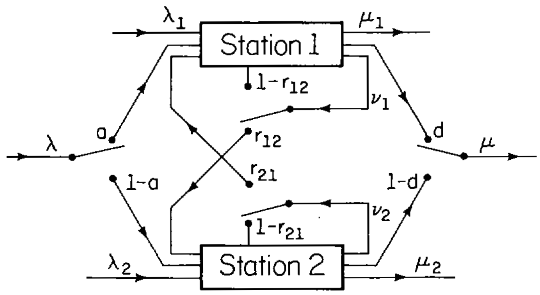


Fig. 1. A two-station network with parameters λ, λ_1, λ_2, μ, μ_1, μ_2, ν_1, and ν_2 and control $z = (a, d, r_{12}, r_{21})$ in $[0.1]^4$.

I. Introduction

The main model considered in this paper is pictured in Fig. 1 and is described in detail in the second section. Using little more than the well-known inductive dynamic programming approach (see [24, p. 282]) we establish the existence of optimal controls which have a switch curve structure. As a side product, convexity properties are established for the value functions. Our convexity proof is an alternative to the convex duality method applied in [19].

We consider both finite horizon and long run average cost problems with linear (hence, unbounded) instantaneous cost. The long run average cost problem is treated by direct passage to the limit in the finite horizon problem as in [19]. A related approach has been used by Borkar [3] (also, see Remark at the end of Section IV).

Lest we get too optimistic, an example of a slightly more general network is given in Section V for which nonconvex value functions appear. As indicated for a single station system in Section VI, the inductive approach can still be used to establish switch structure, but rather delicate inductive hypotheses appear to be needed. An apparently new approach, based on policy iteration and stochastic coupling, is also given in Section VI to establish the optimality of threshold policies for the single station system.

If certain parameters in our model are set equal to zero, then any one of three previously studied models can be obtained. First, if λ_1, λ_2, and μ are the only nonzero parameters, then a service priority problem is obtained. For this problem it was shown by Harrison [10] that the optimal rule, even with nonexponential service, is a fixed priority rule (equivalently, the switch curve lies on the boundary of the state space). Priorities are assigned on the basis of an easily calculated index for each class. Problems of this sort were studied in more generality by Gittens [8], Nash [17], Whittle [28], and Varaiya, Walrand, and Buyukkoc [26]. Lin and Kumar [14] consider a related problem in which two servers work on a single queue.

Second, if λ_1, ν_1, and μ_1 are the only nonzero parameters, then a control problem for tandem queues is obtained. If the holding cost per customer is the same in each queue, then Sobel [23] has shown that the long run average cost is minimized by a full-service policy, even for some nonexponential service distributions, for arbitrary arrival processes, and for more general networks. Rosberg, Varaiya, and Walrand [19] demonstrate that an optimal policy with switch structure exists when the holding cost per customer is larger in the second queue.

Finally, if λ, λ_1, λ_2, μ_1, and μ_2 are the nonzero parameters, then a routing problem is obtained. For this problem Foschini and Salz [6] revealed the striking effectiveness of state feedback policies in the case of heavy traffic. Assuming that the service-time distribution is the same for both nodes, Winston [29] and Weber [27] have shown that the policy of joining the shortest queue is strongly optimal in that it minimizes the discounted number of departures up to any time t. Ephremides, Varaiya, and Walrand [5] obtained similar results.

Ghoneim [7] established the optimality of switch controls for a similar model in which exogenous arrivals are controlled at each of two queues, and one of the two queues feeds into the other.

II. The Basic Control Problem

Consider the system of two interacting service stations pictured in Fig. 1. We will first informally describe the system, and then give a mathematical formulation. The evolution of the system is influenced by a possibly time-varying, state-dependent control with values $z = (a, d, r_{12}, r_{21})$ in $[0,1]^4$.

New customers arrive at station 1 in a Poisson stream of rate λ_1 and at station 2 in a Poisson stream of rate λ_2. A third sequence of arrivals occurs in a Poisson stream of rate λ. A customer in this stream arriving at time t is routed to station 1 with probability $a(t)$ and to station 2 otherwise.

Manuscript received November 10, 1982; revised June 29, 1983. Paper recommended by P. R. Kumar, Past Chairman of the Stochastic Control Committee. This work was supported by the Office of Naval Research under Contract N00014-82-K-0359. This paper was presented in part at the IEEE Decision and Control Conference, 1982.

The author is with the Department of Electrical Engineering and the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801.

0018-9286/84/0600-0491$01.00 ©1984 IEEE

Customers are served at station 1 by an exponential server at rate μ_1 and they are served at station 2 by an exponential server at rate μ_2. In addition, the stations share an exponential server of rate μ which devotes a proportion of effort $d(t)$ to station 1 and proportion $1-d(t)$ to station 2. After being served by one of these three servers a customer departs from the system.

Finally, there are two more exponential servers—one of rate ν_1 at station 1 and one of rate ν_2 at station 2. If at time t such server at station 1 (respectively, station 2) completes service for a customer, that customer is immediately routed to the other station with probability $r_{12}(t)$ [respectively, with probability $r_{21}(t)$], and otherwise it remains in the same station.

A semi-Markov decision model will now be introduced to rigorously define our problem. (See Stidham and Prabhu [24], Serfozo [22], and Bertsekas [1] for a general background.) First, define mappings from $\mathscr{S}$ to $\mathscr{S}$, where $\mathscr{S}=\mathbb{Z}_+^2$, which correspond to state transitions. For $i=1$ or 2, let A_i and D_i correspond to an arrival or *potential* departure at station i and let R_{ij} correspond to a *potential* transfer from queue i to queue j. (Use of potential transitions gives rise to a constant total event rate [16], [21].) For example $A_1x=(x_1+1,x_2)$, $D_1x=((x_1-1)^-,x_2)$, and $R_{12}x=(x_1-1,x_2+1)$ if $x_1>0$ and $R_{12}x=x$ if $x_1=0$. Define the "total event rate" γ by

$$\gamma=\lambda+\lambda_1+\lambda_2+\mu+\mu_1+\mu_2+\nu_1+\nu_2$$

and for each control value $z=(a,d,r_{12},r_{21})$ define the transition probability function $p(\cdot|\cdot,z)$ on $\mathscr{S}\times\mathscr{S}$ by

$$\begin{aligned}p(y|x,z)\gamma=&(\lambda a+\lambda_1)I\{y=A_1x\}\\&+(\lambda(1-a)+\lambda_2)I\{y=A_2x\}\\&+(\mu d+\mu_1)I\{y=D_1x\}\\&+(\mu(1-d)+\mu_2)I\{y=D_2x\}\\&+\nu_1r_{12}I\{y=R_{12}x\}+\nu_2r_{21}I\{y=R_{21}x\}.\end{aligned}$$

A control policy $\boldsymbol{u}$ is a sequence $(u_0,u_1,\cdots)$ such that u_k is a function from $\mathscr{S}^{k+1}$ to $[0,1]^4$ for each k. Given a control policy $\boldsymbol{u}$ and an initial state x in $\mathscr{S}$, a semi-Markov decision process $(x(t): t\geq 0)$ with jump rate γ and imbedded transition probabilities p is given by

$$x(t)=x_{n(t)}$$

where $\boldsymbol{x}=(x_0,x_1,\cdots)$ is a sequence of random variables with $P(x_0=x)=1$ and

$$P[x_{k+1}=j|x_k=i_k,\cdots,x_0=i_0]=p(j|i_k,u_k(i_0,\cdots,i_k)) \tag{2.1}$$

and $(n(t): t>0)$ is a rate γ Poisson process which is independent of $\boldsymbol{x}$. We use the notation E_x^u to denote expectations relative to $(x(t))$.

We will assume that the instantaneous cost is a linear combination $c(x)=c'x=c_1x_1+c_2x_2$ where c_1 and c_2 are nonnegative, and that a nonnegative discount rate α is in effect. Let τ_n denote the (random) time at which $(n(t))$ jumps for the nth time. Then the cost for a control policy $\boldsymbol{u}$ over the random time interval $[0,\tau_n)$ is

$$E_x^u\int_0^{\tau_n}e^{-\alpha t}c'x_t\,dt.$$

We easily compute [19] that this cost is equal to the constant factor $(\alpha+\gamma)^{-1}$ times

$$E_x^u\sum_{k=0}^{n-1}\beta^kc'x_k \tag{2.2}$$

where $\beta=\gamma/(\alpha+\gamma)$. We will ignore the constant factor and take the cost to be (2.2). In view of (2.1) this cost can also be viewed as the cost over n time steps for a discrete time decision process with discount factor β.

For a given initial state x define

$$V_n^\beta(x)=\min_u E_x^u\sum_{k=0}^{n-1}\beta^kc'x_k \tag{2.3}$$

for n less than or equal to $+\infty$. (Except for certain special cases, $V_{+\infty}^\beta(x)=+\infty$ unless $\beta<1$.) By convention, set $V_0^\beta(x)=0$. Then V_n^β is characterized by the dynamic programming optimality equation $V_{n+1}^\beta=TV_n^\beta$ for $0\leq n<+\infty$ where T is the operator defined on real-valued functions on $\mathscr{S}$ by

$$\begin{aligned}Tf(x)=c'x+\beta\gamma^{-1}\{&\lambda_1f(A_1x)\\&+\lambda_2f(A_2x)+\mu_1f(D_1x)+\mu_2f(D_2x)\\&+\lambda\min(f(A_1x),f(A_2x))\\&+\mu\min(f(D_1x),f(D_2x))\\&+\nu_1\min(f(x),f(R_{12}x))\\&+\nu_2\min(f(x),f(R_{21}x))\}.\end{aligned} \tag{2.4}$$

Since at most one arrival occurs per step, we have for $\beta<1$ that

$$\lim_{n\to\infty}\sup_u E_x^u\sum_{k=n}^{\infty}\beta^kc'x_k=0.$$

This readily implies (see Schäl [20] for generalizations) that for each x,

$$V_{+\infty}^\beta(x)=\lim_{n\to\infty}V_n^\beta(x). \tag{2.5}$$

Furthermore, when there are k steps to go (if $\beta<1$ we can allow $k=+\infty$) and the current state is x, then an optimal action z, $z=(a,d,r_{12},r_{21})$, is determined by the condition $z\in\{0,1\}^4$ and

$$\begin{aligned}a=0&\Leftrightarrow V_k^\beta(A_1x)>V_k^\beta(A_2x)\\d=0&\Leftrightarrow V_k^\beta(D_1x)>V_k^\beta(D_2x)\\r_{12}=0&\Leftrightarrow V_k^\beta(R_{12}x)>V_k^\beta(x)\\r_{21}=0&\Leftrightarrow V_k^\beta(R_{21}x)>V_k^\beta(x).\end{aligned} \tag{2.6}$$

The first condition in (2.6), for example, means that a controlled arrival will be sent to station 2 (i.e., $a=0$) if and only if the resulting cost-to-go $V_k(A_2x)$ is smaller than what the cost-to-go would be [namely, $V_k(A_1x)$] if instead the arrival were sent to station 1. In fact, each of the four types of decisions in our model boil down to whether a certain customer should be placed into station 1 or into station 2, and the conditions in (2.6) mean that optimal decisions minimize the resulting cost-to-go.

III. An Inductive Proof of Switch Structure

In this section we deduce from the optimality equations that the optimal policy is switch type. The finite horizon problem and the infinite horizon problem with $\beta<1$ are considered here. The key is to show that the optimal value functions satisfy conditions c1) and c2) below. (Conditions a) and b) are also considered to aid in the induction proof. Conditions equivalent to a), b), c1), and c2) were used by Davis [4] and Ghoneim [7] for their related problems.) Conditions c1) and c2) reflect the fact that the difference in cost between having one more customer at station 1 and having one more customer at station 2 is a decreasing (increasing) function of the number already at station 1 (station 2). Hence, we switch from preferring to have an additional customer at station 1 to having him/her at station 2 as either the number at station 1 increases or as the number at station 2 decreases. Therefore, a switching curve exists.

In order to deal with boundaries, it is convenient to work with functions on $\mathbb{Z}^2$ rather than on $\mathbb{Z}_+^2$. For a function f on $\mathbb{Z}_+^2$, define the function $\hat{f}$ on $\mathbb{Z}^2$ by $\hat{f}(x_1, x_2) = f(x_1^+, x_2^+)$. Define mappings from $\mathbb{Z}^2$ to $\mathbb{Z}^2$ by

$$\hat{A}_1 x = (x_1 + 1, 1)$$
$$\hat{D}_1 x = (x_1 - 1, 1)$$
$$\hat{R}_{12} x = (x_1 - 1, x_2 + 1)$$

and define $\hat{A}_2$, $\hat{D}_2$, and $\hat{R}_{21}$ similarly. Let $\hat{T}$ be the operator acting on functions on $\mathbb{Z}^2$ which is defined by the right side of (2.4) with A_i, D_i, and R_{ij} everywhere replaced by $\hat{A}_i$, $\hat{D}_i$, and $\hat{R}_{ij}$. (Since $A_i x = \hat{A}_i x$ for x in $\mathbb{Z}_+^2$, we will sometimes write $A_i x$ for $\hat{A}_i x$ even if x is not in $\mathbb{Z}_+^2$. We will also adopt the convention that "increasing" means nondecreasing, and we will specifically say "strictly increasing" when we mean it. A similar convention should be understood for "decreasing.") Finally, define $\mathscr{V}$ (respectively, $\hat{\mathscr{V}}$) to be the set of real-valued functions f on $\mathbb{Z}_+^2$ (respectively, on $\mathbb{Z}^2$) such that:

a) $f(x)$ is increasing in x_1 and x_2

b) $f(x+y) - f(x)$ is increasing in x_1 and in x_2 for each fixed y in $\mathbb{Z}_+^2$

c1) $f(A_2 x) - f(A_1 x)$ is decreasing in x_1

c2) $f(A_2 x) - f(A_1 x)$ is increasing in x_2.

We remark that property b) is equivalent to the two properties: $f(x)$ is convex in x_1 and in x_2 [this comes from $f(x+y) - f(x)$ increasing in x_1 for $y = (1,0)$ and in x_2 for $y = (0,1)$, respectively]; $f(x)$ is supermodular in x [this comes from $f(x+y) - f(x)$ increasing in x_1 for $y = (0,1)$, or, equivalently, from $f(x+y) - f(x)$ increasing in x_2 for $y = (1,0)$]. The convexity part of b is a consequence of property c1) [c2)] and supermodularity.

Lemma 3.1: i) The restriction of a function f in $\hat{\mathscr{V}}$ to $\mathbb{Z}_+^2$ is in $\mathscr{V}$. ii) If f is in $\mathscr{V}$, then $\hat{f}$ is in $\hat{\mathscr{V}}$. iii) If f is a function on $\mathbb{Z}_+^2$, which is increasing in both x_1 and x_2, then Tf is the restriction of $\hat{T}\hat{f}$ to $\mathbb{Z}_+^2$.

Proof: The proofs of assertions i) and iii) are left to the reader. The proof of assertion ii) is equally easy once one notes that if $f \in \mathscr{V}$, then $\hat{f}$ satisfies property c1) for $x_2 = -1$ since, by property b1), $f(x_1 + 1, 0) - f(x_1, 0)$ is increasing in x_1. □

Lemma 3.2: If $g \in \hat{\mathscr{V}}$, then $\hat{T}g \in \hat{\mathscr{V}}$.

Proof: Let $g \in \hat{\mathscr{V}}$. Clearly $\hat{A}_1 g$, $\hat{A}_2 g$, $\hat{D}_1 g$, and $\hat{D}_2 g$ are also in $\hat{\mathscr{V}}$. Furthermore, since the terms in the definition of $\hat{T}$ which involve a minimization are equivalent up to translations, it suffices to check that f defined by

$$f(x) = \min(g(A_1 x), g(A_2 x))$$

is in $\hat{\mathscr{V}}$. It is easy to check that f has property a) since g does. Fix $x \in \mathbb{Z}^2$. We will show that

$$f(A_1^2 x) - f(A_1 x) \geq f(A_1 x) - f(x) \tag{3.1}$$
$$f(A_1 A_2 x) - f(A_2 x) \geq f(A_1 x) - f(x) \tag{3.2}$$

and

$$f(A_1 A_2 x) - f(A_1^2 x) \leq f(A_2 x) - f(A_1 x). \tag{3.3}$$

Inequalities (3.1) and (3.2) imply that f has property b) for $y = (1,0)$. By symmetry f also has property b) for $y = (0,1)$, and therefore for all y $\mathbb{Z}^2$. Inequality (3.3) implies that f has property c1) and by symmetry, f also has property c2). Therefore, $f \in \hat{\mathscr{V}}$ so the lemma will be proved once inequalities (3.1)–(3.3) are established.

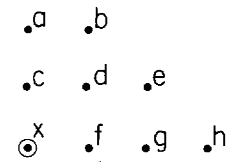


Fig. 2. Points in $\mathbb{Z}^2$ at which the function g is evaluated to define (a, b, c, d, e, f, g, h).

Let the variables a, b, c, d, e, f, g, and h denote g evaluated at $A_2^2 x$, $A_2^2 A_1 x$, $A_2 x$, $A_2 A_1 x$, $A_2 A_1^2 x$, $A_1 x$, $A_1^2 x$, and $A_1^3 x$, respectively (see Fig. 2). Then the inequalities (3.1)–(3.3) to be established become (using "$\wedge$" to denote minimum)

$$e \wedge h - d \wedge g \geq d \wedge g - c \wedge f$$
$$b \wedge e - a \wedge d \geq d \wedge g - c \wedge f$$
$$b \wedge e - e \wedge h \geq a \wedge d - d \wedge g,$$

respectively. Since $g \in \hat{\mathscr{V}}$ we have that

$$c - f \geq - g \geq e - h; \quad e - d \geq d - c$$
$$h - g \geq g - f; \quad e - g \geq d - f \tag{3.4}$$
$$a - d \geq b - e \geq d - g$$
$$a - d \geq e - f \geq d - g$$
$$a - c \leq b - d; \quad d - f \leq e - g$$
$$d - c \leq e - d; \quad d - f \leq b - d \tag{3.5}$$

and

$$a - d \geq b - e \geq e - h$$
$$a - d \geq d - g \geq e - h. \tag{3.6}$$

Inequality (3.1) is then easily established by using (3.4) and by considering separately the four cases: $0 \geq c - f$; $c - f \geq 0 \geq d - g$; $d - g \geq 0 \geq e - h$; and $e - h > 0$. (For example, if $0 \geq c - f$, then also $0 \geq d - g$ and $0 \geq e - h$ by the first part of (3.4). Therefore, inequality (3.1) becomes $h - g \geq g - f$ which is given in (3.4). All the other cases proceed similarly.) Inequality (3.2) is easily established by using (3.5) and by considering the three cases: $0 \geq a - d$; $a - d > 0 \geq d - g$; and $d - g > 0$, and the second of these cases should be divided into four subcases according to whether or not each of $b - e$ and $c - f$ is positive. Finally, inequality (3.3) is easily established by using (3.6) and by considering the three cases: $0 \geq a - d$; $a - d \geq 0 \geq e - h$; and $e - h \geq 0$. □

Lemma 3.3: If $f \in \mathscr{V}$, then $Tf \in \mathscr{V}$.

Proof: If $f \in \mathscr{V}$, then by part ii) of Lemma 3.1, $\hat{f} \in \hat{\mathscr{V}}$. Then, by Lemma 3.2, $\hat{T}\hat{f} \in \hat{\mathscr{V}}$. Now by iii) of Lemma 3.1 Tf is the restriction of $\hat{T}\hat{f}$ to $\mathbb{Z}_+^2$. Finally, apply part i) of Lemma 3.1. □

Lemma 3.4: $V_n^\beta \in \mathscr{V}$ for $0 \leq n < +\infty$ and if $\beta < 1$, then $V_{+\infty}^\beta \in \mathscr{V}$.

Proof: Since $V_0^\beta \equiv 0$, it is in $\mathscr{V}$. Thus, V_n^β is in $\mathscr{V}$ for all finite n by Lemma 3.3 and induction. If $\beta < 1$, then also $V_{+\infty}^\beta$ is in $\mathscr{V}$ by (2.5) and the fact that $\mathscr{V}$ is closed under pointwise limits. □

For the rest of this section we allow $n = +\infty$ if $\beta < 1$. Define the switching function

$$s_n(x_1) = \min\{x_2 \colon V_n^\beta(A_2 x) - V_n^\beta(A_1 x) < 0\} \tag{3.7}$$

and the associated region (see Fig. 3)

$$S_n = \{x \in \mathbb{Z}_+^2 \colon x_2 \leq s_n(x_1)\}. \tag{3.8}$$

Theorem 3.1: The function s_n is increasing for each n. When there are n steps to go the optimal action is given by $z = (a, d, r_{12}, r_{21}) \in \{0,1\}^4$ such that

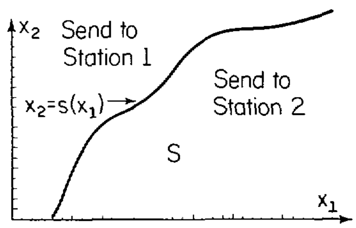


Fig. 3. A typical switch curve.

$$\begin{aligned} a=0 &\Leftrightarrow x\in S_n\\ d=1 &\Leftrightarrow D_1D_2x\in S_n\\ r_{12}=1 &\Leftrightarrow D_1x\in S_n\\ r_{21}=0 &\Leftrightarrow D_2x\in S_n. \end{aligned} \tag{3.9}$$

Proof: As discussed at the beginning of this section, the first assertion is a consequence of the fact, implied by Lemma 3.4, that $V_n^\beta(A_2x)-V_n^\beta(A_1x)$ is decreasing in x_1, and the second from (2.6) and the fact that $V_n^\beta(A_2x)-V_n^\beta(A_1x)$ is increasing in x_2.

Remark: The results and their proofs regarding the finite horizon control problem in this section clearly remain valid if the assumption that the cost c is linear is relaxed to the requirement that c be in $\mathscr{V}$.

IV. Long Run Average Cost

Define the long run average cost for a control policy $\boldsymbol{u}$ and initial state x to be

$$\lim_{n\to\infty}\inf\frac{1}{n}E_x^u\sum_{k=0}^{n-1}c'x_k.$$

In this section we show that there is a stationary policy of switching type which minimizes the average cost. To ensure that there exists a control with a finite long run average cost we make the following assumptions:

$$\begin{aligned} \lambda+\lambda_1+\lambda_2&<\mu+\mu_1+\mu_2\\ \lambda_1&<\mu+\mu_1+\nu_1\\ \lambda_2&<\mu+\mu_2+\nu_2. \end{aligned}\tag{4.1}$$

Since throughout this section $\beta=1$ we will drop the superscript when writing V_n^β.

Lemma 4.1: Let m be a positive integer multiple of the integer n. Then for any policy $\boldsymbol{u}$ and initial state x

$$\frac{1}{m}E_x^u\sum_{k=0}^{m-1}c'x_k\geqslant\frac{1}{n}V_n(0)$$

where $0=(0,0)$.

Proof: The left side of the inequality is the cost over an interval of m consecutive decision points. Suppose that the interval is divided into subintervals of length n and that at the beginning of each such subinterval, all customers in the system are removed. Given this modification, an optimal policy would be to use the optimal policy for an interval of length n during each of the subintervals, and the resulting average cost would be $(1/n)V_n(0)$. Since, by a stochastic domination argument, the minimum average cost cannot be strictly increased by having customers artificially removed for free, the lemma is established. □

Corollary 4.1: $(1/k)V_k(0)$ converges to v^* as k tends to infinity, where

$$v^*=\sup_{k\geqslant 0}\frac{1}{k}V_k(0).\tag{4.2}$$

v^* lower bounds the long run average cost for any control policy $\boldsymbol{u}$ and any initial state x (or, more generally, any initial distribution). v^* is finite if condition (4.1) is satisfied.

Proof: Only the last assertion does not follow directly from Lemma 1. To prove the last assertion it suffices to find a policy which has a finite long run average cost for some initial distribution. Now condition (4.1) ensures that there exists a control value z such that, under the control policy $\boldsymbol{u}$ with $u_k=z$ for each k, the process $(x(t))$ is a Markov process corresponding to a stable Jackson queueing network [11]. The chain is ergodic and under the invariant distribution the number of customers in the two queues are independent geometrically distributed random variables. Thus, the long run average cost is finite when the initial distribution is chosen as the invariant distribution. □

For x and y in $\mathscr{S}$ we write $x\prec y$ if $x_1\leqslant y_1$ and $x_2\leqslant y_2$.

Lemma 4.2: For $n\geqslant 0$,

$$V_{n+1}(y)-V_{n+1}(x)\geqslant V_n(y)-V_n(x)\qquad \text{for } x\prec y\quad x,y \text{ in } \mathbb{Z}_+^2.\tag{4.3}$$

Proof: The proof is by induction on n. Equation (4.3) is true for $n=0$. Now suppose for some $n\geqslant 1$ that

$$V_n(y)-V_n(x)\geqslant V_{n-1}(y)-V_{n-1}(x)\qquad\text{for } x\prec y.\tag{4.4}$$

To complete the induction proof condition (4.3) must then be established. It suffices to establish condition (4.3) for $y=A_1x$ or for $y=A_2x$, and by symmetry only the case $y=A_1x$ needs to be considered.

That is, it suffices to prove that $\Delta(x)\geqslant 0$ for $x\in\mathbb{Z}_+^2$ where

$$\Delta(x)=V_{n+1}(A_1x)-V_{n+1}(x)-[V_n(A_1x)-V_n(x)].$$

By Lemma 3.1,

$$\Delta(x)=\hat{T}\tilde{V}_n(A_1x)-\hat{T}\tilde{V}_n(x)-[\hat{T}\tilde{V}_{n-1}(A_1x)-\hat{T}\tilde{V}_{n-1}(x)]$$

for $x\in\mathbb{Z}_+^2$, and for later use we conclude from condition (4.4) that

$$\tilde{V}_n(y)-\tilde{V}_n(x)\geqslant\tilde{V}_{n-1}(y)-\tilde{V}_{n-1}(x)\qquad\text{for } x\prec y,\quad x,y\in\mathbb{Z}^2.\tag{4.5}$$

Now the operator $\hat{T}$ is the sum of operators of three types. One operator simply adds the function $c'x$, which gives no contribution to $\Delta(x)$. Operators of the second type in $\hat{T}$ are translations in $\mathbb{Z}^2$. By condition (4.5) such operators give a nonnegative contribution to $\Delta(x)$. The remaining four operators in the sum defining T involve minimization, such as the operator U defined by

$$Uf(x)=\min(f(A_1x),f(A_2x))$$

and these four operators are equivalent modulo translations. It thus suffices to show that the contribution to $\Delta(x)$ of one of these four operators, say U, is nonnegative. Equivalently, we need to show that

$$a\wedge c-b\wedge d-(e\wedge g-f\wedge h)\geqslant 0\tag{4.6}$$

where (a,b,c,d) [respectively, (e,f,g,h)] is given by V_n (respectively, V_{n-1}) evaluated at the four points A_1^2x, A_1x, A_1A_2x, and A_2x.

By (4.4), we have $a-b\geqslant e-f$; $c-d\geqslant g-h$; and $c-b\geqslant g-f$. Furthermore, since V_{n-1} and V_n are in $\mathscr{V}$, we have that

$d-b \geqslant c-a$ and $h-f \geqslant g-e$. Now, these last two inequalities imply that one of the three conditions $(a \leqslant c,\ b \leqslant d)$, $(a > c,\ b \leqslant d)$, or $(a > c,\ b > d)$ is true and that one of the three conditions $(e \leqslant g,\ f \leqslant h)$, $(e > g,\ f \leqslant h)$, or $(e > g,\ f > h)$ is true. Choosing one condition from each set yields nine possible cases, and in each case (4.6) can be readily verified. □

Define the relative cost functions $h_n(x)$ for $n \geqslant 0$ by

$$h_n(x) = V_n(x) - V_n(0) \qquad \text{for } x \in \mathscr{S}.$$

Corollary 4.2: For each x in $\mathscr{S}$, $h_n(x)$ is nonnegative and is increasing in n.

For a control policy $\boldsymbol{u}$ define

$$M(x) = E_x^{\boldsymbol{u}} \sum_{k=0}^{\tau} c'x_k \tag{4.7}$$

where

$$\tau = \min\{k: x_k = 0\} \leqslant +\infty. \tag{4.8}$$

Lemma 4.3: Under condition (4.1) there exists a control $\boldsymbol{u}$ so that $M(x)$ is finite for all x.

Proof: Using [9, Theorem 2.3 and Proposition 2.5] it is easy to show that the controls which give rise to a stable Jackson network give finite values for $M(x)$. □

Lemma 4.4: Let $M(x)$ be defined as in (4.7) and (4.8) for some arbitrary policy $\boldsymbol{u}$. Then

$$h_n(x) \leqslant M(x).$$

Proof: In terms of V_n this inequality reads

$$V_n(x) \leqslant V_n(0) + M(x).$$

A possible policy for initial state zero is to use the policy $\boldsymbol{u}$ up until time τ and then (if it should happen that $\tau < n$) switch to the optimal policy for horizon n and initial state 0, shifted over to start at time τ. The cost over the horizon $[0, n]$ for this combination policy is at most $V_n(0) + M(x)$, so the lemma is established. □

Theorem 4.1: The limit

$$h(x) = \lim_{n \to \infty} h_n(x) \tag{4.9}$$

exists and is finite for each x, and is the minimal nonnegative solution to

$$h = Th - v^*. \tag{4.10}$$

If $\boldsymbol{u}^*$ is the control policy defined by (2.6) with V_n^β replaced by h, or equivalently if $\boldsymbol{u}^*$ is given by (3.9) for the switching region defined by (3.7) and (3.8) with V_n replaced by h, then

$$v^* = \lim_{n \to \infty} \frac{1}{n} E_x^{\boldsymbol{u}^*} \sum_{k=0}^{n-1} c'x_k \tag{4.11}$$

so that $\boldsymbol{u}^*$ minimizes the long run average cost.

Proof: The limit h exists and is finite by Corollary 4.2 and Lemmas 4.3 and 4.4. The definitions of h_n and h_{n+1} and the dynamic programming equation $V_{n+1} = TV_n$ yield that

$$h_{n+1} = Th_n - (V_{n-1}(0) - V_n(0)).$$

As n tends to infinity h_{n+1} tends to h and Th_n tends to Th. Therefore, $V_{n+1}(0) - V_n(0)$ also converges as n tends to infinity and the limit must be v^* in view of Corollary 4.1. This establishes (4.10).

Let g be any other nonnegative solution to (4.10). The monotonicity of the operator T implies [using induction and then (4.2)] that (pointwise)

$$h_n \leqslant g + (nv^* - V_n(0)) \leqslant g.$$

The limit h is therefore also bounded by g so h is the minimal solution to (4.10).

By induction and Lemma 3.3, h_n, and therefore the limit h, is contained in $\mathscr{V}$. The two descriptions of $\boldsymbol{u}^*$ in the theorem are thus indeed equivalent. Finally, repeated use of (4.10) implies that

$$h(x) = E_x^{\boldsymbol{u}^*} \sum_{k=0}^{n-1} c'x_k - nv^* + E_x^{\boldsymbol{u}^*} h(x_n)$$

so that

$$\frac{1}{n} E_x^{\boldsymbol{u}^*} \sum_{k=0}^{n-1} c'x_k = h(x)/n + v^* - E_x^{\boldsymbol{u}^*} h(x_n)/n.$$

Since $h \geqslant 0$ this implies that

$$\limsup_{n \to \infty} \frac{1}{n} E_x^{\boldsymbol{u}^*} \sum_{k=0}^{n-1} c'x_k \leqslant v^*.$$

Equation (4.11) is a consequence of this equation and Corollary 4.1. □

Remarks: 1) By treating separately cases when $c_1 = 0$ or $c_2 = 0$ we can easily establish that $\boldsymbol{u}^*$ minimizes the long run cost with probability one. (See [3].) 2) In a much more general setting Borkar [3] established a useful set of conditions which directly apply to our problem (assuming that $c_1 > 0$ and $c_2 > 0$). He proved that there exists an optimal policy which is stationary and is given by a relative value function h which satisfies (4.10). To apply the inductive approach we deduced (4.9). It would be useful to establish conditions in Borkar's setting which imply (4.9) or which imply the closely related property that h is a unique (up to an additive constant), or a minimal positive, solution to (4.10).

V. Some Nonconvex Value Functions

The system considered in the previous sections should be slightly generalized in order to be considered "the" general controlled version of a two-station Jackson network. The natural generalization would be to have "forced transfers" in the following sense. There would be an additional uncontrolled server at each queue such that after being served by such a server a customer must join the other queue. Equivalently, we could simply require the control values r_{ij} in our original model to take values in fixed intervals which are bounded away from zero. Unfortunately, we have not been able to establish the existence of a switching curve for this model. In fact, the example below shows that the convexity properties we relied on in earlier sections can be violated if forced transfers are included in the model. We conjecture that the optimal controls are still switch type, but that a more delicate inductive argument (or other method) is needed to prove it. In this direction we show in the next section how the optimality of threshold policies can be established for a related single station system which also can lead to nonconvex value functions.

Example 1: Consider the system pictured in Fig. 4 for which there are forced transfers from station 1 to station 2 at rate δ and there are departures from station 1 which leave the system at rate μ_1. Suppose that $\mu_1 + \delta = 1$ and that $\delta > 0$. Let $(c_1, c_2) = (0, 1)$. Then the dynamic programming operator T is given by

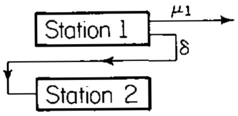


Fig. 4. A two-station Markov network with forced transfers which can lead to a nonconvex value function.

Fig. 5. A single station with parameters λ, λ_1, μ, and μ_1 and control $a = (a, d)$ in $[0,1]^2$.

$$Tf(x) = x_2 + \beta\{\delta f(R_{12}x) + \mu_1 f(D_1 x)\}.$$

Since there are no controls for this example it is not difficult to compute that

$$V_n^1(j,k) = kn + \frac{\delta}{2}\{n(n-1) - (n-j)^+(n-j-1)^-\}$$

and

$$V_{+\infty}^{\beta}(j,k) = \frac{k}{1-\beta} + \frac{\delta\beta(1-\beta^j)}{(1-\beta)^2}.$$

These functions do not satisfy condition c1) (which was crucial for deducing switch structure) nor are they convex in j for k fixed.

VI. Optimality of Threshold Control for a General Single Station Markov Network

Consider the single station system pictured in Fig. 5. In analogy with the two-station system described in Section II the system is influenced by a control with values $z = (a, d)$ in $[0,1]^2$. The arrival rate is $\lambda + a\lambda_1$ and the departure rate is $\mu + d\mu_1$. We will consider a cost function c which is convex on the state space $\mathbb{Z}_+$. The function c is *not* assumed to be monotone, primarily for two reasons. First, nonmonotone cost functions arise in flow control problems for which there is a penalty for having either too few (since throughput suffers) or too many (since delay becomes large) customers in the system. Second, for a closed network of two stations a convex increasing (in the $\mathbb{Z}_+^2$ ordering) cost function need not be monotone as a function of the number of customers in one of the stations (for a fixed total number in the network). Thus, the analysis in this section provides insight into the behavior of the two-station systems considered in the preceding sections. A system with more general cost structures and decision spaces was considered by Serfozo [22]. He assumed, however, that the instantaneous cost is a monotone function of the state.

Assume that $\lambda + \lambda_1 + \mu + \mu_1 = 1$. Then the dynamic programing operator T is defined by

$$\begin{aligned} Tf(x) = c(x) + \beta\{&\lambda_1 \min(f(x), f(x+1)) \\ &+ \mu_1 \min(f(x), f((x-1)^+)) \\ &+ \lambda f(x+1) + \mu f((x-1)^+)\} \end{aligned}$$

and value functions V_n^β are defined by $V_0^\beta = 0$ and $V_{n-1}^\beta = TV_n^\beta$ for $n \geqslant 0$.

Example 2: Suppose $\mu = 1$ and $c(x) = \max(0, k_0 - x)$ for some constant k_0. The value function V_{10}^1 is shown in Fig. 6 for $k_0 = 5$ and the relative value function h for the long run average cost can be taken to be

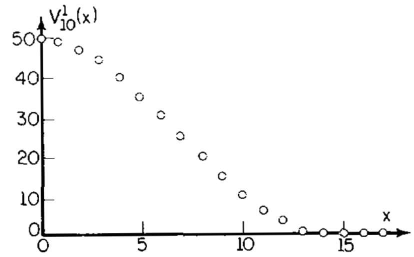


Fig. 6. Value function for $n = 10$ steps when $k_0 = 5$ and $\mu = 1$ in Example 2.

$$h(x) \triangleq \lim_{n\to\infty} V_n^1(x) - nk_0 = \begin{cases} \dfrac{-x(x+1)}{2} & \text{if } 0 \leqslant x \leqslant k_0 \\ -k_0\left(x - \dfrac{1+k_0}{2}\right) & \text{if } x \geqslant k_0. \end{cases}$$

Neither of these functions is convex.

When the minimum cost to go at the next decision time is given by V, then action $w(x) = (a(x), d(x))$ is optimal if $w = (a, d)$ is any $[0,1]^2$-valued function on $\mathbb{Z}_+$ (such functions are called state feedback control functions, or simply control functions) satisfying

$$\begin{aligned} a(x) &= 1 && \text{if } V(x) > V(x+1) \\ a(x) &= 0 && \text{if } V(x) < V(x+1) \\ d(x) &= 1 && \text{if } V(x) > V((x-1)^+) \\ d(x) &= 0 && \text{if } V(x) < V((x-1)^+). \end{aligned} \tag{6.1}$$

We call a function f on $\mathbb{Z}_+$ "unimin" (short for "unique minimum") if $f(x+1) - f(x) \geqslant 0$ for all $x > y$ whenever y satisfies $f(y+1) - f(y) > 0$. We call a control function w threshold type if $a(x) = I\{x < l\}$ and $d(x) = I\{x > l\}$ for some l in $\mathbb{Z}_+ \cup \{+\infty\}$ and l is called the threshold. Clearly there exists a threshold type control function w satisfying conditions (6.1) if and only if V is unimin.

A convenient way to prove that a value function is unimin is to prove that it is convex. This approach was used by Serfozo [22]. This approach does not work for the system considered in this section, in view of Example 2. The next proposition shows that, at least in a special case, the inductive method based on value iteration can nevertheless be used to prove that the value functions are unimin.

Proposition 6.1: Suppose that $\lambda = 0$ (i.e., no forced arrivals). Then for each $n \geqslant 1$:

a) if $V_n^\beta(x_0) < V_n^\beta(x_0+1)$, then V_n restricted to $(x_0, x_0 + 1, \cdots)$ is convex.

b) $(V_n^\beta(x+1) - V_n^\beta(x))^- \geqslant (V_{n-1}^\beta(x+1) - V_{n-1}^\beta(x))^-$ for x in $\mathbb{Z}_+$.

Remark: Property a) ensures that V_n^β is unimin. The proposition can be readily proved by induction by the method used to verify Lemmas 3.2 and 4.2. We conjecture that even if $\lambda > 0$, then V_n^β is unimin and that more sophisticated induction statements can be used to prove it. In the remainder of this section it is shown by a different method that $V_{+\infty}^\beta$ is unimin without restrictions on λ, λ_1, μ, and μ_1.

Induction Based on Policy Iteration and Stochastic Coupling

Policy iteration and stochastic coupling are applied in this subsection to prove that optimal threshold policies exist for the general single station model for the infinite horizon discounted cost criteria. Since we will consider infinite horizon problems and nonlinear c, a growth assumption will be placed on c to ensure that policy iteration works. Fix the discount factor β and a number η with $0<\beta<\eta<1$, and define the norm of a function f on $\mathbb{Z}_+$ by

$$\|f\| = \sup_{x\in\mathbb{Z}^+} |f(x)|\eta^x.$$

Henceforth, we assume that c is convex and that $\|c\|$ is finite. A control $\boldsymbol{u}$ is called a stationary state feedback control if $u_k = w(x_k)$ for a control function w. For such $w=(a,d)$ we define the operator T^w by

$$T^w f(x) = c(x) + \beta\{(\lambda_1 a(x)+\lambda)f(x+1) + (\mu_1 d(x)+\mu)f((x-1)^-)\}.$$

It is readily checked that the operators T and T^w are contractions relative to the norm $\|\ \ \|$. By the elementary fixed point theorem for contractions in complete metric spaces (see [1, pp. 249–251]) there exist unique functions V and U^w with finite norms such that

$$V = c + TV$$

and

$$U^w = c + T^w U^w.$$

Furthermore,

$$V(x) = \min_{\boldsymbol{u}} E_x^{\boldsymbol{u}} \sum_{k=0}^{\infty} c(x_k)\beta^k$$

and, if $\boldsymbol{u}$ denotes the stationary control associated with control function w, then

$$U^w(x) = E_x^{\boldsymbol{u}} \sum_{k=0}^{\infty} c(x_k)\beta^k.$$

U^w is called the value function for control function w.

Theorem 6.1: V is unimin (and so there exists an optimal control function of threshold type).

The proof will be given after some lemmas are presented.

In the following lemma, P_x^l denotes the probability distribution of the Markov random process resulting from using initial state x and the stationary feedback control with threshold l.

Lemma 6.1—Coupled Processes for Threshold Control: Given an integer l, let $\tilde{P}_{xy}$ denote the probability distribution of the time homogeneous Markov chain (x_k, y_k) with state space

$$\tilde{\mathcal{S}} = \{(x,y)\colon x\in\mathbb{Z}_-,\ y=x \text{ or } y=x+1\},$$

one-step transition probabilities

$$\begin{aligned} &p(x',y'|x,y)\\ &= \lambda I\{x'=x+1, y'=y+1\} + \mu I(x'=(x-1)^-, y'=(y-1)^+\}\\ &+ \lambda_1 I\{x'=x+I\{x<l\}, y'=y+I\{y<l\}\}\\ &+ \mu_1\{x'=(x-I\{x>l\})^+, y'=(y-I\{y>l\})^+\},\end{aligned}$$

and initial state (x,y) in $\tilde{\mathcal{S}}$. Then for (x,y) in $\tilde{\mathcal{S}}$,

1) (x_k) under $\tilde{P}_{xy}$ has distribution P_x^l

2) (y_k) under $\tilde{P}_{xy}$ has distribution P_y^l

3) $(y_k - x_k)$ under $\tilde{P}_{xy}$ takes values 0 and 1 and is nonincreasing in k.

Proof: The proof is elementary and is left to the reader. See [12], [13], or [18, pp. 19–24] for other examples of coupled Markov chains. □

Henceforth, we will write $\tilde{E}_x$ to denote $\tilde{E}_{x,x+1}$.

Lemma 6.2: Let U be the value function for a stationary control of threshold type. Then U is unimin.

Proof: Let the threshold be l and consider the process defined in Lemma 6.1. Define the random time τ by

$$\tau = \min\{k\geq 0\colon x_k = y_k\}$$

(set $\tau=+\infty$ if the set is empty) and define

$$r(x) = \frac{U(x+1)-U(x)}{\tilde{E}_x \sum_{k=0}^{\tau-1}\beta^k} \qquad \text{for } x\in\mathbb{Z}_+.$$

We will prove the fact that r is increasing on $\mathbb{Z}_+$, for this fact immediately implies the lemma.

Set $\delta(x) = c(x+1)-c(x)$ and then

$$\begin{aligned} U(x+1)-U(x) &= \tilde{E}_x \sum_{k=0}^{\infty} \beta^k(c(y_k)-c(x_k))\\ &= \tilde{E}_x \sum_{k=0}^{\tau-1}\beta^k\delta(x_k) \end{aligned}\tag{6.2}$$

so that

$$r(x) = \frac{\tilde{E}_x \sum_{k=0}^{\tau-1}\delta(x_k)\beta^k}{\tilde{E}_x\sum_{k=0}^{\tau-1}\beta^k}.$$

Since the function r depends on the function δ in a linear way and since δ (which is increasing since c is convex) is the sum of a constant function and functions of x of the form $I\{x\geq b\}$, it suffices to verify that r is increasing for either of these two choices for δ. Now r is increasing (in fact, r is constant) if δ is constant. It remains to consider the case when $\delta(x)=I\{x\geq b\}$ for some constant b. That is, we must prove for x and b in $\mathbb{Z}_-$ that

$$\frac{\tilde{E}_x \sum_{k=0}^{\tau-1} I\{x_k\geq b\}\beta^k}{\tilde{E}_x\sum_{k=0}^{\tau-1}\beta^k} \leq \frac{\tilde{E}_{x+1} \sum_{k=0}^{\tau-1} I\{x_k\geq b\}\beta^k}{\tilde{E}_{x+1}\sum_{k=0}^{\tau-1}\beta^k}. \tag{6.3}$$

Suppose that $x<b$. Define σ by

$$\sigma = \min\{\min\{k\colon x_k = x+1\}, \tau\}.$$

Using the fact that $x_k<b$ for $k<\sigma$, the fact that $x_\sigma = x+1$ if $\sigma<\tau$, and the strong Markov property of (x_k) under $\tilde{P}_x$, we deduce that

$$\begin{aligned} &\tilde{E}_x \sum_{k=0}^{\tau-1} I\{x_k\geq b\}\beta^k\\ &= \tilde{E}_x \sum_{k=0}^{\sigma-1} I\{x_k\geq b\}\beta^k + \tilde{E}_x I\{\sigma<\tau\}\sum_{k=\sigma}^{\tau-1} I\{x_k\geq b\}\beta^k\\ &= 0 + \tilde{E}_x[\beta^\sigma, \sigma<\tau]\tilde{E}_{x+1}\sum_{k=0}^{\tau-1} I\{x_k\geq b\}\beta^k \end{aligned}$$

and that

$$\tilde{E}_x \sum_{k=0}^{\tau-1} \beta^k = \tilde{E}_x \sum_{k=0}^{\sigma-1} \beta^k + \tilde{E}_x[\beta^\sigma, \sigma > \tau]\tilde{E}_{x+1} \sum_{k=0}^{\tau-1} \beta^k.$$

These two equations readily imply inequality (6.3).

On the other hand, if $x \geqslant b$, then define σ by

$$\sigma = \min\{\min\{k: x_k = x\}, \tau\}.$$

Then deduce (from the facts that $x_k \geqslant b$ for $k < \sigma$ and that $x_\sigma = x$ if $\sigma < \tau$ and from the strong Markov property) that

$$\tilde{E}_{x+1} \sum_{k=0}^{\tau-1} I\{x_k \geqslant b\}\beta^k$$
$$= \tilde{E}_{x+1} \sum_{k=0}^{\sigma-1} \beta^k + \tilde{E}_{x-1}[\beta^\sigma, \sigma < \tau]\tilde{E}_x \sum_{k=0}^{\tau-1} I\{x_k \geqslant \beta\}\beta^k$$

and that

$$\tilde{E}_{x+1} \sum_{k=0}^{\tau-1} \beta^k = \tilde{E}_{x+1} \sum_{k=0}^{\tau-1} \beta^k$$
$$+ \tilde{E}_{x+1}[\beta^\sigma, \sigma < \tau]\tilde{E}_k \sum_{k=0}^{\tau-1} I\{x_k \geqslant b\}\beta^k.$$

These two equations readily imply inequality (6.3).

Inequality (6.3) is thus true for all x, b in $\mathbb{Z}_{-}$, so r is indeed increasing, and the theorem is proved. □

Remark: The fact that $r(x)$ defined in the proof of Lemma 6.2 is increasing means that U does yield a convex function when the x axis is "rescaled" so that the denominator in the definition of r is the distance between points x and $x+1$.

Lemma 6.3: [Recall that $\delta(x) = c(x+1) - c(x)$.] Suppose that for some $\epsilon > 0$ and $K \geqslant 0$ that $\delta(x) \geqslant \epsilon$ for all $x \geqslant K$. Then there is a constant L depending only on β, ϵ, K, and $\delta(0)$ such that the value function U for any stationary threshold policy satisfies $U(x+1) - U(x) > 0$ for $x \geqslant L$.

Proof: Since $x_{k+1} - x_k \geqslant -1$ for each k, for any initial state x with $x \geqslant K$ we have that $x_k \geqslant K$ (and hence, $\delta(x_k) \geqslant 0$) for $1 \leqslant k \leqslant x - K$. In addition, $\delta(x_k) \geqslant \delta(0)$ for $k > x - K$. Using these inequalities in (6.2) yields that

$$U(x+1) - U(x) = \delta(x) + \tilde{E}_x \sum_{k=1}^{\tau-1} \delta(x_k)\beta^k$$
$$\geqslant \delta(x) - \left(\sum_{k=x-K-1}^{\infty} \beta^k\right)\delta(0)^-$$
$$\geqslant \epsilon - (\beta^{x-K+1}\delta(0)^-)/(1-\beta).$$

The lemma is a consequence of this inequality. □

Proof of Theorem 6.1: If c is decreasing, then the inductive approach used in the previous sections can be easily used to show that V is also decreasing and is therefore unimin. Hence, we need only consider the case when $c(x+1) - c(x) \geqslant \epsilon$ for all sufficiently large x for some $\epsilon > 0$.

Let w^0 denote a threshold type control function with threshold l_0 chosen arbitrarily. Then define:

U^n = value function using control function w^n

w^{n-1} = threshold type control function with threshold

$$l_{n+1} = \min\{x: U^n(x+1) > U^n(x)\}.$$

By Lemma 6.2 and induction, U^n is unimin for each n and (w^{n+1}, U^n) satisfies the conditions (6.1).

Therefore, the sequence of functions U^n is decreasing (componentwise) in n, and $U^n \equiv U^{n+1}$ if and only if w^{n+1} is an optimal control policy [1, pp. 245–246]. Now each of the threshold values l_n for $n \geqslant 1$ is contained in a finite set $\{1, 2, \cdots, L\}$ by Lemma 6.3. Therefore, the functions in the decreasing sequence (U^n) are all contained in a finite set of functions. Thus $U^n \equiv U^{n-1}$ for some n, so the threshold type control function w^{n+1} is optimal, $V = U^n$, and V is unimin. □

ACKNOWLEDGMENT

The author wishes to thank G. Sasaki, S. Stidham, and an anonymous reviewer for very helpful suggestions.

REFERENCES

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

[2] R. Boel, P. Varaiya, and E. Wong, "Martingales on jump processes II: Applications," *SIAM J. Contr.*, vol. 13, pp. 1022–1060, 1975.

[3] V. S. Borkar, "Controlled Markov chains and stochastic networks," Dept. Appl. Math., Twente University, Enschede, The Netherlands, Memo. 378, 1982.

[4] E. Davis, "Optimal control of arrivals to a two-server queueing system with separate queues," Ph.D. dissertation, Progr. Oper. Res., North Carolina State Univ., Raleigh, NC, 1977.

[5] A. Ephremides, P. Varaiya, and J. Walrand, "A simple dynamic routing problem," *IEEE Trans. Automat. Contr.*, vol. AC-25, no. 4, pp. 690–693, 1980.

[6] G. J. Foschini and J. Salz, "A basic dynamic routing problem and diffusion," *IEEE Trans. Commun.*, vol. COM-26, pp. 320–347, 1978.

[7] H. Ghoneim, "Optimal control of arrivals to a network of two queues in series," Ph.D. dissertation, Progr. Oper. Res., North Carolina State Univ., Raleigh, NC, 1980.

[8] J. C. Gittins, "Bandit processes and the dynamic allocation indices (with discussion)," *J. R. Statist. Soc. B.*, vol. 41, pp. 148–164, 1979.

[9] B. Hajek, "Hitting time and occupation time bounds implied by drift analysis with applications," *Advances in Appl. Prob.*, vol. 14, pp. 502–525, Sept. 1982.

[10] J. M. Harrison, "Dynamic scheduling of a multiclass queue: Discount optimality," *Oper. Res.*, vol. 23, pp. 270–282, 1975.

[11] J. P. Jackson, "Networks of waiting lines," *Oper. Res.*, vol. 5, pp. 518–521, 1957.

[12] T. Kamae, U. Krengel, and G. L. O'Brien, "Stochastic inequalities on partially ordered spaces," *Ann. Prob.*, vol. 5, pp. 899–912, 1977.

[13] T. M. Liggett, "Coupling the simple exclusion process," *Ann. Prob.*, vol. 4, pp. 339–356, 1976.

[14] W. Lin and P. R. Kumar, *Stochastic Control of a Queue with Two Servers of Different Rates* (Lecture Notes in Control and Inform. Sci., vol. 44), A. Bensoussan and J. L. Lions, Eds. New York: Springer-Verlag, 1982, pp. 719–728.

[15] S. A. Lippman, "Semi-Markov decision processes with unbounded rewards," *Management Sci.*, vol. 19, pp. 717–731, 1973.

[16] S. Lippman, "Applying a new device in the optimization of exponential queueing systems," *Oper. Res.*, vol. 23, pp. 687–710, 1975.

[17] P. Nash, "Optimal allocation of resources between research projects," Ph.D. dissertation, Cambridge University, Cambridge, England, 1973.

[18] C. Preston, *Gibbs States on Countable Sets* (Cambridge Tracts in Mathematics, No. 68). London: Cambridge Univ. Press, 1974.

[19] Z. Rosberg, P. P. Varaiya, and J. C. Walrand, "Optimal control of service in tandem queues," *IEEE Trans. Automat. Contr.*, vol. AC-27, pp. 600–609, 1982.

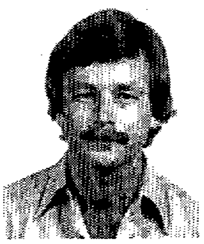
[20] M. Schäl, "Conditions for optimality and for the limit of n-stage optimal policies to be optimal," *Z. Wahrschein. verw. Geb.*, vol. 32, pp. 179–196, 1975.

[21] R. Serfozo, "An equivalence between discrete and continuous time Markov decision processes," *Oper. Res.*, vol. 27, pp. 616–620, 1979.

[22] —, "Optimal control of random walks, birth and death processes, and queues," *Adv. Appl. Prob.*, vol. 13, pp. 61–83, 1981.

[23] M. J. Sobel, "The optimality of full-service policies," *Oper. Res.*, vol. 30, pp. 636–649, 1982.

[24] S. Stidham and N. U. Prabhu, "Optimal control of queueing systems," in *Mathematical Methods in Queueing Theory* (Lecture Notes in Economics and Math. Syst., Vol. 98), A. B. Clarke, Ed. New York: Springer-Verlag, 1974, pp. 263–294.
[25] S. Stidham, "Optimal control of arrivals to queues and networks of queues," Program in Oper. Res., North Carolina State Univ., Raleigh, NC, preprint; also presented at IEEE Conf. on Dec. and Contr., Dec. 1982.
[26] P. Varaiya, J. Walrand, and C. Buyukkoc, "Extensions of the multi-armed bandit problem," Electron. Res. Lab., Berkeley, CA, Memo. UCB/ERL M83/14, 1983.
[27] R. Weber, "On the optimal assignment of customers to parallel servers," *J. Appl. Prob.*, vol. 15, pp. 406–413, 1978.
[28] P. Whittle, "Arm-acquiring bandits," *Ann. Prob.*, vol. 9, pp. 284–292, 1981.
[29] W. Winston, "Optimality of the shortest-processing-time discipline," *J. Appl. Prob.*, vol. 14, pp. 181–189, 1977.



Bruce Hajek (M'79) was born in 1955. He received the B.S. degree in mathematics and the M.S. and Ph.D. degrees in electrical engineering in 1976, 1977, and 1979, respectively.

He is currently an Associate Professor of Electrical Engineering at the University of Illinois at Urbana–Champaign, where he has been since August 1979. His research interests include multiple-user communication theory (e.g., random access and network routing problems), information theory, random fields and other stochastic processes, stochastic control, and combinatorial optimization.

Dr. Hajek was a winner of the 1973 USA Mathematical Olympiad. He was a National Science Foundation Graduate Fellow from 1976 to 1979, and received the Eckman Award of the American Automatic Control Council.

Minimax Linear Observers and Regulators for Stochastic Systems with Uncertain Second-Order Statistics

SERGIO VERDÚ, STUDENT MEMBER, IEEE, AND H. VINCENT POOR, SENIOR MEMBER, IEEE

***Abstract*—The problem of minimax design of linear observers and regulators for linear time-varying multivariable stochastic systems with uncertain models of their second-order statistics is treated in this paper. General classes of allowable covariance matrices and means of the process and observation noises and of the random initial condition are considered. A game formulation of the problem is adopted and it is shown that the optimal filter for the least favorable set of covariances is minimax robust for each of the filtering situations analyzed. Conditions satisfied by the saddle-point solutions are given, and their utility for finding the worst case covariances is illustrated by way of several examples of uncertainty classes of practical interest.**

I. Introduction

THE problem of designing robust observers and regulators for linear stochastic systems with modeling uncertainties in their dynamics, input functions, or statistical characterization has received considerable attention in the technical literature. The majority of the proposed solutions to this problem fall into the categories of adaptive and minimax systems. While adaptive systems utilize the incoming information to learn a more accurate model of the unknown system, the minimax approach provides a fixed system whose worst performance among an assumed possible uncertainty set is the best possible. The mutual advantages and disadvantages of both approaches have been widely discussed (e.g., see [14], [26], [33], and [35]). To name a few, adaptive systems are usually more complex, have worse performance for small-sample-size problems (and thus, inferior dynamical response), and commonly are ad hoc solutions; minimax systems are criticized as being pessimistic (although often this is not the least advisable engineering approach), as having poor performance in the nominal (and maybe most probable) model, and as being dependent on the specification of an often arbitrary uncertainty class. Generally speaking, adaptive systems are preferable when the uncertainty is large and there is sufficient time to make decisions, however, even in this case, the use of a minimax design as a final stage of an indirect adaptive system is reasonable since the uncertainty of the identification vanishes only asymptotically.

Two major types of uncertainties in the statistical characterization of the process and observation noises in linear stochastic systems have been studied, namely deviations from the Gaussian assumption (see [11] and the references therein) and inaccurate knowledge of the first and second moments (cf. [1] for a Bayesian

Manuscript received August 23, 1982; revised March 17, 1983. This paper is based on prior submission of January 4, 1982. Paper recommended by P. R. Kumar, Past Chairman of the Stochastic Control Committee. This work was supported in part by the Office of Naval Research under Contract N00014-81-0014 and by the National Science Foundation under Grant ECS-82-12080.

The authors are with the Department of Electrical Engineering and the Coordinated Science Laboratory, University of Illinois at Urbana–Champaign, Urbana, IL 61801.

0018-9286/84/0600-0499$01.00 ©1984 IEEE