February 1982 LIDS-R-1178
(revised June 1982)

# CONVEXITY AND CHARACTERIZATION OF OPTIMAL POLICIES IN A DYNAMIC ROUTING PROBLEM[1]

by

John N. Tsitsiklis[2]

## ABSTRACT

An infinite horizon, expected average cost, dynamic routing problem is formulated for a simple failure prone queueing system, modelled as a continuous time, continuous state controlled stochastic process. We prove that the optimal average cost is independent of the initial state and that the cost-to-go functions of dynamic programming are convex. These results, together with a set of optimality conditions lead to the conclusion that optimal policies are switching policies, characterized by a set of switching curves (or regions), each curve corresponding to a particular state of the nodes (servers).

Key words: Stochastic control; unreliable queueing systems; average cost; jump disturbances.

---

1) I would like to thank Prof. M. Athans for his help and Dr. S. Gershwin for his suggestions and for reading the manuscript. This research was supported by the National Science Foundation under grant DAR-78-17826.

2)Doctoral Student, Laboratory for Information and Decision Systems, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA.

# 1. INTRODUCTION

## Overview

The main body of queueing theory has been concerned with the properties of queueing systems that are operated in a certain, fixed fashion (Ref. 1). Considerable attention has also been given to optimal static (stationary) routing strategies in queueing networks (Refs. 2–4) which are often found from the solution of a nonlinear programming problem (flow assignment problem).

Concerning dynamic control strategies, most of the literature (Refs. 5–6 are good surveys) deals with the control of the queueing discipline (priority setting) or with the control of the arrival and/or service rate in a M/M/1 (Ref. 7) or M/G/1 (Ref. 8) queue. Ref. 9 considers the problem of controlling the service rate in a two-stage tandem queue.

Results for queueing systems where customers have the choice of selecting a server are fewer. Ref. 10 considers multi-server queueing models with lane selection and derives mean waiting times but does not consider the optimization problem. Some problems with a high degree of symmetry have been solved (Refs. 11–13) leading to intuitively appealing strategies like, for example, "join the shortest queue". Results for systems without any particular symmetry are rare. Ref. 14 contains a qualitative analysis of a dual purpose system. In Ref. 15, a routing problem (very similar to ours) where the servers are allowed to be failure prone is solved numerically. A simpler failure-prone system is studied in Ref. 16 , and some analytical results are derived. Finally, the dynamic control problem for a class of flexible manufacturing systems, as defined in Ref. 17, has significant qualitative similarities with our problem.

In this paper we intend, through the study of a particular queueing system, to display a methodology which may be used to establish certain properties of dynamically controlled queueing systems. We consider an unreliable (failure prone) system (Figure 1) with arrivals modelled as a continuous flow. Consequently, our model concentrates more on the effects of failures rather than the effects of random arrivals and service times, as is the case in mainstream queueing theory. We prove convexity of the cost-to-go functions of dynamic programming (and hence the optimality of switching policies). Our methodology readily extends to more complex configurations.

## Problem Description.

We study a queueing control problem corresponding to the unreliable queueing system depicted in Figure 1. We let $M_o$, $M_1$ and $M_2$ be failure prone nodes (servers, machines, processors) and $B_1$, $B_2$ be buffers (queues) with finite storage capacity. Machine $M_o$ receives external input (assumed to be always available) which it processes and sends to either of the buffers $B_1$ and $B_2$. Machines $M_1$ and $M_2$ then process the material in the buffers that precede them. We assume that each of the machines may fail and get repaired in a random manner. The failure and repair processes are modelled as memoryless stochastic processes (continuous time Markov chains). We also assume that the maximum processing rate of a machine which is in working condition is finite.

With this system, there are two kinds of decisions to be made: a) Decide on the actual processing rate of each machine, at any time when it is in working condition and input to it is available; b) Decide, at any time, on how to route the output of machine $M_o$.

We consider a rather general performance criterion corresponding to the production rate (throughput) of the system, together with a storage cost. Moreover, we formulate the problem as an infinite horizon, expected average cost minimization problem and we are looking for dynamic control policies in which the decision at any time is allowed to depend on all information on the system available at that time. In particular, the values of the control variables are allowed to be quite arbitrary functions of the current states of the machines and the buffer levels.

The above introduced configuration arises in certain manufacturing systems (Refs. 15, 18) (from which our terminology is borrowed) and also in communication networks where the nodes may be thought as being computers and the material being processed as messages (packets) (Ref. 11). Note that the Markovian assumption on the failure and repair process of the nodes implies that a node may fail even at a time when it is not operating. This is a realistic assumption, in unreliable communication networks and in those manufacturing systems where failures may be ascribed to external causes (Refs. 18–19). On the other hand, in some manufacturing systems, failure probabilities increase with the degree of utilization of the machines (Ref. 20). Such systems are not captured by our model and require a substantially different mathematical approach.

We model the queue levels as continuous variables and the flow through the machines as a continuous flow (the fluid approximation of Ref. 2). This is a good model when the workpieces in an actual manufacturing system, or the messages in a communications network, are very small compared with the storage capacity of the queues and when the processing time of any single workpiece (or message) is very small compared with the natural time scales of the system. The latter are determined by the failure and repair rates of the machines and the time needed to fill an empty queue or to empty a full queue.

Let us now comment on the methodology and the mathematical issues involved. Our main objective is to establish that optimal policies are switching policies. Namely, that the state space is divided into two main regions corresponding to the two buffers, separated by a simply connected dividing region. Whenever the state of the system lies in one of the two main regions, it is optimal to send all output of $M_0$ to the buffer corresponding to that region. (Optimality conditions on the dividing region are slightly more complicated and may depend on the actual shape of that region.) To achieve our objective, we need to prove that an (appropriately defined) cost-to-go function is convex.

One method (the most usual) to prove convexity of cost-to-go functions exploits the dynamic programming recursions (value iteration algorithm) and involves an inductive argument. This method has been used in inventory control problems (Ref. 21) and may be easily applied to queueing control problems formulated in discrete time. However, in this paper we are dealing with a continuous state space and a system running in continuous time; the dynamic programming recursions are no more available. Another indirect method (using the duality theory of linear programming) was developed in Ref. 9 but is also applicable only in discrete systems.

In our approach, we define the stochastic process $s^u(t)$ describing the evolution of the state of the system on a single probability space $(\Omega, \mathcal{A}, \mathcal{P})$, independent of the control law being used, or the initial state. Any fixed $\omega \in \Omega$ gives rise to a different sample path $s^u(\omega, t)$ for every different control law and every different initial state. Keeping $\omega$ fixed, we may compare these different sample paths for different control laws and for different initial states. Then, by taking expectations, we can deduce properties of the cost–to–go functions.

We should point out that this approach contrasts with some recent trends in the theory of controlled stochastic processes. In that theory, the mapping $u{:}\omega \mapsto s^u(\omega, t)$ is the same for all $u$ but different control laws lead to different probability measures $\mathcal{P}^u$ (Ref. 22). In our approach, the probability measures are kept fixed, but the mapping $u{:}\omega \mapsto s^u(\omega, t)$ varies.

We study the average cost problem by introducing and solving an auxiliary total cost problem. Therefore, total cost problems may be analyzed with the same tools presented in this paper and, in fact, with less difficulty. In fact, our treatment transcends the scope of the particular routing problem. Most results to be proved are valid for a broad class of dynamically controlled stochastic processes driven by jump Markov disturbances, with linear dynamics, convex costs and convex constraint sets.

## 2. THE DYNAMIC ROUTING PROBLEM.

In this section we formulate mathematically the dynamic routing problem and define the set of admissible control laws and the performance criterion to be minimized.

Consider the queueing system of Figure 1, as described in Section 1. Let $x_i$ be a continuous variable indicating the amount of material in buffer $B_i$ $(i = 1, 2)$ and let $N_i$ be the maximum allowed level in that buffer. We denote $(x_1, x_2)$ by $\underline{x}$. We define an indicator variable $\alpha_i$ for each machine by

$$\alpha_i = \begin{cases} 0 & \text{if machine } M_i \text{ is down (under repair)} \\ 1 & \text{if machine } M_i \text{ is up (operational)} \end{cases} \tag{2.1}$$

We let $\underline{\alpha} \equiv (\alpha_o, \alpha_1, \alpha_2)$ and we refer to it as the "state of the machines". We assume that the time to failure and the time to repair of machine $M_i$ are independent, exponentially distributed random variables with rates $p_i$ and $r_i$, respectively. Then, the process $\underline{\alpha}(t)$ (as well as each of the processes $\alpha_i(t)$) is a continuous time Markov chain. Let $\Omega$ be the set of three-component funtions of one nonnegative real variable $t$, such that each component is right-continuous in $t$ and takes the values 0 or 1. Any sample point $\omega \in \Omega$ may be identified with a unique sample path $\underline{\alpha}(\omega, t)$. Let $\mathcal{A}$ be a $\sigma$-algebra of subsets of $\Omega$ and for any $\underline{\alpha}(0)$ let $\mathcal{P}(\underline{\alpha}(0))$ be a measure defined on $(\Omega, \mathcal{A})$ such that $(\Omega, \mathcal{A}, \mathcal{P}(\underline{\alpha}(0)))$ is a probability space corresponding to the Markov process $\underline{\alpha}(t)$ with initial state $\underline{\alpha}(0)$. Then, $\underline{\alpha}(t)$ is a strong Markov process and any $\omega \in \Omega$ determines uniquely the jump times of each component of $\underline{\alpha}(t)$.

Let $\mathcal{A}_t \subset \mathcal{A}$ be the smallest $\sigma$-algebra such that $\underline{\alpha}(\tau)$ is an $\mathcal{A}_t$-measurable random variable, $\forall \tau \in [0, t]$. A stopping time $T$ is a random variable such that $T \geq 0$ and $\{\omega : T(\omega) \leq t\} \in \mathcal{A}_t$, $\forall t$. For any stopping time $T$, we define a $\sigma$-algebra $\mathcal{A}_T$ as follows:

$$A \in \mathcal{A}_T \quad \text{if and only if} \quad A \cap \{\omega : T(\omega) \leq t\} \in \mathcal{A}_t, \qquad \forall t \geq 0. \tag{2.2}$$

Then, $\mathcal{A}_T$ contains all events that depend only on the history of the process up to the stopping time $T$.

We define the state space $S$ of the system by

$$S \equiv [0, N_1] \times [0, N_2] \times \{0, 1\}^3 \tag{2.3}$$

The state $s(t) \in S$ of the system at time $t$ is defined as

$$s(t) = ((x_1, x_2), (\alpha_o, \alpha_1, \alpha_2))(t) = (\underline{x}, \underline{\alpha})(t) \tag{2.4}$$

Let $\lambda^*$, $\mu_1{}^*$, $\mu_2{}^*$ be the maximum allowed flow rates through machines $M_o$, $M_1$, $M_2$, respectively; let $\lambda(t)$, $\mu_1(t)$, $\mu_2(t)$ be the actual flow rates, at time $t$, through machines $M_o$, $M_1$, $M_2$, respectively; finally, let $\lambda_1(t)$, $\lambda_2(t)$ be the flow rates, at time $t$, from machine $M_o$ to the buffers $B_1$ and $B_2$, respectively. No flow may go through a machine which is down:

$$\alpha_i(t) = 0 \qquad \Rightarrow \begin{cases} \lambda(t) = \lambda_1(t) = \lambda_2(t) = 0 & i = 0 \\ \mu_i(t) = 0 & i = 1, 2 \end{cases} \tag{2.5}$$

Conservation of flow implies

$$\lambda(t) = \lambda_1(t) + \lambda_2(t) \tag{2.6}$$

$$x_i(t) = x_i(0) + \int_0^t (\lambda_i(\tau) - \mu_i(\tau))\, d\tau. \tag{2.7}$$

The integral in (2.7) is well–defined, for each sample path, as long as $\lambda_i(t)$ and $\mu_i(t)$ are right–continuous functions of time, for all $\omega \in \Omega$. This will be guaranteed by our definition of admissible control laws.

We view an admissible control law $u$ as a mapping which to any initial state $s(0) = (\underline{x}(0), \underline{\alpha}(0))$ assigns a stochastic process $u(\omega, t) = (\lambda_1^u(\omega, t),\ \lambda_2^u(\omega, t),\ \mu_1^u(\omega, t),\ \mu_2^u(\omega, t))$ defined on the previously introduced probability space $(\Omega, \mathcal{A}, \mathcal{P}(\underline{\alpha}(0)))$ with the following properties:

(S1) Each of the random variables $\lambda_i^u(t)$, $\mu_i^u(t)$ is $\mathcal{A}_t$-measurable.

(S2) For any $\omega \in \Omega$, $s(0) \in S$, the sample functions $\lambda_i^u(\omega, t)$, $\mu_i^u(\omega, t)$ are right–continuous in time.

(S3)
$$0 \le \lambda_i^u(\omega, t) \qquad i = 1, 2 \tag{2.8a}$$

$$\lambda_1^u(\omega, t) + \lambda_2^u(\omega, t) \le \alpha_0(\omega, t)\lambda^* \tag{2.8b}$$

$$0 \le \mu_i^u(\omega, t) \le \alpha_i(\omega, t)\mu_i{}^* \qquad i = 1, 2 \tag{2.8c}$$

(S4) The solution of the (stochastic) differential equation

$$dx_i^u(\omega, t) = (\lambda_i^u(\omega, t) - \mu_i^u(\omega, t))\, dt \tag{2.9}$$

with initial conditions $\underline{x}(0)$ (which exists and is unique for any $\omega \in \Omega$ and initial state $s(0)$ by assumption (S2)) satisfies

$$0 \le x_i^u(\omega, t) \le N_i \qquad i = 1, 2, \quad \forall \omega \in \Omega, \quad \forall t \ge 0 \tag{2.10}$$

and $x_i^u(\omega, t)$ is a measurable function of the initial state $s(0)$, $\forall \omega \in \Omega$, $\forall t \geq 0$.

It is a consequence of assumption (S1) that between any two jump times of the process $\underline{\alpha}(t)$, equation (2.9) is essentially a deterministic differential equation. More precisely, if $t_n$ is the time of the $n$-th jump of $\underline{\alpha}(t)$, then $\underline{x}^u(t)$, for $t \in [t_n, t_{n+1}]$ is uniquely determined by the history of the system up to time $t_n$.

We let $U$ be the set of all admissible control laws and $U_M \subset U$ the set of those control laws such that $u(t)$ depends only on $s(t)$ and $s(t)$ is a strong Markov process with stationary transition probabilities.

For any initial state $(\underline{x}, \underline{\alpha})$, let $\underline{x}_\alpha^u(t)$ be the value of $\underline{x}$ at time $t$ if control law $u$ is used and if no jump of $\underline{\alpha}$ occurs until time $t$. If $u \in U_M$, the weak infinitesimal generator $\mathcal{L}^u$ of the Markov process $s^u(t) \equiv (\underline{x}^u(t), \underline{\alpha}(t))$ is defined by the pointwise limit

$$(\mathcal{L}^u f)(s) = \lim_{t \downarrow 0} \frac{E[f(s^u(t)) \mid s^u(0) = s] - f(s)}{t} \tag{2.11}$$

The domain of $\mathcal{L}^u$ is the set of the real, bounded and measurable functions on $S$ such that the ratios in the right hand side of (2.11) are uniformly bounded, for all $t$ in some neighborhood of zero, and converge pointwise. For the system being studied, $\mathcal{L}^u$ has a very simple form:

Proposition 2.1: Let $f$ be a real valued, continuous function on the state space such that, for all $\underline{\alpha} \in \{0, 1\}^3$, the ratio $(f(\underline{x}_\alpha^u(t), \underline{\alpha}) - f(\underline{x}_\alpha^u(0), \underline{\alpha}))/t$ converges pointwise and boundedly to a bounded function of $\underline{x}_\alpha^u(0)$. Then, $f$ is in the domain of the weak infinitesimal generator $\mathcal{L}^u$ of $s^u(t)$ and

$$(\mathcal{L}^u f)(\underline{x}, \underline{\alpha}) = \lim_{t \downarrow 0} \frac{f(\underline{x}_\alpha^u(t), \underline{\alpha}) - f(\underline{x}, \underline{\alpha})}{t} + \sum_{\alpha^*} p_{\alpha\alpha^*}[f(\underline{x}, \underline{\alpha}^*) - f(\underline{x}, \underline{\alpha})] \tag{2.12}$$

where $\underline{\alpha}^*$ ranges over those elements of $\{0, 1\}^3$ that differ from $\underline{\alpha}$ in a single component and where $p_{\alpha\alpha^*}$ is the transition rate from $\underline{\alpha}$ to $\underline{\alpha}^*$.

Proof: This result may be derived in straightforward manner from equation (2.11). ∎

It will be notationally convenient to define an operator $\mathcal{L}^u$ even when $u$ is not a Markovian control law, by equation (2.11). In that case, Proposition 2.1 is still valid.

**Performance Criterion.**

We are interested in minimizing the long-run (infinite horizon) average cost resulting from the operation of the system. Let $k(s, \lambda_1, \lambda_2, \mu_1, \mu_2)$ be a function of the state and control variables representing the instantaneous cost. For notational convenience, we define

$$k^u(s^u(\omega, t)) \equiv k(s^u(\omega, t), \lambda_1^u(\omega, t), \lambda_2^u(\omega, t), \mu_1^u(\omega, t), \mu_2^u(\omega, t)) \tag{2.13}$$

We introduce the following assumptions:

$$k^u(s^u(\omega, t)) = f(\underline{x}^u(\omega, t), \underline{\alpha}(\omega, t)) - c_1\mu_1^u(\omega, t) - c_2\mu_2^u(\omega, t) \tag{2.14}$$

where $c_1, c_2 > 0$ and for any $\alpha \in \{0, 1\}^3$, $f(\underline{x}, \underline{\alpha})$ is (i) Nondecreasing in $x_1$ and $x_2$, (ii) Convex and (iii) Lipschitz continuous. Let $f_\alpha(\underline{x})$ be an alternative notation for $f(\underline{x}, \underline{\alpha})$.

The function to be minimized is

$$g^u(s) = \limsup_{T\to\infty} E\left[\frac{1}{T}\int_0^T k^u(s^u(\omega, t))\,dt \,|\, s^u(0) = s\right] \tag{2.15}$$

We define the optimal average cost by

$$g^*(s) = \inf_{u\in U} g^u(s) \tag{2.16}$$

In section 3, we show that $g^*$ is independent of $s$.

## 3. REDUCTION OF THE SET OF ADMISSIBLE CONTROL LAWS.

Suppose that at some time the lead machine is down and both downstream machines are up. If this configuration does not change for a large enough time interval, we expect that any reasonable control law would eventually empty both buffers. Indeed, Theorem 3.1 shows that we may so restrict the set of admissible control laws without worsening the optimal performance of the system (i.e. without increasing the optimal value of the cost functional).

We then show that there exists a particular state which is recurrent under any control law that satisfies the above introduced constraint (Theorem 3.2). The existence of such a recurrent state permits a significant simplification of the mathematical issues typically associated with average cost problems.

We end this section by introducing the class of regenerative control laws. This is the class of control laws for which the stochastic process regenerates (i.e. forgets the past history and starts afresh) each time the particular recurrent state is reached. In that case, $g^u$ admits a simple and useful representation and is independent of $s$ (Theorem 3.3). We show that we may restrict to regenerative control laws without any loss of performance (Theorem 3.4).

Definition 3.1: Let $U_A$ be the set of control laws in $U$ with the following property: If at some time $t_o$, $\underline{\alpha}(t_o) = (0, 1, 1)$ and $\underline{\alpha}(t)$ does not change for a further time interval of $\max\{N_1/\mu_1{*}, N_2/\mu_2{*}\}$ time units then

$$s^u\left(t = t_o + \max\left\{\frac{N_1}{\mu_1{*}}, \frac{N_2}{\mu_2{*}}\right\}\right) = ((0,0),(0,1,1))$$

Remark: A sufficient (but not necessary) condition for a control law $u$ to belong in $U_A$ is that downstream machines operate at full capacity whenever $\underline{\alpha} = (0, 1, 1)$. However, we do not want to impose the latter condition because in the course of the proofs in sections 5 and 6 we will use control laws that violate it.

Theorem 3.1: For any $u \in U$, $s(0) \in S$, there exists some $w \in U_A$ such that

$$\int_0^t k^w(s^w(\omega,\tau))\,d\tau \leq \int_0^t k^u(s^u(\omega,\tau))\,d\tau \qquad \forall t \geq 0, \quad \forall \omega \in \Omega. \tag{3.1}$$

Proof: Fix some initial state $s(0)$ and a control law $u \in U$. Let $w \in U$ be a control law such that, with the same initial state, we have

$$\lambda_i^w(\omega,t) = \lambda_i^u(\omega,t), \qquad i = 1,2, \quad \forall \omega, t. \tag{3.2}$$

$$\mu_i^w(\omega,t) = \begin{cases} \mu_i^* & \text{if} \quad x_i^w(\omega,t) \neq 0, & \underline{\alpha} = (0,1,1) \\ 0 & \text{if} \quad x_i^w(\omega,t) = 0, & \underline{\alpha} = (0,1,1) \\ \mu_i^u(\omega,t) & \text{if} \quad x_i^w(\omega,t) = x_i^u(\omega,t), & \underline{\alpha} \neq (0,1,1) \\ 0 & \text{if} \quad x_i^w(\omega,t) \neq x_i^u(\omega,t), & \underline{\alpha} \neq (0,1,1) \end{cases} \tag{3.3}$$

where $x_i^w(\omega,t)$ is determined by

$$x_i^w(\omega,t) = x_i(0) + \int_0^t (\lambda_i^w(\omega,\tau) - \mu_i^w(\omega,\tau))\, d\tau. \tag{3.4}$$

<u>Lemma 3.1</u>: A control law $w \in U_A$ with the above properties exists and satisfies

$$0 \leq x_i^w(\omega,t) \leq x_i^u(\omega,t), \qquad \forall \omega, t, \quad i = 1,2. \tag{3.5}$$

<u>Proof:</u> (Outline) The trajectories $s^u(\omega,t)$ determine uniquely (and in a nonanticipative way) the trajectories of $s^w(\omega,t)$ by means of (3.2)-(3.4). From (3.3) we can see that whenever $x_i^w(\omega,t) = x_i^u(\omega,t)$ we have $\mu_i^w(\omega,t) \geq \mu_i^u(\omega,t)$ which implies that $x_i^w(\omega,t) \leq x_i^u(\omega,t)$ for all times. From (3.3) again, it is easy to see that $x_i^w(\omega,t)$ never becomes negative. Right continuity of $\lambda_i^w$, $\mu_i^w$ follows from (3.2), (3.3) and right-continuity of $\lambda_i^u$, $\mu_i^u$. ∎

From (3.5) and the monotonicity of $f_\alpha$, we have

$$\int_0^t f_{\alpha(\omega,\tau)}(\underline{x}^w(\omega,\tau))\, d\tau \leq \int_0^t f_{\alpha(\omega,\tau)}(\underline{x}^u(\omega,\tau))\, d\tau, \qquad \forall \omega, t. \tag{3.6}$$

Using (3.2), (3.4) and (3.5), we have

$$\begin{aligned} \int_0^t \mu_i^w(\omega,\tau)\, d\tau &= x_i(0) - x_i^w(\omega,t) + \int_0^t \lambda_i^w(\omega,\tau)\, d\tau \geq \\ & x_i(0) - x_i^u(\omega,\tau) + \int_0^t \lambda_i^u(\omega,\tau)\, d\tau = \int_0^t \mu_i^u(\omega,\tau)\, d\tau, \qquad \forall \omega, t. \end{aligned} \tag{3.7}$$

Adding inequalities (3.6) and (3.7), for $i = 1,2$, we obtain the desired result. ∎

From Theorem 3.1 and the definition of $g^u(s)$, it follows that

<u>Corollary 3.1</u>: $$\inf_{u \in U_A} g^u(s) = \inf_{u \in U} g^u(s) = g^*(s), \qquad \forall s \in S. \tag{3.8}$$

We now proceed with the recurrence properties of control laws in $U_A$. For the rest of this paper we let $s_o$ denote the special state $(\underline{x}, \underline{\alpha}) = ((0,0),(0,1,1))$. Let $u \in U_A$. We define the stopping time

$T^u_n$ as the $n$-th time that the state $s_o$ is reached, given that control law $u$ is used. More precisely, we let $T^u_o = 0$ and

$$T^u_{n+1} \quad = \quad \inf\{t > T^u_n\colon\ s^u(t) = s_o,\ \exists \tau \in (T^u_n, t)\ s.t.\ s^u(\tau) \neq s_o\} \qquad (3.9)$$
$$\infty \qquad \text{if the above set is empty}$$

$T^u_n$ is a stopping time, for all $n$, because $s^u(t)$ is a right-continuous stochatic process and $\{s_o\}$ is a closed subset of $S$.

Let $s^u(0) = (\underline{x}^u(0), \underline{\alpha})$, $s^w(0) = (\underline{x}^w(0), \underline{\alpha})$ be elements of $S$ with the same value of $\underline{\alpha}$; let $u, w \in U_A$ and let $s^u(t)$, $s^w(t)$ be the corresponding stochastic processes, with initial states $s^u(0)$, $s^w(0)$. We define the stopping time $T^{uw}$ by

$$T^{uw} \quad = \quad \inf\{t > 0\colon\ s^u(t) = s^w(t) = s_o\} \qquad (3.10)$$
$$\infty \qquad \text{if the above set is empty}$$

If we are given a third element of $S$, $s^v(0) = (\underline{x}^v(0), \underline{\alpha})$, (with the same value of $\underline{\alpha}$) and a third control law $v \in U$ we may define $T^{uvw}$ in a similar way, as the first time that $s^u(t) = s^w(t) = s^v(t) = s_o$.

Theorem 3.2: Let $u$, $v$, $w \in U_A$ and let $s^u(0)$, $s^v(0)$, $s^w(0)$ be three initial states with the same value of $\underline{\alpha}$. Assume that $p_o \neq 0$, $r_1 \neq 0$, $r_2 \neq 0$. Then,

a) $$E[T^u_{n+1} - T^u_n] \leq B \qquad (3.11)$$

b) $$E[T^{uvw}] \leq B \qquad E[T^{uw}] \leq B \qquad (3.12)$$

where $B$ is a constant independent of $u$, $v$, $w$ and the initial states $s^u(0)$, $s^v(0)$, $s^v(0)$.

Proof: Let $Q_n$ be the $n$-th time that the continuous time Markov chain $\underline{\alpha}(t)$ reaches the state $\underline{\alpha} = (0, 1, 1)$. Since $p_o$, $r_1$, $r_2$ are nonzero, there exists a constant $A$ such that $E[Q_n] \leq nA$, for all initial states $\underline{\alpha}(0)$, and $E[Q_n - Q_m] \leq (n-m)A$. If $\underline{\alpha}(t) = (0, 1, 1)$ and if no jumps of $\underline{\alpha}$ occur for a further time interval of $T \equiv \max\{N_1/\mu_1{}^*, N_2/\mu_2{}^*\}$ time units (which is the case with probability equal or larger to $q \equiv \exp(-(r_0 + p_1 + p_2)T)$ ), the state becomes $s^u(t+T) = s_o$, for any $u \in U_A$ and regardless of the initial state. It follows that $E[T^u_{n+1} - T^u_n] \leq \sum_{k=1}^{\infty}(kA + T)q(1-q)^{k-1} \leq B$, for some finite constant $B$. Similar inequalities hold for $E[T^{uw}]$, $E[T^{uwv}]$. ∎

It will be assumed throughout this paper that $p_o \neq 0$, $r_i \neq 0$, $i = 1, 2$. If we allowed $p_o = 0$, all subsequent results would be still valid, but the recurrent state $s_o$ should be differently chosen.

Theorem 3.2 allows us to break down the infinite horizon into a sequence of almost surely finite and disjoint time intervals $[T^u_n, T^u_{n+1})$. If, in addition, the stochastic process $s^u(t)$ regenerates at the times $T_n^u$, the infinite horizon average cost admits a simple and useful representation in terms of the statistics of $s^u(t),\ t \in [T_1^u, T_2^u)$.

We now define what we mean by a regenerative control law. Intuitively said, regenerative control laws forget the past each time that the state is equal to $s_o$ and start afresh. We first define a regeneration time to mean an almost surely finite stopping time $T$, such that $s^u(T) = s_o$, with probability 1. Our first condition on regenerative control laws is that the past is forgotten at regeneration times. Formally,

(S5) The stochastic process $\{(s^u(T+t), u(T+t)),\ t \geq 0\}$ is independent of $\mathcal{A}_T$, for any regeneration time $T$.

The second requirement is that the stochastic process in (S5) is the same (in a probabilistic sense) for all regeneration times $T$. Namely,

(S6) For any two regeneration times $S$, $T$, the stochastic processes $\{(s^u(T+t), u(T+t)),\ t \geq 0\}$ and $\{(s^u(S+t), u(S+t)),\ t \geq 0\}$ are identically distributed.

Definition 3.2: We let $U_R$ denote the set of regenerative control laws, that is the set of control laws in $U_A$ satisfying (S5) and (S6).

Markovian control laws in $U_A$ certainly belong to $U_R$. However, the proofs of the results of section 5 require us to consider non-Markovian control laws as well. It turns out that $U_R$ is a suitable framework.

Theorem 3.3: Let $u \in U_R$. Then,

$$g^u = \lim_{t\to\infty} E\left[\frac{1}{t}\int_0^t k^u(s^u(\tau))\,d\tau\right] = \frac{E\left[\int_{T_n^u}^{T_{n+1}^u} k^u(s^u(\tau))\,d\tau\right]}{E[T_{n+1}^u - T_n^u]} \qquad n = 1, 2, \ldots \tag{3.18}$$

(Note that the first equality implies that the limit exists and is independent of the initial state.)

Proof: Define

$$W_m \equiv T_{m+1}^u - T_m^u \qquad m = 1, 2, \ldots \tag{3.19}$$

$$U_m \equiv \int_{T_m^u}^{T_{m+1}^u} k^u(s^u(\tau))\,d\tau \qquad m = 1, 2, \ldots \tag{3.20}$$

The random vectors $(W_m, U_m)$, $m = 1, 2, \ldots$ are independent (by S5), identically distributed (by S6). Then, an ergodic theorem (Ref. 23, Vol. 2) implies that

$$\lim_{m\to\infty} \frac{\sum_{k=1}^m U_k}{\sum_{k=1}^m W_k} = \frac{E[U_n]}{E[W_n]}, \qquad \text{almost surely.} \tag{3.21}$$

Now,

$$\lim_{m\to\infty}\frac{\sum_{k=1}^{m}U_k}{\sum_{k=1}^{m}W_k}=\lim_{m\to\infty}\frac{\int_{T_1^u}^{T_m^u}k^u(s^u(\tau))\,d\tau}{T_m^u-T_1^u}=\lim_{m\to\infty}\left[\frac{1}{T_m}\int_0^{T_m^u}k^u(s^u(\tau))\,d\tau\right]+$$
$$\lim_{m\to\infty}\left[\left(\frac{1}{T_m^u-T_1^u}-\frac{1}{T_m^u}\right)\int_0^{T_m^u}k^u(s^u(\tau))\,d\tau\right]-\lim_{m\to\infty}\left[\frac{1}{T_m^u-T_1^u}\int_0^{T_1^u}k^u(s^u(\tau))\,d\tau\right] \tag{3.22}$$

We claim that the second and third summands are almost surely equal to zero. Let $M$ be a bound on $|k^u|$. Then,

$$\left|\left(\frac{1}{T_m^u-T_1^u}-\frac{1}{T_m^u}\right)\int_0^{T_m^u}k^u(s^u(\tau))\,d\tau\right|\leq MT_m^u\frac{T_1^u}{T_m^u(T_m^u-T_1^u)} \tag{3.23}$$

Now, $T_1^u<\infty$ (almost surely) and $\lim_{m\to\infty}T_m^u=\infty$ (almost surely). Therefore, the right hand side of (3.23) converges to zero, almost surely. Also,

$$\left|\frac{1}{T_m^u-T_1^u}\int_0^{T_1^u}k^u(s^u(\tau))\,d\tau\right|\leq\frac{MT_1^u}{T_m^u-T_1^u}\to 0 \qquad \text{a.s.} \tag{3.24}$$

for the same reasons. We now take expectations in (3.22) and invoke (3.21) to obtain

$$E\lim_{m\to\infty}\left[\frac{1}{T_m^u}\int_0^{T_m^u}k^u(s^u(\tau))\,d\tau\right]=\frac{E[U_n]}{E[W_n]} \tag{3.25}$$

Let $T^u(t)=\inf\{\tau\geq t:\exists n \text{ such that } \tau=T_n^u\}$ and observe that the sequence $(1/T_m^u)\int_0^{T_m^u}k^u(s^u(\tau))\,d\tau$ and the function $(1/T^u(t))\int_0^{T^u(t)}k^u(s^u(\tau))\,d\tau$ take the same values in the same order; therefore, they have the same limit and may be interchanged in (3.25). We then use the dominated convergence theorem to interchange the limit and the expectation at the left hand side to obtain

$$\lim_{t\to\infty}E\left[\frac{1}{T^u(t)}\int_0^{T^u(t)}k^u(s^u(\tau))\,d\tau\right]=\frac{E[U_n]}{E[W_n]} \tag{3.26}$$

Finally,

$$\lim_{m\to\infty}\left|E\left[\frac{1}{t}\int_0^t k^u(s^u(\tau))\,d\tau\right]-E\left[\frac{1}{T^u(t)}\int_0^{T^u(t)}k^u(s^u(\tau))\,d\tau\right]\right|\leq$$
$$\lim_{t\to\infty}\left|E\left[\frac{1}{T^u(t)}\int_t^{T^u(t)}k^u(s^u(\tau))\,d\tau\right]\right|+\lim_{t\to\infty}\left|E\left[\left(\frac{1}{t}-\frac{1}{T^u(t)}\right)\int_t^{T^u(t)}k^u(s^u(\tau))\,d\tau\right]\right|=0 \tag{3.27}$$

The two summands in the right hand side of (3.27) converge to zero because they are bounded above by $E[T^u(t) - t]M/t$ which is bounded by $BM/t$ (Theorem 3.2). Equations (3.26), (3.27) complete the proof of (3.18). ∎

Remark: If $s(0) = s_o$, then (3.18) is obviously true for $n = 0$ as well.

The last result of this section shows that we may restrict to control laws in $U_R$ without increasing the optimal value of the cost functional. To avoid certain technicalities, we only present an indirect and informal argument.

Theorem 3.4
$$\inf_{u \in U_R} g^u = \inf_{u \in U_A} g^u = g^* \tag{3.28}$$

Proof: (Outline) View our control problem as follows: Each time $T_n$ the state $s_o$ is reached, a policy $u_n \in U_A$ to be followed in $[T_n, T_{n+1})$ is chosen. We then have a single state semi-Markov renewal programming problem with an infinite action space and bounded costs per stage; regenerative control laws correspond to stationary policies of the semi–Markov problem. Moreover, $T_n - T_{n-1}$ is uniformly bounded, in expected value, for all policies of the semi–Markov problem. It follows that stationary policies exist that come arbitrarily close to being optimal. By translating this statement to the original problem, we obtain (3.28). ∎

## 4. THE VALUE FUNCTION OF DYNAMIC PROGRAMMING

Using the recurrence properties of control laws in $U_R$, we may now define value (cost–to–go) functions of dynamic programming. This is done by using the recurrent state $s_o$ as a reference state. Moreover, we exploit theorem 3.3 to convert the average cost problem to a total cost problem.

Similarly with Ref. 24, we define the value function $V^u{:}S \mapsto R$, corresponding to a control law $u \in U_R$ by

$$V^u(s) = E\left[\int_0^{T_1^u} (k^u(s^u(\tau)) - g^u)\, d\tau \,|\, s^u(0) = s\right] \tag{4.1}$$

In view of Theorem 3.3, we have $V^u(s_o) = 0$, for all $u \in U_R$. We also define an auxiliary value function $\hat{V}^u(s)$ by

$$\hat{V}^u(s) = E\left[\int_0^{T_1^u} (k^u(s^u(\tau)) - g^*)\, d\tau \,|\, s^u(0) = s\right] \tag{4.2}$$

and the optimal value function $V^*(s)$ by

$$V^*(s) = \inf_{u \in U_R} \hat{V}^u(s) \tag{4.3}$$

The above defined functions are all bounded by $2BM$, where $M$ is a constant bounding $|k^u(s)|$ and $B$ is the constant of theorem 3.2.

Lemma 4.1: a) $0 \le \hat{V}^u(s) - V^u(s) \le (g^u - g^*)B, \quad \forall s \in S$.

b) $\hat{V}^u(s_o) = (g^u - g^*)E[T_1^u \,|\, s^u(0) = s_o]$.

c) $g^u = g^*$ iff $\hat{V}^u(s_o) = 0$.

d) $\hat{V}^u(s_o) \ge 0, \quad V^*(s_o) = 0$.

Proof: Follows directly from the definitions and the inequality $E[T_1^u] \le B$. ∎

Lemma 4.1c shows that an average cost optimal control law is one that minimizes $\hat{V}^u(s_o)$. This will certainly be the case if a control law minimizes $\hat{V}^u(s)$ for all $s \in S$, which is a stronger requirement. It is possible to show that if $u \in U_R \cap U_M$ is optimal, $u$ should minimize $\hat{V}^u(s)$ for all $s$, except, possibly, for a subset of $S$ of zero steady-state probability measure. This shows that minimization of $\hat{V}^u(s)$, for all $s$, is not a much stronger requirement than minimization of $g^u$. Moreover, minimization of $\hat{V}^u(s)$ is now a problem of total expected cost minimization which may be handled through traditional techniques.

We will say that a control law $u \in U_R$ is everywhere optimal if $\hat{V}^u(s) = V^*(s)$, $\forall s \in S$; optimal if $g^u = g^*$.

We conclude this section with a few properties of $\hat{V}^u$ and $V^u$ that will be needed in the next section.

Lemma 4.2: a) For any positive integers $m$ and $n$ such that $n \geq m$ and any $u \in U_R$

$$E\left[\int_{T_m^u}^{T_n^u} (k^u(s^u(\tau)) - g^u)\, d\tau\right] = 0 \tag{4.4}$$

b) For any positive integer $n$ and any $u \in U_R$

$$V^u(s) = E\left[\int_0^{T_n^u} (k^u(s^u(\tau)) - g^u)\, d\tau \,|\, s^u(0) = s\right] \tag{4.5}$$

Proof: Both parts follow immediately from Theorem 3.3. ∎

The following result is essentially a version of equality (4.5) with random sampling of the upper limit of integration.

Lemma 4.3: Let $u$, $v$, $w \in U_R$; let $s^u(0)$, $s^v(0)$, $s^w(0)$ be three states with the same value of $\alpha$. Let $T^{uvw}$ be as defined in section 3. Then,

$$V^u(s) = E\left[\int_0^{T^{uvw}} (k^u(s^u(\tau)) - g^u)\, d\tau \,|\, s^u(0) = s\right] \tag{4.6}$$

Proof: Let $T = \min\{T_n^u\colon T_n^u \geq T^{uvw}\}$ and let $\chi_n$ be the characteristic (indicator) function of the set of those $\omega \in \Omega$ such that $T > T_n$. We then have (using the dominated convergence theorem to interchange summation and expectation)

$$E\left[\int_0^{T} (k^u(s^u(\tau)) - g^u)\, d\tau\right] = E\left[\int_0^{T_1^u} (k^u(s^u(\tau)) - g^u)\, d\tau\right] + \sum_{n=1}^{\infty} E\left[\chi_n \int_{T_n^u}^{T_{n+1}^u} (k^u(s^u(\tau)) - g^u)\, d\tau\right] \tag{4.7}$$

The random variable $\chi_n$ is $\mathcal{A}_{T_n^u}$-measurable. Therefore,

$$E\left[\chi_n \int_{T_n^u}^{T_{n+1}^u} (k^u(s^u(\tau)) - g^u)\, d\tau\right] = E\left[\chi_n E\left[\int_{T_n^u}^{T_{n+1}^u} (k^u(s^u(\tau)) - g^u)\, d\tau \,|\, \mathcal{A}_{T_n^u}\right]\right] = 0 \tag{4.8}$$

The second equality in (4.8) follows from Lemma 4.2a and the assumption that $u$ regenerates at time $T^u_n$, assumption (S6) in particular. For the same reasons we obtain:

$$E\left[\int_{T^{uvw}}^{T} (k^u(s^u(\tau)) - g^u)\, d\tau\right] = E\left[\int_{T_1^u}^{T_2^u} (k^u(s^u(\tau)) - g^u)\, d\tau\right] = 0. \qquad (4.9)$$

Combining (4.7), (4.8), (4.9) and using the definition of $V^u$, we obtain

$$E\left[\int_{0}^{T^{uvw}} (k^u(s^u(\tau)) - g^u)\, d\tau\right] = E\left[\int_{0}^{T_1^u} (k^u(s^u(\tau)) - g^u)\, d\tau\right] = V^u(s^u(0)). \qquad (4.10)$$

∎

The last Lemma is an elementary consequence of our definitions:

Lemma 4.4: Given some $s \in S$ and $\epsilon > 0$, $\exists\, u \in U_R$ such that $\hat{V}^u(s) \leq V^*(s) + \epsilon$ and $g^u \leq g^* + \epsilon$.
Proof: (Outline) Assume $s \neq s_o$. Then $\hat{V}^u$ depends on the choice of the control variables up to time $T_1^u$ and $g^u$ depends on the choice after that time. The control variables before and after $T_1^u$ may be independently chosen so as to satisfy both inequalities. If $s = s_o$, choose $u$ such that $g^u \leq g^* + \min\{\epsilon, \epsilon/B\}$. Then, $\hat{V}^u(s_o) \leq V^*(s_o) + \epsilon$. ∎

## 5. CONVEXITY AND OTHER PROPERTIES OF $V^*$.

In this section, we exploit the structure of our system to obtain certain basic properties of $V^*$. These properties, together with the optimality conditions, to be derived in section 7, lead directly to the characterization of optimal control laws.

Theorem 5.1: $V^*(\underline{x}, \underline{\alpha})$ is convex, $\forall \underline{\alpha} \in \{0,1\}^3$.

Proof: Let $s^u(0) = (\underline{x}^u, \underline{\alpha})$ and $s^v(0) = (\underline{x}^v, \underline{\alpha})$ be two states in $S$ with the same value of $\underline{\alpha}$. Let $c \in [0,1]$ and $s^w(0) = (c\underline{x}^u + (1-c)\underline{x}^v, \underline{\alpha})$. Then $s^w(0) \in S$, because $[0, N_1] \times [0, N_2]$ is a convex set, and we need to show that

$$V^*(s^w(0)) \leq cV^*(s^u(0)) + (1-c)V^*(s^u(0)) \tag{5.1}$$

Fix some $\epsilon > 0$ and let $u$, $v$ be control laws in $U_R$ such that

$$g^u \leq g^* + \epsilon \qquad g^v \leq g^* + \epsilon \tag{5.2}$$

$$\hat{V}^u(s^u(0)) \leq V^*(s^u(0)) + \epsilon \qquad \hat{V}^v(s^v(0)) \leq V^*(s^v(0)) + \epsilon \tag{5.3}$$

(Such control laws exist by Lemma 4.4.) Let $s^u(\omega, t)$ and $s^v(\omega, t)$ be the corresponding sample paths. We now define a control law $w$ to be used starting from the initial state $s^w(0)$. Let, for $i = 1, 2$,

$$\lambda_i^w(\omega, t) = c\lambda_i^u(\omega, t) + (1-c)\lambda_i^v(\omega, t) \tag{5.4}$$

$$\mu_i^w(\omega, t) = c\mu_i^u(\omega, t) + (1-c)\mu_i^v(\omega, t) \tag{5.5}$$

With $w$ defined by (5.4), (5.5) assumptions (S1)-(S4) are satisfied because these assumptions are satisfied by $u$ and $v$. Moreover, by linearity of the dynamics,

$$x_i^w(\omega, t) = cx_i^u(\omega, t) + (1-c)x_i^v(\omega, t) \tag{5.6}$$

Since $\underline{x} = (0,0)$ is an extreme point of $[0, N_1] \times [0, N_2]$, equation (5.6) implies that whenever $s^w(t) = s_o$, we also have $s^u(t) = s^v(t) = s_o$. Therefore, $T_1^w = T^{uvw}$ and consequently, $w \in U_A$. Moreover, $u$

and $v$ regenerate whenever $s^w(t) = s_o$ and, therefore, $w \in U_R$. Using (5.6) and the convexity of the cost function we obtain

$$\int_0^{T_1^w} (k^w(s^w(\omega,\tau)) - g^*)\,d\tau \leq c\int_0^{T_1^w} (k^u(s^u(\omega,\tau)) - g^*)\,d\tau + (1-c)\int_0^{T_1^w} (k^v(s^v(\omega,\tau)) - g^*)\,d\tau \quad (5.7)$$

We take expectations of both sides of inequality (5.7) and rearrange it to obtain

$$V^*(s^w(0)) \leq \hat{V}^w(s^w(0)) \leq cE\left[\int_0^{T_1^w} (k^u(s^u(\omega,\tau)) - g^u)\,d\tau \,|\, s^u(0)\right] + \\ (1-c)E\left[\int_0^{T_1^w} (k^v(s^v(\omega,\tau)) - g^v)\,d\tau \,|\, s^v(0)\right] + (cg^u + (1-c)g^v - g^*)E[T_1^w] \quad (5.8)$$

Since $T^{uvw} = T_1^w$, Lemma 4.3 applies. Using also inequalities (5.2), (5.3) and Lemma 4.1c, we obtain

$$V^*(s^w(0)) \leq cV^u(s^u(0)) + (1-c)V^v(s^v(0)) + \epsilon B \leq c\hat{V}^u(s^u(0)) + (1-c)\hat{V}^v(s^v(0)) + \epsilon B \leq \\ cV^*(s^u(0)) + (1-c)V^*(s^v(0)) + \epsilon(1+B) \quad (5.9)$$

Since $\epsilon$ was arbitrary, we may let $\epsilon \downarrow 0$ in (5.9) to obtain inequality (5.1). ∎

It is not hard to show that if $f_\alpha$ (defined by equation (2.14)) is srictly convex, then the inequality (5.1) is strict. In fact, it is also true that (5.1) is a strict inequality even if $f_\alpha$ is linear. A detailed proof would be fairly involved and we only give here an outline.

With control law $w$, defined by (5.4), (5.5), there is positive probability that $\underline{\alpha}(t) = (0,1,1)$, $x_1^w(t) \neq 0$, $x_1^u(t) = 0$, in which case $\mu_1^w(t) < \mu_1{}^*$, for all $t$ belonging to a time interval of positive length. We can also show (in a way similar to the proof of Theorem 3.1) that any control law with the above property does not minimize $\hat{V}^w(s^w(0))$ and that $V^*(s^w(0)) < \hat{V}^w(s^w(0)) - \delta$, for some $\delta$ independent of $\epsilon$. Using this inequality in (5.8) and (5.9), (5.1) becomes a strict inequality.

Let $f_\alpha$ be the function defined by (2.15) and let $M$ be such that

$$|f_\alpha(x_1,x_2) - f_\alpha(x_1+\Delta_1, x_2+\Delta_2)| \leq M(|\Delta_1| + |\Delta_2|) \quad (5.10)$$

Such a $M$ exists, since $f_\alpha$ is Lipschitz continuous. We let $V_\alpha{}^*$ denote $V^*(\underline{x},\alpha)$ and we have the following result:

<u>Theorem 5.2:</u> Let $\Delta_1 \geq 0$, $\Delta_2 \geq 0$, $\Delta_1\Delta_2 \neq 0$. Then,

$$-c_1\Delta_1 - c_2\Delta_2 < V_\alpha{}^*(x_1+\Delta_1, x_2+\Delta_2) - V_\alpha{}^*(x_1,x_2) < MB(\Delta_1+\Delta_2) \quad (5.11)$$

$\forall \alpha \in \{0,1\}^3$. In particular, $V_\alpha{}^*$ is Lipschitz continuous and if $f_\alpha \equiv 0$, then $V_\alpha{}^*$ is strictly decreasing in each variable.

Proof: The two inequalities in (5.11) will be proved separately. Without loss of generality, we assume that $\Delta_2 = 0$ and we start by proving the second inequality.

a) Fix two initial states $s^u(0) = (x_1, x_2, \underline{\alpha})$ and $s^w(0) = (x_1 + \Delta, x_2, \underline{\alpha})$, $\Delta > 0$ with the same value of $\underline{\alpha}$. Let $u \in U_R$ be such that (Lemma 4.4)

$$\hat{V}^u(s^u(0)) \leq V^*(s^u(0)) + \epsilon, \qquad g^u \leq g^* + \epsilon. \tag{5.12}$$

We now define a new control law $w \in U_R$ to be used starting from $s^w(0)$ as follows:

$$\lambda_1^w(\omega,t) = \begin{cases} 0 & \text{if } x_1^w(\omega,t) \neq x_1^u(\omega,t) \\ \lambda_1^u(\omega,t) & \text{if } x_1^w(\omega,t) = x_1^u(\omega,t) \end{cases} \tag{5.13a}$$

$$\mu_1^w(\omega,t) = \begin{cases} \alpha_1(\omega,t)\mu_1{}^* & \text{if } x_1^w(\omega,t) \neq x_1^u(\omega,t) \\ \mu_1^u(\omega,t) & \text{if } x_1^w(\omega,t) = x_1^u(\omega,t) \end{cases} \tag{5.13b}$$

$$\lambda_2^w(\omega,t) = \lambda_2^u(\omega,t) \qquad \mu_2^w(\omega,t) = \mu_2^u(\omega,t) \tag{5.13c}$$

(Intuitively said, with control law $w$, no material is routed to buffer $B_1$ and machine $M_1$ operates at full capacity until the buffer level $x_1^w(\omega,t)$ decreases enough to become the same as $x_1^u(\omega,t)$. From that time on, the two sample paths coincide.) Then, $w \in U_R$ and has the following properties:

$$x_1^u(\omega,t) \leq x_1^w(\omega,t) \leq x_1^u(\omega,t) + \Delta \tag{5.14}$$

$$\mu_1^w(\omega,t) \geq \mu_1^u(\omega,t) \tag{5.15}$$

Then, $T^{uw} = T_1^w$ and

$$\begin{aligned} \int_0^{T_1^w} k^w(s^w(\omega,\tau))\,d\tau = &\int_0^{T_1^w} \big(f_\alpha(x^w(\omega,\tau)) - c_1\mu_1^w(\omega,\tau) - c_2\mu_2^w(\omega,\tau)\big)\,d\tau \leq \\ &\int_0^{T_1^w} \big(f_\alpha(x^u(\omega,\tau)) + M\Delta - c_1\mu_1^u(\omega,\tau) - c_2\mu_2^u(\omega,\tau)\big)d\tau = \\ &\int_0^{T_1^w} k^u(s^u(\omega,\tau))\,d\tau + M\Delta T_1^w \end{aligned} \tag{5.16}$$

We claim that there exists a set $A \subset \Omega$ of positive probability measure such that inequality (5.16) is strict for all $\omega \in A$. Namely consider all those $\omega$ for which $\underline{a}(\omega, t)$ becomes $(0,1,1)$ before time $\Delta/2(\lambda^* + \mu_1^*)$ and stays equal to $(0,1,1)$ until time $T_1^w$. Let $\delta > 0$ be such that $\Pr(\omega \in A) > \delta$. For all $\omega \in A$, we have

$$\int_0^{T_1^w} c_1 \mu_1^w(\omega, \tau)\, d\tau > \int_0^{T_1^w} c_1 \mu_1^u(\omega, \tau)\, d\tau + c_1 \frac{\Delta}{2} \tag{5.17}$$

and consequently,

$$\int_0^{T_1^w} k^w(s^w(\omega, \tau))\, d\tau < \int_0^{T_1^w} k^u(s^u(\omega, \tau))\, d\tau + M\Delta T_1^w - c_1 \frac{\Delta}{2}, \qquad \omega \in A. \tag{5.18}$$

Then, (5.16) may be strengthened as follows: Taking expectations in (5.16) and using (5.18) we obtain

$$E\left[\int_0^{T_1^w} (k^w(s^w(\omega, \tau)) - g^*)\, d\tau\right] \leq E\left[\int_0^{T_1^w} (k^u(s^u(\omega, \tau)) - g^*)\, d\tau\right] + M\Delta E[T_1^w] - \frac{c_1 \delta \Delta}{2} \tag{5.19}$$

Using Lemma 4.3 and following the same steps as in the proof of Theorem 5.1 we obtain

$$V^*(s^w(0)) \leq \hat{V}^w(s^w(0)) \leq V^*(s^u(0)) + \epsilon(1 + B) + M\Delta B - \delta c_1 \Delta/2 \tag{5.20}$$

Since $\epsilon$ was arbitrary, we may let $\epsilon$ decrease to zero to obtain the second inequality in (5.11).
b) For the proof of the left-hand-side of (5.11), let $s^u(0)$ and $s^w(0)$ be as before and let $w \in U_R$ be such that

$$\hat{V}^w(s^w(0)) \leq V^*(s^w(0)) + \epsilon, \qquad g^w \leq g^* + \epsilon, \tag{5.21}$$

and define $u \in U_R$ (to be used starting from $s^u(0)$) as follows:

$$\lambda_1^u(\omega, t) = \begin{cases} \lambda^* - \lambda_2^w(\omega, t) & \text{if } x_1^u(\omega, t) \neq x_1^w(\omega, t) \\ \lambda_1^w(\omega, t) & \text{if } x_1^u(\omega, t) = x_1^w(\omega, t) \end{cases} \tag{5.22a}$$

$$\mu_1^u(\omega, t) = \begin{cases} 0 & \text{if } x_1^u(\omega, t) \neq x_1^w(\omega, t) \\ \mu_1^w(\omega, t) & \text{if } x_1^u(\omega, t) = x_1^w(\omega, t) \end{cases} \tag{5.22b}$$

$$\lambda_2^u(\omega,t) = \lambda_2^w(\omega,t) \qquad \mu_2^u(\omega,t) = \mu_2^w(\omega,t) \tag{5.22c}$$

(So, control law $u$ sends as much as possible material to $B_1$ and machine $M_1$ is not being operated until the level $x_1^u(\omega,t)$ rises to $x_1^w(\omega,t)$.) Similarly with part (a) we obtain $T^{uw} = T_1^u = T_1^w$ and

$$\int_0^{T^{uw}} k^u(s^u(\omega,\tau))\,d\tau \leq \int_0^{T^{uw}} k^w(s^w(\omega,\tau))\,d\tau + c_1\Delta \tag{5.23}$$

Consider the set $A \subset \Omega$ of those $\omega$ such that $\underline{\alpha}$ becomes (1,0,0) within $\Delta/2(\lambda^* + \mu_1^*)$ time units and stays equal to (1,0,0) for at least $(N_1 + N_2)/\lambda^*$ additional time units. For any $\omega \in A$ we will have

$$\int_0^{T^{uw}} k^u(s^u(\omega,\tau))\,d\tau \leq \int_0^{T^{uw}} k^w(s^w(\omega,\tau))\,d\tau + c_1\Delta/2 \tag{5.24}$$

Taking expectations and following the same procedure as in part (a), we establish the desired result. ∎

Corollary 5.1: If $f_\alpha \equiv 0$ (and therefore $M = 0$ in (5.11)) then

$$-c_1 \leq \lim_{\Delta \downarrow 0} \frac{V_\alpha^*(x_1+\Delta, x_2) - V_\alpha^*(x_1,x_2)}{\Delta} \leq 0 \tag{5.25}$$

The first inequality is strict for all $x_1 \neq 0$; the second, for all $x_1 \neq N_1$. Similar inequalities hold for the second argument of $V_\alpha^*(x_1,x_2)$ or if $\Delta \uparrow 0$.

Proof: The existence of the limit follows from the convexity of $V_\alpha^*$. Then, (5.25) follows from theorem 5.2. The strict inequalities follow from the strictness of the inequalities in (5.11) and the convexity of $V_\alpha^*$. ∎

## 6. DEPENDENCE OF $g^*$ ON THE SYSTEM PARAMETERS

We have shown (Theorem 3.3) that the optimal cost $g^*$ is independent of the initial state. In this section, we view $g^*$ as a function of the parameters of the system and examine the form of the functional dependence. In particular, we consider the dependence of $g^*$ on the buffer sizes $N_1$ and $N_2$ as well as the machine capacities $\lambda^*$, $\mu_1^*$, $\mu_2^*$. To illustrate this dependence, we write $g^*(N_1, N_2, \lambda^*, \mu_1^*, \mu_2^*)$. Our result states that $g^*$ is a convex function of its parameters.

The proof is similar to the proofs given in section 5, but simpler. Despite the simplicity of the proof, we are not aware of any similar result on controlled queueing systems.

Theorem 6.1: $g^*(N_1, N_2, \lambda^*, \mu_1{}^*, \mu_2{}^*)$ is a convex function of its arguments.

Proof: Let $c \in [0, 1]$ and let $(N_{1j}, N_{2j}, \lambda_j{}^*, \mu_{1j}{}^*, \mu_{2j}{}^*)$, $j = 1, 2, 3$ be three sets of (positive) parameters such that

$$N_{i2} = cN_{i1} + (1-c)N_{i3} \qquad i = 1, 2 \tag{6.1a}$$

$$\lambda_2{}^* = c\lambda_1{}^* + (1-c)\lambda_3{}^* \tag{6.1b}$$

$$\mu_{i2}{}^* = c\mu_{i1}{}^* + (1-c)\mu_{i3}{}^* \qquad i = 1, 2 \tag{6.1c}$$

Let, for simplicity, the initial state be $s_o$. Let $u_1$, $u_3$ be two control laws satisfying (S1)-(S4) with the parameters of the system being $(N_{11}, N_{21}, \lambda_1{}^*, \mu_{11}{}^*, \mu_{21}{}^*)$ and $(N_{13}, N_{23}, \lambda_3{}^*, \mu_{13}{}^*, \mu_{23}{}^*)$, respectively. We define a new control law $u_2$ by:

$$\lambda_i^{u_2}(\omega, t) = c\lambda_i^{u_1}(\omega, t) + (1-c)\lambda_i^{u_3}(\omega, t) \qquad i = 1, 2 \tag{6.2a}$$

$$\mu_i^{u_2}(\omega, t) = c\mu_i^{u_1}(\omega, t) + (1-c)\mu_i^{u_3}(\omega, t) \qquad i = 1, 2 \tag{6.2b}$$

This definition of $u_2$, for each $\omega$, in terms of $u_1$, $u_3$ is legitimate because the underlying probability space $(\Omega, \mathcal{A}, \mathcal{P})$ is fixed and independent of the control law being used or the parameters considered in this section. It follows that $u_2$ satisfies (S1)-(S4), the parameters of the system being $(N_{12}, N_{22}, \lambda_2{}^*, \mu_{12}{}^*, \mu_{22}{}^*)$. Moreover,

$$\underline{x}^{u_2}(\omega, t) = c\underline{x}^{u_1}(\omega, t) + (1-c)\underline{x}^{u_3}(\omega, t) \tag{6.3}$$

The convexity of the cost function implies that

$$\int_0^t k^{u_2}(s^{u_2}(\omega, \tau))\, d\tau \leq c \int_0^t k^{u_1}(s^{u_1}(\omega, \tau))\, d\tau + (1-c) \int_0^t k^{u_3}(s^{u_3}(\omega, \tau))\, d\tau \tag{6.4}$$

We now take expectations in (6.4), divide by $t$ and then take the limit, as $t \to \infty$, to obtain

$$g^*(N_{12}, N_{22}, \lambda_2^*, \mu_{12}^*, \mu_{22}^*) \le g^{u_2} \le cg^{u_1} + (1-c)g^{u_3} \tag{6.5}$$

Since $u_1$ and $u_3$ were arbitrarily chosen, we may take the infimum in (6.5) to obtain

$$g^*(N_{12}, N_{22}, \lambda_2^*, \mu_{12}^*, \mu_{22}^*) \le cg^*(N_{11}, N_{21}, \lambda_1^*, \mu_{11}^*, \mu_{21}^*) + (1-c)g^*(N_{13}, N_{23}, \lambda_3^*, \mu_{13}^*, \mu_{23}^*) \tag{6.6}$$

∎

We now consider the consequences of this theorem on the problem of optimally chosing the parameters of the system (capacity assignment problem). It is trivial to prove that $g^*$ decreases as each of the system parameters increases. On the other hand, higher values of the parameters usually mean higher capital costs. If these capital costs $K(N_1, N_2, \lambda^*, \mu_1^*, \mu_2^*)$ are assumed to be convex, the problem consists of minimizing $(g^* + K)(N_1, N_2, \lambda^*, \mu_1^*, \mu_2^*)$, a convex function. Convexity is a nice property to have in this situation, because any local minimum will also be a global one and iterative procedures are applicable. Of course, this presupposes that we are able to evaluate $g^*$ for any given parameter set, as well as the direction of change of $g^*$ as the parameter values are varied. While this is in principle possible, the computational requirements may become very demanding and further research is needed on this issue.

## 7. NECESSARY CONDITIONS FOR OPTIMALITY

In this section we prove the necessary conditions for optimality that will be used in the next section. We start by demonstrating that $V^*$ is in the domain of $\mathcal{L}^u$ (defined in section 2), for any admissible control law $u$.

<u>Lemma 7.1:</u> Let $\underline{x}(t)$ be a trajectory in $[0, N_1] \times [0, N_2]$ and suppose that $\underline{d} \equiv \lim_{t\downarrow 0}(\underline{x}(t) - \underline{x}(0))/t$ exists. Then,

$$\lim_{t\downarrow 0} \frac{V_\alpha{}^*(\underline{x}(t)) - V_\alpha{}^*(\underline{x}(0))}{t} = \lim_{t\downarrow 0} \frac{V_\alpha{}^*(\underline{x}(0) + t\underline{d}) - V_\alpha{}^*(\underline{x}(0))}{t} \tag{7.1}$$

(The existence of the limits is part of the result.)

<u>Proof:</u> We first note that the limit in the right–hand–side of (7.1) exists, by convexity of $V_\alpha{}^*$ and is finite, by the Lipschitz continuity of $V_\alpha{}^*$. By Lipschitz continuity again, there exists a constant $C$ such that

$$\left| \frac{V_\alpha{}^*(\underline{x}(t)) - V_\alpha{}^*(\underline{x}(0))}{t} - \frac{V_\alpha{}^*(\underline{x}(0) + t\underline{d}) - V_\alpha{}^*(\underline{x}(0))}{t} \right| \leq C \left\| \frac{\underline{x}(t) - \underline{x}(0)}{t} - \underline{d} \right\| \tag{7.2}$$

The right–hand–side of (7.2) converges to zero, as $t\downarrow 0$, from the definition of $\underline{d}$, thus proving the lemma. ∎

Let $u \in U_R$. For any fixed $\underline{\alpha}$, let, as in section 2, $\underline{x}_\alpha^u(t)$ be the value of $\underline{x}$ at time $t$ if no jump occurs until time $t$. By right–continuity and boundedness of the control variables, the trajectory $\underline{x}_\alpha^u(t)$ possesses right–hand–side derivatives which are uniformly bounded. Then, Lemma 7.1 and Proposition 2.1 imply:

<u>Theorem 7.1:</u> $V^*$ belongs to the domain of $\mathcal{L}^u$, for any $u \in U_R$.

<u>Lemma 7.2</u> For any $\epsilon > 0$, there exists some $w \in U_R$ such that

$$\hat{V}^w(s) \leq V^*(s) + \epsilon, \qquad \forall s \in S. \tag{7.3}$$

<u>Proof:</u> (Outline) Partition the state space $S$ into a finite collection of disjoint and small enough rectangles $R_1, \ldots, R_k$. Choose a state $s_j \in R_j$ and a control law $w_j \in U_R$ such that $\hat{V}^{w_j}(s_j) \leq V^*(s_j) + \epsilon_1$, where $\epsilon_1$ is small enough. Define $w_j$ for all initial states on the rest of the rectangle $r_j$ so that all sample paths $s^{w_j}(\omega, t)$, $w_j(\omega, t)$ starting from $R_j$ stay close enough. In particular, choose $w_j$ in such

a way that $s^{w_j}(\omega, t)$ and $\mu_i^{w_j}(\omega, t)$ are continuous functions of the initial state, for any $\omega, t$. In that case, $|\hat{V}^{w_j}(s_j) - \hat{V}^{w_j}(s)| \leq \epsilon_2, \quad \forall s \in R_j$, for some small enough $\epsilon_2$. Then, define a control law $w$ by lumping together control laws $w_j$, $j = 1, \ldots, k$. Given that $V^*$ is Lipschitz continous and since $\epsilon_1$, $\epsilon_2$ may be chosen as small as desired, $w$ satisfies (7.3). ∎

Lemma 7.3: $\lim_{t\downarrow 0}(1/t)E[V^*(s^u(t))\chi] = 0$, where $\chi$ is the indicator function of the event $T_1^u < t$.

Proof:

$$\lim_{t\downarrow 0} \frac{E[V^*(s^u(t))\chi]}{t} = \lim_{t\downarrow 0} \frac{Pr(T_1^u < t)}{t} \lim_{t\downarrow 0} E[V^*(s^u(t)) \,|T_1^u < t] \tag{7.4}$$

The first limit in the r.h.s. of (7.4) is bounded by the transition rates $p_i$, $r_i$; the second one is equal to $V^*(s_o) = 0$, unless a jump occurs in $[T_1^u, t]$, which is an event whose probability goes to zero, as $t$ goes to zero. ∎

Lemma 7.4: $\mathcal{L}^u V^* + k^u \geq g^* \qquad \forall s \in S, \quad \forall u \in U_R$

Proof: Let $u \in U_R$, $t > 0$, $s \in S$ be fixed and let $w$ be the control law of Lemma 7.2. Consider a new control law $v$ with the following properties: $v$ coincides with $u$ up to time $t$; at that time the past is "forgotten" and the process is restarted using control law $w$. Then,

$$\begin{aligned} V^*(s) \leq \hat{V}^v(s) =& E\left[\int_0^{\min\{t,T_1^u\}} (k^u(s^u(\tau)) - g^*)\, d\tau\right] + E[\hat{V}^w(s^u(t))(1-\chi)] \leq \\ & E\left[\int_0^{\min\{t,T_1^u\}} (k^u(s^u(\tau)) - g^*)\, d\tau\right] + E[V^*(s^u(t))(1-\chi)] + \epsilon \end{aligned} \tag{7.5}$$

Since $\epsilon$ was arbitrary, we may let $\epsilon \downarrow 0$, then divide by $t$, take the limit, as $t \downarrow 0$, and invoke Lemma 7.3 to obtain

$$\mathcal{L}^u V^*(s) = \lim_{t\downarrow 0} \frac{E[V^*(s^u(t))] - V^*(s)}{t} \geq -\lim_{t\downarrow 0} \frac{1}{t} E\left[\int_0^{\min\{t,T_1^u\}} (k^u(s^u(\tau)) - g^*)\, d\tau\right] = -k^u(s) + g^* \tag{7.6}$$

(the last equality follows from the right–continuity of $k^u$ and the dominated convergence theorem). ∎

Theorem 7.2: If $u \in U_R \cap U_M$ is everywhere optimal, then

$$(\mathcal{L}^u V^* + k^u)(s) \leq (\mathcal{L}^w V^* + k^w)(s) \qquad \forall w \in U_R, \forall s \in S \tag{7.7}$$

Proof: We start with the equation $\mathcal{L}^u\hat{V}^u + k^u = g^*$. (This equation is derived the same way as Lemma 7.4, except that inequalities become equalities.) Since $u$ is everywhere optimal, $\hat{V}^u = V^*$ and (using Lemma 7.4) $\mathcal{L}^u V^* + k^u = g^* \leq \mathcal{L}^w V^* + k^w$, for all $w \in U_R$. ∎

It is also possible to prove that if $u \in U_M$ and (7.6) holds for all $s \in S$ then $u$ is everywhere optimal. However, this result will not be needed and the proof is omitted.

## 8. CHARACTERIZATION OF OPTIMAL CONTROL LAWS

In this section we use the optimality conditions (theorem 7.2) together with the properties of $V^*$ (theorems 5.1 and 5.2) to characterize everywhere optimal control laws. We mainly consider Markovian control laws for which the control variables $\lambda_i$, $\mu_i$ can be viewed as functions of the state. The first two theorems (8.1 and 8.2) state that the machines should be always operated at the maximum rate allowed, as it should be expected. Theorem 8.4 is much more substantial, as it characterizes the way that the flow through machine $M_o$ should be split.

Theorem 8.1: If $u \in U_M \cap U_R$ is everywhere optimal, then

a) $$\mu_i^u(\underline{x}, \underline{a}) = a_i \mu_i^* \quad \text{if} \quad x_i \neq 0, \; i = 1, 2. \tag{8.1}$$

b) $$\mu_i^u(\underline{x}, \underline{a}) = a_i \min\{\mu_i^*, \lambda_i^u(\underline{x}, \underline{a})\} \quad \text{if} \quad x_i = 0, \; i = 1, 2. \tag{8.2}$$

Proof: Let $u \in U_M \cap U_R$ be everywhere optimal . Then, $u$ must minimize $(\mathcal{L}^u V^* + k^u)(s)$, $\forall s \in S$. Using Proposition 2.1 and dropping those terms that do not depend on $u$ we conclude that $\mu_1^u$, $\mu_2^u$ must be chosen so as to minimize

$$\lim_{\Delta \downarrow 0} \frac{V_a^*(x_1 + (\lambda_1^u - \mu_1^u)\Delta, x_2 + (\lambda_2^u - \mu_2^u)\Delta) - V_a^*(x_1, x_2)}{\Delta} - c_1 \mu_1^u - c_2 \mu_2^u \tag{8.3}$$

Let $x_i \neq 0$ and assume that $a_i \neq 0$. By Corollary 5.1, the slopes of $V_a^*$ are strictly larger than $-c_1$, $-c_2$ and, as a result, $\mu_i^u$ must be set to its highest admissible values which is $a_i \mu_i^*$. If $x_i = 0$, (8.2) follows because otherwise the right–continuity assumption (S2) would be violated. ∎

Theorem 8.2: If $V_a^*$ is strictly decreasing in each variable (or in particular, by Theorem 5.2, if $k^u = -c_1 \mu_1^u - c_2 \mu_2^u$) and if $u \in U_M \cap U_R$ is everywhere optimal, then

a) $$\lambda^u(\underline{x}, \underline{a}) = a_o \lambda^* \quad \text{if} \quad \underline{x} \neq (N_1, N_2) \tag{8.4}$$

b) $$\lambda^u(\underline{x}, \underline{a}) = a_o \min\{\lambda^*, a_1 \mu_1^* + a_2 \mu_2^*\} \quad \text{if} \quad \underline{x} = (N_1, N_2) \tag{8.5}$$

Proof: Let $u \in U_M \cap U_R$ be everywhere optimal. Then, $u$ must minimize $(\mathcal{L}^u V^* + k^u)(s)$, $\forall s \in S$. Theorem 8.1 determines $\mu_i^u$ uniquely and $k^u$ is no more dependent on $u$. By dropping those terms that do not depend on $u$ we conclude that $\lambda_1^u$, $\lambda_2^u$ must be chosen so as to minimize

$$\lim_{\Delta \downarrow 0} \frac{V_a^*(x_1 + (\lambda_1^u - \mu_1^u)\Delta, x_2 + (\lambda_2^u - \mu_2^u)\Delta) - V_a^*(x_1, x_2)}{\Delta} \tag{8.6}$$

Let $(x_1, x_2) \neq (N_1, N_2)$. Since $V_a^*$ is strictly decreasing, $\lambda^u = \lambda_1^u + \lambda_2^u$ must be set equal to its highest admissible value which is $a_o \lambda^*$. If $(x_1, x_2) = (N_1, N_2)$, equation (8.5) follows because otherwise the right–continuity requirement (S2) would be violated. ∎

If $f_\alpha(\underline{x})$ is nonzero and large enough, compared to $c_1$, $c_2$, then $V_\alpha{}^*$ need not be decreasing. Equivalently, the penalty for having large buffer levels will be larger than the future payoff in terms of increased production. In that case, for any optimal control law, $\lambda^u$ should be set to zero whenever the buffer levels exceed some threshold.

From now on, we assume that $V_\alpha{}^*$ is strictly decreasing, for all $\underline{\alpha}$. Theorems 8.1 and 8.2 define $\mu_1^u$, $\mu_2^u$, $\lambda^u$ uniquely. It only remains to decide on how is $\lambda^u$ going to be split. It is here that convexity of $V_\alpha{}^*$ plays a major role. Note that there is no such decision to be made whenever $\alpha_o = 0$. We will therefore assume that $\alpha_o \neq 0$.

Let

$$h_\alpha(\rho) = \inf\{V_\alpha{}^*(x_1, x_2): x_1 + x_2 = \rho\}, \quad \rho \in [0, N_1 + N_2]. \tag{8.7}$$

Because of the continuity of $V_\alpha{}^*$, the infimum is attained, for each $\rho$, and the set

$$H_\alpha(\rho) = \{(x_1, x_2): x_1 + x_2 = \rho,\ V_\alpha{}^*(x_1, x_2) = h_\alpha(\rho)\} \tag{8.8}$$

is nonempty. Finally, let

$$H_\alpha = \bigcup_{\rho \in [0, N_1+N_2]} H_\alpha(\rho) \tag{8.9}$$

For an illustration of $H_\alpha$, $U_\alpha$ and $L_\alpha$ (the latter are defined by (8.15), (8.16)), see Fig. 2. (We should point out that the points on the $x_1$ axis to the left of point $C$ – point $B$ in particular – belong to $H_\alpha$.)

Lemma 8.1: a) $H_\alpha(\rho)$ is connected for any $\rho$, $\underline{\alpha}$.

b) If $V_\alpha{}^*$ is strictly convex, then $H_\alpha(\rho)$ is a singleton.

Proof: Straightforward consequences of convexity. ∎

Theorem 8.3: a) $h_\alpha$ is Lipschitz continuous.

b) $H_\alpha$ is closed.

c) $H_\alpha$ is connected.

Proof: a) Given some $\rho$, let $(x_1, x_2) \in H_\alpha(\rho)$. Then, for some $F > 0$ and for any real $\Delta$,

$$h_\alpha(\rho - \Delta) \leq V_\alpha{}^*(x_1 - \Delta/2, x_2 - \Delta/2) \leq V_\alpha{}^*(x_1, x_2) + F|\Delta| = h_\alpha(\rho) + F|\Delta| \tag{8.10}$$

(The first inequality follows from the definition of $h_\alpha$; the second, from the Lipschitz continuity of $V_\alpha{}^*$). Similarly,

$$h_\alpha(\rho) \leq h_\alpha(\rho - \Delta) + F|\Delta| \tag{8.11}$$

Inequalities (8.10) and (8.11) prove part (a).

b) Let $\{(x_1(n), x_2(n))\}$ be a sequence of elements of $H_\alpha$, and let $(x_1, x_2)$ be a limit point of the sequence. Let $\rho(n) \equiv x_1(n) + x_2(n)$, $\rho = x_1 + x_2$. Then, by continuity of $h_\alpha$ and $V_\alpha{}^*$,

$$h_\alpha(\rho) = \lim_{n\to\infty} h_\alpha(\rho(n)) = \lim_{n\to\infty} V_\alpha{}^*(x_1(n), x_2(n)) = V_\alpha{}^*(x_1, x_2) \tag{8.12}$$

This shows that $(x_1, x_2) \in H_\alpha(\rho)$ and consequently, $(x_1, x_2) \in H_\alpha$. Therefore, $H_\alpha$ is a closed set.

c) Suppose that $H_\alpha$ is not connected. Then (Ref. 25), there exists a pair of disjoint, nonempty sets $A$, $B$ whose union is $H_\alpha$, neither of which contains a limit point of the other. From Lemma 8.1, $H_\alpha(\rho)$ is connected, for any $\rho$. Therefore, for any $\rho$, we either have $H_\alpha(\rho) \subset A$ or $H_\alpha(\rho) \subset B$. Let

$$C = \{\rho \in [N_1 + N_2]: H_\alpha(\rho) \subset A\} \tag{8.12}$$

$$D = \{\rho \in [N_1 + N_2]: H_\alpha(\rho) \subset B\} \tag{8.13}$$

Since $A \neq \emptyset$, $B \neq \emptyset$, we have $C \neq \emptyset$, $D \neq \emptyset$. Since $[0, N_1 + N_2]$ is connected, one of $C$, $D$ contains a limit point of the other. Assume, without loss of generality, that $\{\rho_n\}$ is a sequence of elements of $C$ that converges to some $\rho \in D$. Let $(x_1(n), x_2(n)) \in H_\alpha(\rho_n)$. In particular, $(x_1(n), x_2(n)) \in A$. Then, the sequence $\{x_1(n), x_2(n)\}$ has a subsequence that converges to a limit point $(x_1, x_2)$ and $x_1 + x_2 = \rho$. By the continuity of $V_\alpha{}^*$ and $h_\alpha$,

$$V_\alpha{}^*(x_1, x_2) = \lim_{n\to\infty} V_\alpha{}^*(x_1(n), x_2(n)) = \lim_{n\to\infty} h_\alpha(\rho_n) = h_\alpha(\rho) \tag{8.14}$$

which shows that $(x_1, x_2) \in H_\alpha(\rho)$. Since $\rho \in D$, definition (8.13) implies that $(x_1, x_2) \in B$. So, $B$ contains a limit point of $A$ and the contradiction shows that $H_\alpha$ is connected. ∎

Now let

$$U_\alpha(\rho) = \{(x_1, x_2): x_1 + x_2 = \rho,\ x_1 - x_2 < y_1 - y_2,\ \forall (y_1, y_2) \in H_\alpha(\rho)\} \tag{8.15}$$

$$L_\alpha(\rho) = \{(x_1, x_2): x_1 + x_2 = \rho,\ x_1 - x_2 > y_1 - y_2,\ \forall (y_1, y_2) \in H_\alpha(\rho)\} \tag{8.16}$$

and (see Figure 2)

$$U_\alpha = \bigcup_{\rho \in [0, N_1+N_2]} U_\alpha(\rho), \qquad L_\alpha = \bigcup_{\rho \in [0, N_1+N_2]} L_\alpha(\rho). \tag{8.17}$$

Since $H_\alpha(\rho)$ is connected, it follows that

$$U_\alpha(\rho) \bigcup L_\alpha(\rho) \bigcup H_\alpha(\rho) = \{(x_1, x_2): x_1 + x_2 = \rho\} \tag{8.18}$$

and consequently

$$U_\alpha \bigcup L_\alpha \bigcup H_\alpha = [0, N_1] \times [0, N_2] \tag{8.19}$$

Finally, note that (keeping $(x_1, x_2) \in U_\alpha$ fixed) the function $V_\alpha{}^*(x_1+\Delta, x_2-\Delta)$ is a strictly decreasing function of $\Delta$ (for small enough $\Delta$), because of the convexity of $V_\alpha{}^*$ and the definition of $U_\alpha$. With this remark, we have the following characterization of the optimal values of $\lambda_1^u$, $\lambda_2^u$ in the interior of the state space:

<u>Theorem 8.4</u>: If $V_\alpha{}^*$ is decreasing, $u \in U_M \cap U_R$ is everywhere optimal and $\underline{x}$ is in the interior of $[0, N_1] \times [0, N_2]$, then

a) If $\underline{x} \in U_\alpha$, then $\lambda_1^u(\underline{x}, \underline{\alpha}) = \lambda^* \alpha_o$.

b) If $\underline{x} \in L_\alpha$, then $\lambda_2^u(\underline{x}, \underline{\alpha}) = \lambda^* \alpha_o$.

<u>Proof:</u> Let $\underline{x}$ belong to the interior of $U_\alpha$. We must again minimize the expression (8.6). Because of the monotonicity property mentioned in the last remark, it follows that $\lambda_1^u$ has to be set equal to its maximum value $\underline{\alpha}_o \lambda^*$. Part (b) follows from a symmetrical argument. ∎

We now discuss the optimality conditions on the separating set $H_\alpha$. We assume that $V_\alpha{}^*$ is strictly convex and (by Lemma 8.1) $H_\alpha(\rho)$ is a singleton, for any fixed $\rho$. Equivalently, $H_\alpha$ is a continuous curve. According to the remarks following theorem 5.1, $V_\alpha{}^*$ is always strictly convex, but since we haven't given a proof of this fact, we introduce it as an assumption.

Fix $(x_1, x_2) \in H_\alpha$ and suppose that $0 < x_i < N_i$, $i = 1, 2$, (interior point). Given a control law $u$, let

$$A(u) = \{\tau > 0: \underline{x}_\alpha^u(\tau) \in U_\alpha\} \tag{8.20}$$

$$B(u) = \{\tau > 0: \underline{x}_\alpha^u(\tau) \in L_\alpha\} \tag{8.21}$$

where $\underline{x}_{\alpha}^{u}(\tau)$ is the path followed starting from $((x_1, x_2), \underline{\alpha})$ if no jump of $\underline{\alpha}$ occurs. We distinguish four cases:

a)Suppose that for all $u \in U_M \cap U_R$, time $t = 0$ is a limit point of $A(u)$. For all $\tau \in A(u)$, we have $\lambda_1^u(\tau) = \lambda^*$ (by Theorem 8.4). Then, by right continuity of $\lambda_1^u$ (assumption S2, section 2), we must have $\lambda_1^u(0) = \lambda^*$.

b) Similarly, if for all $u \in U_M \cap U_R$, $t = 0$ is a limit point of $B(u)$, we must have $\lambda_2^u(0) = \lambda^*$.

c) If $t = 0$ is a limit point of both $A(u)$ and $B(u)$, for all $u \in U_M \cap U_R$, then no everywhere optimal control law exists. Fortunately, this will never be the case if $H_\alpha$ is a sufficiently smooth curve.

d) Finally suppose that there exists some $u$ such that $t = 0$ is not a limit point of either $A(u)$ or $B(u)$. In that case $\underline{x}_{\alpha}^{u}(t) \in H_\alpha$, $\forall t \in [0, \Delta]$ for some small enough $\Delta > 0$. An argument similar to that in theorem 8.4 will show that this control law satisfies the optimality conditions at $(x_1, x_2)$. Such a control law travels on $H_\alpha$, i.e. stays on the deepest part of the valley-like convex function $V_\alpha$*.

The optimality conditions on the boundaries are slightly more complicated because the constraints on $\lambda_i$ and $\mu_i$ are interrelated through the requirement that $x_i$ stays in $[0, N_i]$. The exact form of these conditions depends , in general, on the relative magnitudes of the parameters $\lambda^*$, $\mu_1$* and $\mu_2$*. However, for any particular problem, Theorem 7.2 leads to an unambiguous selection of the values of the control variables.

## 9. CONCLUSIONS – GENERALIZATIONS.

Let us start by pointing out the main properties of our queueing system on which our development has been based: 1) We first have the existence of a special state which is recurrent when we restrict to a class of control laws that have equally good performance as the original set of admissible control laws. 2) We have the convexity of the optimal cost–to–go function which only depends on the following facts: a) The state space is convex, b) the set of admissible values of the control variables is convex and c) the cost function is convex. Our methodology is therefore applicable, with minor adjustments, to the large class of linear dynamical systems in which the above enumerated properties are present.

We now indicate a few alternative configurations for which all steps of our development would remain valid. We may let the buffer capacities be infinite. Then, provided that storage costs increase fast enough with $x_i$, it is still possible to obtain a recurrence result. The convexity theorem would be still valid. A few derivations would need some more care, because $V^*$ and $f$ will no more be bounded functions of the state space but the main results of section 8 would remain unchanged.

We may also have three (instead of two) downstream buffers and machines, in which case the state space is three–dimensional. Convexity of $V^*$ and the optimality conditions then imply that, for any fixed $\underline{\alpha}$, the three dimensional state space is divided into three regions, separated by three two-dimensional surfaces that intersect on a one–dimensional curve. In each of the three regions, all material is to be routed to a unique buffer. The switching surfaces have interpretations similar to the switching curves $H_\alpha$ of section 8.

As pointed out earlier, our recurrence results (Theorem 3.2) have been based on the assumption that the lead machine is unreliable ($p_o \neq 0$). While this is a convenient assumption, it is not a necessary one, except that, if $p_o = 0$, the reference state $s_o$ should be differently chosen. This choice should be problem specific and would not present any difficulties for most interesting queueing systems. The only difference that arises when $p_o = 0$ is that $V^*$ need not be strictly convex and the separating set $H_\alpha$ could even be the entire state space (Ref. 26, Ch.6).

As another variation of our problem, we could include a nonlinear, convex and increasing cost on the utilization rates of the machines, to penalize utilization at or near capacity limits. The rationale behind this cost criterion is that high utilization rates are generally undesirable (in the long run). In that case $V^*$ would still be convex but Theorems 8.1, 8.2, would no longer hold. For example, the optimal utilization rates $\mu_i^u$ of the downstream machines wouldn't be equal to $\mu_i^*$ but rather an increasing function of the buffer levels.

The next issue of concern is the computation of $V^*$ and the generation of an optimal control law. One conceivable procedure (resembling the Howard algorithm) is to evaluate $V^u$, for a fixed Markovian $u$, by solving the equation $\mathcal{L}^u V^u + k^u = g^u$ for $V^u$ and $g^u$. This equation has a unique solution within an additive constant for $V^u$. It really consists of eight coupled first order linear partial differential equations with non-constant coefficients and can only be solved numerically. Based on $V^u$ we may generate a control law $w$ which improves performance by minimizing $\mathcal{L}^w V^u + k^w$ and so on. In practice, any such algorithm would involve a discretization procedure, so it might be preferable to formulate the problem on a discrete state space. In that case, the successive approximation algorithm (or accelerated versions of it) would yield a solution relatively efficiently.

An alternative iterative optimizing algorithm, based on an equivalent deterministic optimal control problem has been also suggested in Ref. 27 (see also Refs. 17, 26 for related ideas).

The drawback of any numerical procedure is that the computational requirements become immense, even for moderate sizes of the state space (e.g. $N_1 = N_2 = 20$, see Ref. 15). Fortunately, the existing numerical evidence shows that the performance functional is not very sensitive to variations of the dividing curve, so that rough approximations may be particularly useful. Estimates of the asymptotic slope of $H_a$ as $N_1$ and $N_2$ increase, as well as of the intercepts of $H_a$ with the axes $x_i = 0$ would be very helpful for obtaining an acceptable suboptimal control law.

## REFERENCES

1. KLEINROCK, L., *Queueing Systems, Vol. I*, J. Wiley, New York, 1975.

2. KLEINROCK, L., *Queueing Systems, Vol. II*, J. Wiley, New York, 1976.

3. GALLAGER, R.G., *A Minimum Delay Routing Algorithm Using Distributed Computation,* IEEE Transactions on Communications, Vol. COM-25, No.1, pp. 73–85, 1977.

4. KIMEMIA, J.G. and GERSHWIN, S.B., *Multicommodity Network Flow Optimization in Flexible Manufacturing Systems,* Report ESL-FR-834-2, Electronic Systems Laboratory, MIT, 1980.

5. CRABILL, T., GROSS, D. and MAGAZINE, M.J., *A Classified Bibliography of Research on Optimal Design and Control of Queues,* Operations Research, Vol. 25, No. 2, pp. 219-232, 1977.

6. SOBEL, M.J., *Optimal Operation of Queues,* Mathematical Methods in Queueing Theory, Proceedings of a Conference on Mathematical Methods in Graph Theory, Western Michigan University, Springer Verlag, New York, 1974.

7. CRABILL, T., *Optimal Control of Service Facility with Variable Exponential Service Time and Constant Arrival Rate,* Management Science, Vol. 18, No. 9, pp. 560-566, 1977.

8. GALLISH, E., *On Monotonic Optimal Policies in a Queueing Model of M/G/1 Type with Controllable Service Time Distribution,* Advances in Applied Probability, Vol. 11, pp. 870-887, 1979.

9. ROSBERG, Z., VARAIYA, P. and WARLAND, J., *Optimal Control of Service in Tandem Queues,* Mem. No. UCB/ERL M80/42, Electronics Research Laboratory, University of California, Berkeley, 1980.

10. SCHWARTZ, B., *Queueing Models with Lane Selection: A New Class of Problems,* Operations Research, Vol. 22, pp. 331-339, 1974.

11. FOSCHINI, G.J., *On Heavy Traffic Diffusion Analysis and Dynamic Routing in Packet Switched Networks,* Computer Performance, K.M. Chandy and M. Reiser (eds.), North Holland, 1977.

12. FOSCHINI, G.J. and SALZ, J., *A Basic Dynamic Routing Problem and Diffusion,* IEEE Transactions on Communications, Vol. COM-26, No. 3, pp. 320–327, 1978.

13. EPHREMIDES, A., VARAIYA, P. and WARLAND, J., *A Simple Dynamic Routing Problem* , IEEE Transactions on Automatic Control, Vol. AC-25, No. 4., pp. 690—693, 1980.

14. DEUERMEYER, B.L. and PIERSKALLA, W.P., *A by-Product Production System with an Alternative,* Management Science, Vol. 24, No. 13, pp.1373–1383, 1978.

15. HAHNE, E., *Dynamic Routing in an Unreliable Manufacturing Network with Limited Storage,* Report LIDS-TH-1063, Laboratory for Information and Decision Systems, MIT, 1981.

16. OLSDER, G.J. and SURI, R., *Time-Optimal Control of Parts Routing in a Manufacturing Sustem with Failure Prone Machines,* Proceedings of the 19th IEEE Conference on Decision and Control, 1980.

17. KIMEMIA, J.G. and GERSHWIN, S.B., *An Algorithm for the Computer Control of Production in a Flexible Manufacturing System,* Proceedings of the 20th IEEE Conference on Decision and Control, 1981.

18. KOENIGSBERG, E., *Production Lines and Internal Storage - A Review,* Management Science, Vol. 5, pp. 410-433, 1959.

19. SARMA, V.V.S. and ALAM, M., *Optimal Maintenance Policies for Machines Subject to Deterioration and Intermittent Breakdowns,* IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-4, pp. 396-398, 1975.

20. GERSHWIN, S.B. and SCHICK, I.C., *Modelling and Analysis of Two- and Three-Stage Transfer Lines with Unreliable Machines and Finite Buffers,* Report LIDS-R-979, Laboratory for Information and Decision Systems, MIT, 1980.

21. BERTSEKAS, D.P., *Dynamic Programming and Stochastic Control* , Academic Press, New York, 1976.

22. BOEL, R. and VARAIYA, P., *Optimal Control of Jump Processes,* SIAM J. Control and Optimization, Vol. 15, No. 1, pp. 92–119, 1977.

23. LOEVE, M., *Probability Theory*, Springer Verlag, New York, 1977.

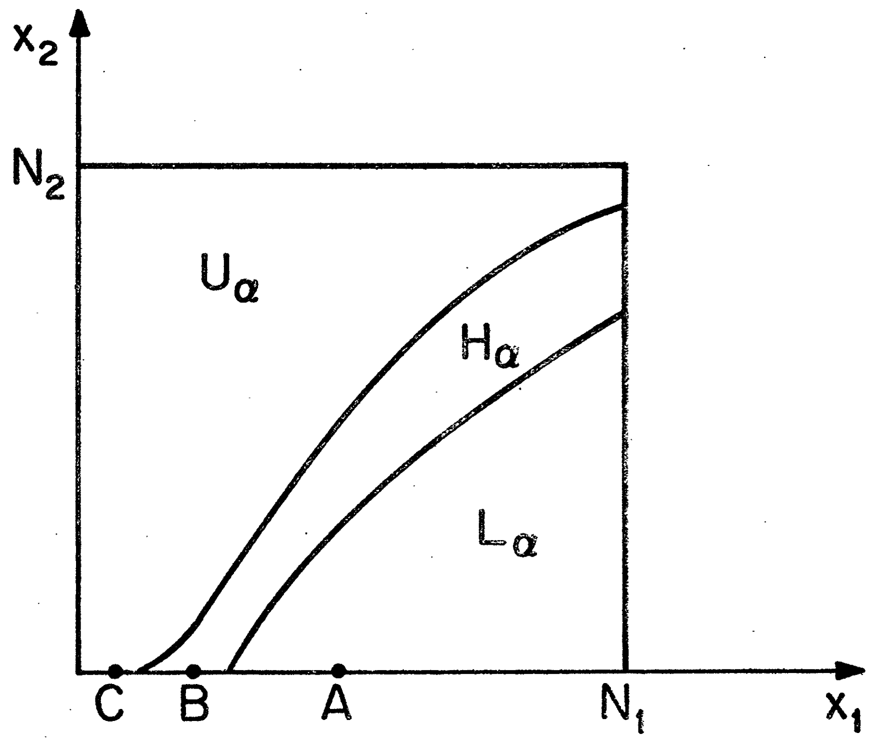
24. KUSHNER, H.J., *Optimality Conditions for the Average Cost per Unit Time Problem with a Diffusion Model,* SIAM J. on Control and Optimization, Vol. 16, No. 2., pp. 330–346, 1978.

25. MUNKRES, J.R., *Topology,* Prentice Hall, Englewood Cliffs, N.J., 1975.

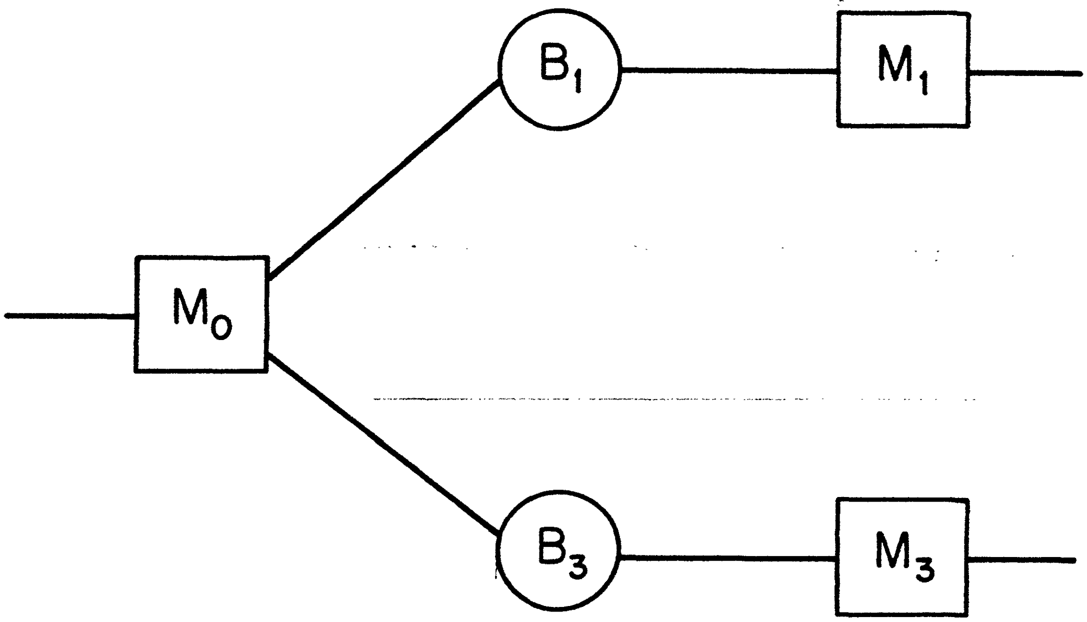
26. TSITSIKLIS, J.N., *Optimal Dynamic Routing in an Unreliable Manufacturing System,* Report LIDS-TH-1069, Laboratory for Information and Decision Systems, MIT, 1981.

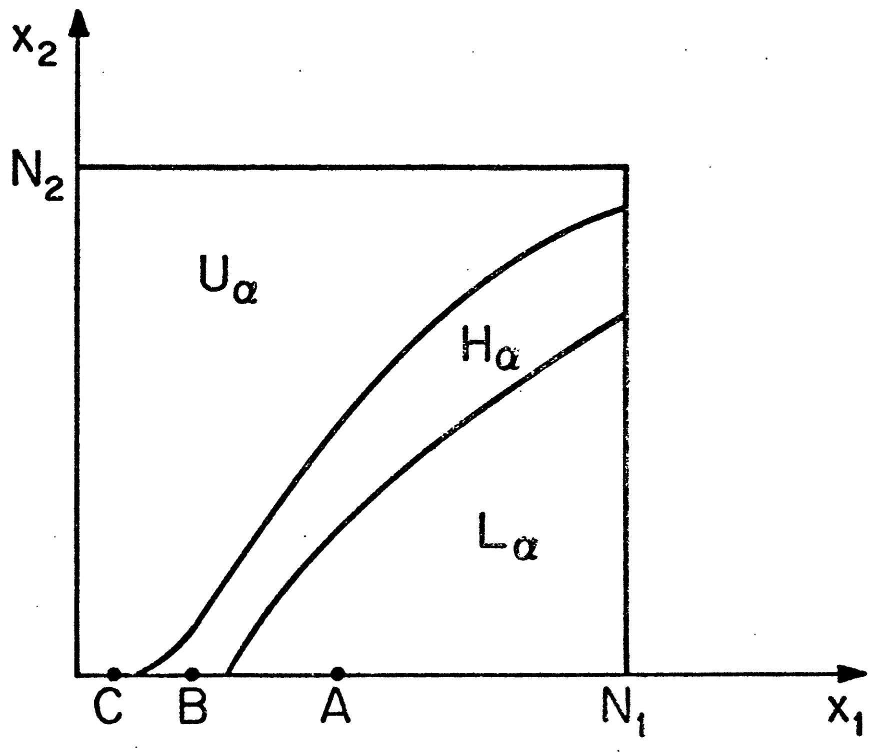
27. RISHEL, R., *Dynamic Programming and Minimum Principles for Systems with Jump Markov Disturbances,* SIAM J. Control, Vol. 13, No. 2, pp. 338-371, 1975.

# LIST OF CAPTIONS



x2
N2
Uα
Hα
Lα
C
B
A
N1
x1


B1
M1
M0
B3
M3



x2
N2
Uα
Hα
Lα
C
B
A
N1
x1