SIAM J. CONTROL AND OPTIMIZATION
Vol. 30, No. 6, pp. 1–23, November 1992
© 1992 Society for Industrial and Applied Mathematics
001

OPTIMAL CONTROL OF SWITCHING DIFFUSIONS WITH APPLICATION TO FLEXIBLE MANUFACTURING SYSTEMS*

MRINAL K. GHOSH†, ARISTOTLE ARAPOSTATHIS‡, AND STEVEN I. MARCUS§

Abstract. A controlled switching diffusion model is developed to study the hierarchical control of flexible manufacturing systems. The existence of a homogeneous Markov non-randomized optimal policy is established by a convex analytic method. Using the existence of such a policy, the existence of a unique solution in a certain class to the associated Hamilton-Jacobi-Bellman equations is established and the optimal policy is characterized as a minimizing selector of an appropriate Hamiltonian.

Key words. flexible manufacturing system, Wiener process, switching diffusion, Poisson measure, Markov policy, dynamic programming equations

AMS(MOS) subject classifications. 93E20

1. Introduction.

We study a controlled switching diffusion process that arises in numerous applications of systems with multiple modes or failure modes, including the hierarchical control of flexible manufacturing systems. A flexible manufacturing system (FMS) consists of a set of workstations capable of performing a number of different operations and interconnected by a transportation mechanism. An FMS produces a family of parts related by similar operational requirements or by belonging to the same final assembly [27]. The rapidly growing range of applicability of FMS includes metal cutting, assembly of printed circuit boards, integrated circuit fabrication, automobile assembly lines, etc. Due to their tremendous flexibility, FMS are significantly more efficient in many ways than traditional manufacturing systems. However, the high capital cost of an FMS demands very efficient management of production and maintenance (repair/replacement) scheduling so that uncertain events such as random demand fluctuations, machine failures, inventory spoilage, sales returns, etc. can be taken care of. The large size of the system and its associated complexities make it imperative to divide the control or management into a hierarchy consisting of a number of levels. Thus, the overall complex problem is reduced to a number of manageable subproblems at each level, and these levels are linked by means of a hierarchical integrative system. We refer to [1], [21], [27] for a detailed description of these hierarchical schemes. We will confine our attention to the top two levels, viz.

(i) Generation of decision tables, which is accomplished by developing a suitable mathematical model describing the dynamical evolution of the system. This is done off-line.
(ii) The flow control level: This plays the central role in the system. It determines, on line, the production and maintenance scheduling and continuously feeds the routing control level which calculates route splits, and which in

*Received by the editors September 19, 1990; accepted for publication (in revised form) March 12, 1992. This research was supported in part by the Texas Advanced Research Program (Advanced Technology Program) under Grants No. 003658-093 and No. 003658-186, in part by the Air Force Office of Scientific Research under Grants AFOSR-91-0033, F49620-92-J-0045 and F49620-92-J-0083, and in part by the National Science Foundation under Grant CDR-8803012.

†Department of Mathematics, Indian Institute of Science, Bangalore 560012.

‡Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, Texas 78712.

§Electrical Engineering Department and Systems Research Center, University of Maryland, College Park, Maryland 20742.

turn governs the sequence controller which determines the scheduling times at which to dispatch parts.

Since the top two levels directly govern the rest, it is of paramount importance to develop and study an appropriate mathematical model which will facilitate to find on line implementable optimal feedback policies.

We first present a heuristic description of our model, which is a modified version of the model in [1], [21], [27]. The FMS consists of L workstations, with each workstation having a number L_m of identical machines ($m = 1, 2, \ldots, L$). A family of N types of different parts is produced. Let $u(t) = [u_1(t), \ldots, u_N(t)]^T \in \mathbb{R}^N$ and $d(t) = [d_1(t), \ldots, d_N(t)]^T \in \mathbb{R}^N$ denote the production rate (a control variable) and the downstream demand rate vectors of this family of parts, respectively. Also, $X(t) = [X_1(t), \ldots, X_N(t)]^T \in \mathbb{R}^N$ denotes the downstream buffer stock. A negative value of $X_j(t)$, $j = 1, \ldots, N$, indicates a backlogged demand for part j, while a positive value is the size of the inventory stored in the buffers. The evolution of $X(t)$ is governed by the following stochastic differential equations

$$\frac{dX(t)}{dt} = u(t) - d(t) + \operatorname{diag}\bigl(\sigma_1, \ldots, \sigma_N\bigr)\xi(t), \tag{1.1}$$

where $\sigma_i > 0$, $i = 1, \ldots, N$ and $\xi(t) = [\xi_1(t), \ldots, \xi_N(t)]^T$ is an N-dimensional white noise which can be interpreted as "sales returns", "inventory spoilage", "sudden demand fluctuations," etc. (see [8]).

If $S_m(t)$ denotes the number of operational machines in station m at time t, then the state of the workstations may be represented by the L-tuple

$$S(t) = \bigl(S_1(t), \ldots, S_L(t)\bigr)\,.$$

The evolution of $S(t)$ is influenced by the inventory size and production scheduling, and can also be controlled by various decisions such as produce, repair, replace, etc. The dynamics of $S(t)$ can be described as follows:

$$\begin{aligned} &P\bigl\{S_m(t+\delta t) = \ell + 1 \bigm| S_m(t) = \ell\bigr\} \\ &\qquad = \begin{cases} (L_m - \ell) v_m(t)\, \delta t + o(\delta t) & \text{for } 0 \le \ell < L_m \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{1.2}$$

where $v_m(t)$, $m = 1, \ldots, L$, are suitable control variables. In the uncontrolled case, $v_m(t) = \gamma_m$, which represents the infinitesimal repair rate at station m. These repair rates may implicitly depend on $X(t)$. This model also allows for a control variable reflecting the decision as to whether to repair or replace on the basis of the inventory size. Also,

$$\begin{aligned} &P\bigl\{S_m(t+\delta t) = \ell - 1 \bigm| S_m(t) = \ell\bigr\} \\ &\qquad = \begin{cases} \ell\, p_m\bigl(X(t), u(t)\bigr)\delta t + o(\delta t) & \text{for } 0 \le \ell < L_m \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{1.3}$$

where p_m models the infinitesimal failure rate at the m^{th} station. Equations (1.2) and (1.3) imply that

$$P\bigl\{S_m(t+\delta t) = \ell_1 \bigm| S_m(t) = \ell_2\bigr\} = 0, \qquad \text{for } \ |\ell_1 - \ell_2| > 1\,.$$

With i and j denoting two states of the system, we define

$$\lambda_{ij}(\cdot)\delta t + o(\delta t) = P\big\{S(t+\delta t) = j \;\big|\; S(t) = i\big\}, \qquad i \neq j$$

and

$$\lambda_{ii}(\cdot) = -\sum_{j\neq i} \lambda_{ij}(\cdot)\,.$$

The machine state $S(t)$ can thus be modeled as a continuous time controlled jump process taking values in a finite state space. In the uncontrolled case, $S(t)$ becomes a continuous time homogeneous Markov chain with infinitesimal generator given by the matrix $\big[\lambda_{ij}\big]$.

The choice of the production rate at each instant is constrained by the capacity of the currently operational machines. This translates into the requirement that at each time t the production rates must lie in some set $\Gamma\big(S(t)\big)$ which depends on the machine state.

Let $y^k_{mn}(t)$ be the number of type n parts which undergo operation k at the m^{th} station per unit interval of time and $\tau^k_{mn}(t)$ the length of time required for the completion of this operation. The product $y^k_{mn}(t)\,\tau^k_{mn}(t)$ is the portion of each unit time interval that one or more operational machines at station m must dedicate to perform operation k on type n parts, as dictated by the flow rate $y^k_{mn}(t)$. Since the amount of work completed at each station per unit time interval cannot exceed the time available at the operational machines, the following constraint applies

$$\sum_n \sum_k y^k_{mn}(t)\tau^k_{mn}(t) \leq S_m(t)\,, \qquad \text{for all } m. \tag{1.4}$$

Also, assuming that no material is allowed to accumulate within the system, the throughput $u_n(t)$ of type n parts must satisfy

$$u_n(t) = \sum_m y^k_{mn}(t)\,, \qquad \text{for all } k \text{ and } n. \tag{1.5}$$

Therefore, for each state i, the set $\Gamma(i)$ is defined as the collection of all production rates $u = [u_1, \dots, u_N]^T$ for which, with the machine state $S(t) = i$, there exist feasible flow rates $y^k_{mn}(t)$ satisfying (1.4) and (1.5).

The flow control problem can now be stated. Given an initial buffer state $X(0) = x$ and machine state $S(0) = i$, we wish to specify a production plan and maintenance (repair/replacement) policy that minimizes the performance index

$$J(x,i,u,v) = E\left[\int_0^\infty e^{-\alpha t} c\big(X(t), S(t), u(t), v(t)\big)\,dt \;\middle|\; X(0) = x, S(0) = i\right], \tag{1.6}$$

where $c(\cdot)$ is a 'cost' function, $\alpha > 0$ is a discount factor, $u(\cdot)$ is the production rate, and $v(\cdot)$ is the maintenance rate. The objective is to find $u(\cdot)$, $v(\cdot)$ for which the minimum is achieved in (1.6). The ideal production and maintenance policy for a wide class of cost functions would minimize J by producing parts at exactly the demand rate, thereby keeping the buffer at zero. Such a policy is generally impossible because of the failures of the machines and various other uncertainties.

This FMS model motivates the study of a stochastic optimization problem in a more abstract setting which subsumes the flow control problem in the FMS as a special case. This abstract problem is manifested in numerous other situations. In [17] it is encountered in a hybrid model proposed for the study of dynamic phenomena in large scale interconnected power networks. Sworder [39], [40] describes possible applications to macroeconomic models and dynamic renewal problems in general. In addition, it should be useful at other levels of the hierarchy described in [21].

We will briefly describe this problem formally; a rigorous description will be given in Section 2. Let $\mathcal{S} = \{1, 2, \ldots, M\}$ and let U_i, $i = 1, \ldots, M$, be prescribed compact metric spaces. For each $i, j \in \mathcal{S}$, let $b(\cdot, \cdot, i, \cdot) : \mathbb{R}_+ \times \mathbb{R}^N \times U_i \to \mathbb{R}^N$ and $\lambda_{ij} : \mathbb{R}_+ \times \mathbb{R}^N \times U_i \to \mathbb{R}$, satisfying $\lambda_{ij} \geq 0$, for $i \neq j$ and $\sum_j \lambda_{ij}(\cdot) = 0$. A stochastic process $\bigl(X(t), S(t)\bigr)$ taking values in $\mathbb{R}^N \times \mathcal{S}$ is given by

$$X(t) = X(0) + \int_0^t b\bigl(\tau, X(\tau), S(\tau), u(\tau)\bigr)\, d\tau + \operatorname{diag}(\sigma_1, \ldots, \sigma_N) W(t) \tag{1.7}$$

$$P\bigl\{S(t + \delta t) = j \bigm| S(t) = i, X(s), S(s), s \leq t\bigr\} = \lambda_{ij}\bigl(t, X(t), u(t)\bigr)\delta t + o(\delta t), \tag{1.8}$$

where $\sigma_i > 0$, $i = 1, \ldots, N$, are constants and $W(\cdot) = [W_1(\cdot), \ldots, W_N(\cdot)]^T$ is an N-dimensional standard Brownian motion. The control $u(\cdot)$ is a $U := \prod_{i=1}^N U_i$-valued process such that when $S(t) = i$, $u(\cdot)$ takes values in U_i and $u(\cdot)$ is non-anticipative with respect to the driving Brownian motion $W(t)$. Let $c : \mathbb{R}_+ \times \mathbb{R}^N \times \mathcal{S} \times U \to \mathbb{R}_+$ be the cost function and $\alpha > 0$ a prescribed discount factor. Define a cost functional of the form

$$E\biggl[\int_0^\infty e^{-\alpha t} c\bigl(t, X(t), S(t), u(t)\bigr)\, dt\biggr]. \tag{1.9}$$

The objective is to find an optimal control policy $u(\cdot)$ which minimizes (1.9) and takes the feedback form $u(t) = \overline{v}\bigl(t, X(t), S(t)\bigr)$, for a suitably defined function $\overline{v}$. In the next Section, we will assume appropriate conditions on b and λ which will guarantee that (1.7), (1.8) are well defined. We note here that for a performance index of the form (1.9), m, λ, c may be assumed to be independent of t without any loss of generality. Also, by replacing each U_k by $\prod_{k=1}^M U_k$ and $b(\cdot, i, \cdot)$ by its composition with the projection $\prod_{k=1}^M U_k \to U_i$, we may assume that each U_i is a replica of a fixed compact metric space.

We now briefly mention some earlier work leading to ours. The class of controlled piecewise deterministic models with jump Markov disturbances have been studied by Sworder [38], Rishel [35], Olsder and Suri [33], Davis [19] and Vermes [42] among many others. The piecewise deterministic FMS model has been studied by Kimenia and Gershwin [27], who have developed a heuristic numerical method based on the maximum principle established in [35]. Akella and Kumar [2] have studied a simplified model and obtained explicit solutions for one machine producing a single commodity. In all these papers the jump process is modeled as a continuous time (uncontrolled) Markov chain. Boukas and Haurie [14], [15] have modified the FMS model of Kimenia and Gershwin by introducing new state variables describing machine wear as well as a control parameter in the jump process; their model incorporates preventive maintenance. They have obtained a maximum principle, thereby extending Rishel's formalism in [35]. They have also considered piecewise deterministic models. To obtain an optimal policy of the feedback type in these models one has to impose very strong conditions on terms like b, λ governing the system and stringent restrictions on the set of allowable policies. At the same time, it is assumed in these models that between any two successive jumps of $S(t)$, the dynamics governing $X(t)$ is deterministic. Thus, certain unavoidable 'environmental' uncertainties are not taken into account. These factors restrict the scope of applicability of these models. We have tried to circumvent these difficulties by adding an additive noise term in the state dynamics. This is specifically done in order to take into account the various sources of environmental randomness. Addition of this noise removes practically all restrictions imposed on the set of allowable control policies, thereby substantially enhancing the range of its applicability. The switching diffusion problem has also been studied by Bensoussan and Lions [7], using a martingale problem formulation. However, our motivation and approach are quite different. In [7], it is assumed that for some $\delta > 0$, $-\lambda_{ii} > \delta > 0$, for each i. We have, instead, used a strong formulation which is very important for practical applications. In our formulation we do not need the condition $-\lambda_{ii} > \delta > 0$. We also refer to [6], [8], [9], [16], [20], [32], [36], [37], [43] for related work.

Our paper is structured as follows. A rigorous description of the mathematical model of the FMS is given in Section 2. The optimization problem is formulated and subsequently reduced to an equivalent convex optimization problem, via the study of associated occupation measures. The compactness of laws is established in Section 3, the convexity and extremality of occupation measures are studied in Section 4, and the proof of existence of optimal policies is given in Section 5. Section 6 deals with the characterization of optimal policies via dynamic programming equations. In Section 7, we apply our theory to a simplified model and derive some interesting results. Finally, Section 8 contains some concluding remarks. Note that we have used a convex analytic approach for this problem, as opposed to the traditional analytic one. For the discounted cost criterion, the latter is more economical and is sketched in the Appendix. However, the convex analytic approach is interesting in its own right and would be more flexible and powerful for certain other purposes, e.g., the pathwise average cost problem or problems with several constraints where the analytic approach does not seem to be amenable. For (nonswitching) controlled diffusions, these problems have been treated in [11, Chap. VI] and [12] by a convex analytic approach. We hope our approach to switching diffusions would be useful in various other situations.

2. Mathematical Model and Preliminaries. Let U be a compact metric space and $\mathcal{S} = \{1, \dots, M\}$. Let $\bar{b} = [\bar{b}_1, \dots, \bar{b}_N]^T : \mathbb{R}^N \times \mathcal{S} \times U \to \mathbb{R}^N$. For each $i \in \mathcal{S}$, $\bar{b}(\cdot, i, \cdot)$ is assumed to be bounded, continuous and Lipschitz in its first argument uniformly with respect to the third. For $i, j \in \mathcal{S}$, let $\bar{\lambda}_{ij} : \mathbb{R}^N \times U \to \mathbb{R}$ be bounded, continuous and Lipschitz in its first argument uniformly with respect to the second. Also, assume that for $i, j \in \mathcal{S}$, $i \neq j$, $\bar{\lambda}_{ij} \geq 0$, and $\sum_{j=1}^{M} \bar{\lambda}_{ij} = 0$, for any $i \in \mathcal{S}$. Let $\sigma_i > 0$, $i = 1, 2, \dots, N$, be prescribed numbers. For a Polish space Y, $\mathfrak{B}(Y)$ will denote its Borel σ-field and $\boldsymbol{\mathcal{P}}(Y)$ the space of probability measures on $\mathfrak{B}(Y)$ endowed with the Prohorov topology, i.e., the topology of weak convergence [10]. Let $\mathfrak{M}(Y)$ be the set of all non-negative integer-valued, σ-finite measures on $\mathfrak{B}(Y)$. Let $\mathfrak{M}_\sigma(Y)$ be the smallest σ-field on $\mathfrak{M}(Y)$ with respect to which all maps from $\mathfrak{M}(Y)$ into $\mathbb{N} \cup \{\infty\}$ of the form $\mu \longmapsto \mu(B)$, with $B \in \mathfrak{B}(Y)$, are measurable. $\mathfrak{M}(Y)$ will always be assumed to be endowed with this measurability structure. Let $\mathcal{V} = \boldsymbol{\mathcal{P}}(U)$ and $b = [b_1, \dots, b_N]^T : \mathbb{R}^N \times \mathcal{S} \times \mathcal{V} \to \mathbb{R}^N$ be defined by

$$b_i(\cdot, \cdot, v) := \int_U \bar{b}_i(\cdot, \cdot, u)\, v(du)\,, \qquad v \in \mathcal{V}, \quad i = 1, \dots, N\,. \tag{2.1}$$

Similarly, for $i,j \in \mathcal{S}$, $\lambda_{ij} : \mathbb{R}^N \times \mathcal{V} \to \mathbb{R}$ is defined as

$$\lambda_{ij}(\cdot, v) := \int_U \overline{\lambda}_{ij}(\cdot, u)\, v(du)\,, \qquad v \in \mathcal{V}, \quad i,j \in \mathcal{S}. \tag{2.2}$$

For $i,j \in \mathcal{S}$, $x \in \mathbb{R}^N$ and $v \in \mathcal{V}$, we construct the intervals $\Delta_{ij}(x,v)$ of the real line in the following manner (see also [13], [17]):

$$\begin{aligned}
\Delta_{12}(x,v) &= \bigl[0, \lambda_{12}(x,v)\bigr) \\
\Delta_{13}(x,v) &= \bigl[\lambda_{12}(x,v), \lambda_{12}(x,v) + \lambda_{13}(x,v)\bigr) \\
&\ \,\vdots \\
\Delta_{1M}(x,v) &= \Biggl[\sum_{j=2}^{M-1} \lambda_{1j}(x,v), \sum_{j=2}^{M} \lambda_{1j}(x,v)\Biggr) \\
\Delta_{21}(x,v) &= \Biggl[\sum_{j=2}^{M} \lambda_{1j}(x,v), \sum_{j=2}^{M} \lambda_{1j}(x,v) + \lambda_{21}(x,v)\Biggr) \\
&\ \,\vdots \\
\Delta_{2M}(x,v) &= \Biggl[\sum_{j=2}^{M} \lambda_{1j}(x,v) + \sum_{\substack{j=1\\ j\neq 2}}^{M-1} \lambda_{2j}(x,v), \sum_{j=2}^{M} \lambda_{1j}(x,v) + \sum_{\substack{j=1\\ j\neq 2}}^{M} \lambda_{2j}(x,v)\Biggr)
\end{aligned}$$

and so on. For fixed x and v, these are disjoint intervals, and the length of $\Delta_{ij}(x,v)$ is $\lambda_{ij}(x,v)$. Now define a function

$$h : \mathbb{R}^N \times \mathcal{S} \times \mathcal{V} \times \mathbb{R} \longrightarrow \mathbb{R}$$

by

$$h(x,i,v,z) = \begin{cases} j - i & \text{if } z \in \Delta_{ij}(x,v) \\ 0 & \text{otherwise.} \end{cases} \tag{2.3}$$

Let $\bigl(X(t), S(t)\bigr)$ be the $(\mathbb{R}^N \times \mathcal{S})$-valued *controlled switching diffusion* process given by the following stochastic differential equations

$$\begin{aligned}
dX(t) &= b\bigl(X(t), S(t), v(t)\bigr)\, dt + \operatorname{diag}(\sigma_1, \dots, \sigma_N)\, dW(t) \\
dS(t) &= \int_{\mathbb{R}} h\bigl(X(t), S(t-), v(t), z\bigr)\, \mathfrak{p}(dt, dz),
\end{aligned} \tag{2.4}$$

for $t \geq 0$, with $X(0) = X_0$ and $S(0) = S_0$, where

(i) X_0 is a prescribed $\mathbb{R}^N$-valued random variable.
(ii) S_0 is a prescribed $\mathcal{S}$-valued random variable.
(iii) $W(\cdot) = [W_1(\cdot), \dots, W_N(\cdot)]^T$ is an N-dimensional standard Wiener process independent of X_0, S_0.
(iv) $\mathfrak{p}(dt, dz)$ is an $\mathfrak{M}(\mathbb{R}_+ \times \mathbb{R})$-valued Poisson random measure with intensity $dt \times m(dz)$, where m is the Lebesgue measure on $\mathbb{R}$ ([25, p. 70]).

(v) $\mathfrak{p}(\cdot\,,\cdot)$ and $W(\cdot)$ are independent.

(vi) $v(\cdot)$ is a $\mathcal{V}$-valued process with measurable sample paths satisfying the following non-anticipativity property: Let $\mathfrak{F}_t^v = \sigma\{v(s) : s \le t\}$ and

$$\mathfrak{F}_{[t,\infty)}^{W,\mathfrak{p}} = \sigma\big\{W(s) - W(t),\, \mathfrak{p}(A,B) \,:\, A \in \mathfrak{B}\big([s,\infty)\big),\, B \in \mathfrak{B}(\mathbb{R}),\, s \ge t\big\}.$$

Then $\mathfrak{F}_t^v$ and $\mathfrak{F}_{[t,\infty)}^{W,\mathfrak{p}}$ are independent.

Such a process $v(\cdot)$ will be called an *admissible* (control) *policy.* If $v(\cdot)$ is a Dirac measure, i.e., $v(\cdot) = \delta_{u(\cdot)}$, where $u(\cdot)$ is a U-valued process, then it is called an admissible *non-randomized* policy. An admissible policy $v(\cdot)$ is called *feedback* if $v(\cdot)$ is progressively measurable with respect to the natural filtration of $\big(X(\cdot), S(\cdot)\big)$. A particular subclass of feedback policies is of special interest. A feedback policy $v(\cdot)$ is called a (non-homogeneous) *Markov* policy if $v(\cdot) = \tilde{v}\big(\cdot, X(\cdot), S(\cdot)\big)$ for a measurable map $\tilde{v} : \mathbb{R}_+ \times \mathbb{R}^N \times \mathcal{S} \to \mathcal{V}$. With an abuse of notation, the map $\tilde{v}$ itself is called a Markov policy. If $\tilde{v}$ has no explicit time dependence, it is called a *homogeneous* Markov policy. Thus, a homogeneous Markov non-randomized policy can be identified with a measurable map $v : \mathbb{R}^N \times \mathcal{S} \to U$.

If $\big(W(\cdot), \mathfrak{p}(\cdot,\cdot), X_0, S_0, v(\cdot)\big)$, satisfying the above, are given on a prescribed probability space $\big(\Omega, \mathfrak{F}, P\big)$, then under our assumptions on b and λ, equation (2.4) will admit an a.s. unique strong solution [22, Chap. 3], [25, Chap. 3, Sect. 2c] and $X(\cdot) \in C(\mathbb{R}_+; \mathbb{R}^N)$, $S(\cdot) \in D(\mathbb{R}_+; \mathcal{S})$, where $D(\mathbb{R}_+; \mathcal{S})$ is the space of right continuous functions on $\mathbb{R}_+$ with left limits taking values in $\mathcal{S}$. However, if $v(\cdot)$ is a feedback policy, then there exists a measurable map

$$f : \mathbb{R}_+ \times C(\mathbb{R}_+; \mathbb{R}^N) \times D(\mathbb{R}_+; \mathcal{S}) \longrightarrow \mathcal{V}$$

such that for each $t \ge 0$, $v(t) = f\big(t, X(\cdot), S(\cdot)\big)$ and is measurable with respect to the σ-field generated by $\big\{X(s), S(s) : s \le t\big\}$. Thus, $v(\cdot)$ cannot be specified a priori in (2.4). Instead, one has to replace $v(t)$ in (2.4) by $f\big(t, X(\cdot), S(\cdot)\big)$ and (2.4) takes the form

$$\begin{aligned} dX(t) &= b\big(X(t), S(t), f(t, X(\cdot), S(\cdot))\big)\, dt + \operatorname{diag}(\sigma_1, \dots, \sigma_N)\, dW(t) \\ dS(t) &= \int_{\mathbb{R}} h\big(X(t), S(t-), f(t, X(\cdot), S(\cdot)), z\big)\, \mathfrak{p}(dt, dz), \end{aligned} \tag{2.5}$$

for $t \ge 0$, with $X(0) = X_0$ and $S(0) = S_0$. In general, (2.5) will not even admit a weak solution. However, if the feedback policy is a Markov policy, then the existence of a unique strong solution can be established. We now introduce some notation which will be used throughout. Define

$$L^1(\mathbb{R}^N \times \mathcal{S}) = \big\{f : \mathbb{R}^N \times \mathcal{S} \longrightarrow \mathbb{R} \,:\, \text{ for each } i \in \mathcal{S}, f(\cdot, i) \in L^1(\mathbb{R}^N)\big\}.$$

$L^1(\mathbb{R}^N \times \mathcal{S})$ is endowed with the product topology of $\big(L^1(\mathbb{R}^N)\big)^M$. Similarly, we define $C_0^\infty(\mathbb{R}^N \times \mathcal{S})$, $W_{\ell oc}^{2,p}(\mathbb{R}^N \times \mathcal{S})$, etc. For $f \in W_{\ell oc}^{2,p}(\mathbb{R}^N \times \mathcal{S})$ and $u \in U$, we write

$$L^u f(x,i) = L_i^u f(x,i) + \sum_{j=1}^M \overline{\lambda}_{ij}(x,u)[f(x,j) - f(x,i)], \tag{2.6}$$

where

$$L_i^u f(x,i) = \frac{1}{2}\sum_{j=1}^{N} \sigma_j^2 \frac{\partial^2 f(x,i)}{\partial x_j^2} + \sum_{j=1}^{N} \bar{b}_j(x,i,u)\frac{\partial f(x,i)}{\partial x_j} \tag{2.7}$$

and more generally, for $v \in \mathcal{V}$,

$$L^v f(x,i) = \int_U L^u f(x,i)\, v(du). \tag{2.8}$$

THEOREM 2.1. *Under a Markov policy v, (2.4) admits an a.s. unique strong solution such that $\bigl(X(\cdot), S(\cdot)\bigr)$ is a Feller process with differential generator L^v.*

Proof (Sketch). This proof is based on the technique involving the removal of drift [41], [10, Thm. 1.4, pp. 10–12]. Clearly, it suffices to prove the result in the interval $[0,T]$, for a fixed $T>0$. For $T>0$, let H be the function space defined by

$$\begin{aligned} H = \Bigl\{ g \in W^{1,2,p}_{\ell oc}\bigl([0,T]\times\mathbb{R}^N\times\mathcal{S}\bigr),\ 2\le p<\infty :\ &\text{for each } i\in\mathcal{S},\\ \sup_{0\le t\le T} |g(t,x,i)| \text{ grows slower than } \exp(k\|x\|^2) &\text{ for all } k>0\Bigr\}. \end{aligned} \tag{2.9}$$

Fix an $i\in\mathcal{S}$. For $1\le j\le N$, let $\varphi_i(t,x,j)$ be the unique solution in H (as in (2.9)) of

$$\begin{aligned} \frac{\partial \varphi_i(t,x,j)}{\partial t} + L_j^{v(t,x,j)}\varphi_i(t,x,j) &= 0\\ \varphi_i(T,x,j) &= x_i\,, \end{aligned} \tag{2.10}$$

where $x=(x_1,\dots,x_N)$. Let $\varphi = [\varphi_1,\dots,\varphi_N]^T$. It can be shown that for fixed j, $\varphi(t,\cdot,j)$ is a homeomorphism onto its range for each $t\in[0,T]$. Set $Y(t) = \varphi\bigl(t,X(t),S(t)\bigr)$, $t\in[0,T]$. Using Ito's formula, it follows that $Y(t)$ satisfies

$$\begin{aligned} Y(t) = Y(0) &+ \int_0^t \Bigl[\bigl(D\varphi_s \operatorname{diag}(\sigma_1,\dots,\sigma_N)\bigr)\circ\varphi_s^{-1}\Bigr]\bigl(Y(s)\bigr)\,dW(s)\\ &+ \int_0^t\int_{\mathbb{R}} \Bigl[\varphi_s\Bigl(\varphi_s^{-1}(Y(s-)) + \tilde{h}\bigl(\varphi_s^{-1}(Y(s-)),z\bigr)\Bigr) - Y(s)\Bigr]\,\mathfrak{p}(ds,dz), \end{aligned} \tag{2.11}$$

where $D\varphi_s$, φ_s^{-1} denote respectively the Jacobian matrix and the inverse map of $\varphi\bigl(s,\cdot,S(s)\bigr)$, '$\circ$' indicates composition of functions and

$$\tilde{h}(\cdot,\cdot,\cdot) = [0,0,\dots,0,h(\cdot,\cdot,\cdot)]^T \in \mathbb{R}^{N+1}.$$

Now by [41], equation (2.11) has an a.s. unique strong solution which is a Markov process. The corresponding claim for $\bigl(X(t),S(t)\bigr)$ follows via the homeomorphic property of φ. It remains to show the Feller property. Pick any bounded continuous function $f:\mathbb{R}^N\times\mathcal{S}\to\mathbb{R}$. The system of equations

$$\begin{aligned} \frac{\partial \psi(t,x,i)}{\partial t} + L^{v(t,x,i)}\psi(t,x,i) &= 0\\ \psi(T,x,i) &= f(x,i) \end{aligned} \tag{2.12}$$

can be shown to have a unique solution in H [18]. Therefore, by Ito's formula, it follows that

$$\psi(t,x,i) = E\big[f\big(X(T),S(T)\big) \;\big|\; X(t)=x, S(t)=i\big],$$

where the expectation is under the Markov policy v. By Sobolev's imbedding theorem [5, p. 53], $H \subset C\big([0,T)\times\mathbb{R}^N\times\mathcal{S}\big)$ and hence $\psi(t,\cdot,i)$ is continuous for each $t\in[0,T)$. □

Some comments are in order now.

Remark 2.1.

(i) We have used Ito's formula for functions in $W^{1,2,p}_{\ell oc}(\mathbb{R}_+\times\mathbb{R}^N\times\mathcal{S})$. This generalization is due to Krylov [28, pp. 121–127] for "classical" diffusions. Its extension for the present system is routine.

(ii) The well-posedness of the Cauchy problem for the weakly coupled parabolic system (2.10) has been established in [18] under slightly stronger conditions on the first order terms. However, in view of the results in [3], [30, Chap. 7], its extension to the present case is straightforward.

Remark 2.2. We have seen in Theorem 2.1 that under a Markov policy the corresponding solution $\big(X(\cdot),S(\cdot)\big)$ of (2.4) is a Markov process. We have the following converse result. Let $v(\cdot)$ be a feedback policy, such that the corresponding solution $\big(X(\cdot),S(\cdot)\big)$ of (2.4) is a Markov process. Then $v(\cdot)$ may be taken to be a Markov policy. Since we do not need this result, we omit the proof.

2.1. The Optimization Problem. Let $\bar{c}:\mathbb{R}^N\times\mathcal{S}\times U\to\mathbb{R}_+$ be a bounded, continuous cost function, and let $c:\mathbb{R}^N\times\mathcal{S}\times\mathcal{V}\to\mathbb{R}_+$ be defined as

$$c(\cdot,\cdot,v) = \int_U \bar{c}(\cdot,\cdot,u)\,v(du)\,.$$

Let $\alpha>0$ be a prescribed discount factor. Let $v(\cdot)$ be an admissible policy and $\big(X(\cdot),S(\cdot)\big)$ the corresponding process. Then the total α-discounted cost under $v(\cdot)$ is defined as

$$J_v(x,i) := E\left[\int_0^\infty e^{-\alpha t} c\big(X(t),S(t),v(t)\big)\,dt \;\middle|\; X(0)=x, S(0)=i\right]. \tag{2.13}$$

If the laws of X_0, S_0 are $\pi\in\boldsymbol{\mathcal{P}}(\mathbb{R}^N)$, $\xi\in\boldsymbol{\mathcal{P}}(\mathcal{S})$ respectively, then

$$J_v(\pi,\xi) = \sum_i \int_{\mathbb{R}^N} J_v(x,i)\,\pi(dx)\,\xi(i). \tag{2.14}$$

Let

$$V(x,i) := \inf_{v(\cdot)}\big\{J_v(x,i)\big\}, \tag{2.15}$$

$$V(\pi,\xi) := \inf_{v(\cdot)}\big\{J_v(\pi,\xi)\big\}. \tag{2.16}$$

The function $V(x,i)$ is called the (α-discounted) value function. An admissible policy $v(\cdot)$ satisfying

$$J_v(\pi,\xi) = V(\pi,\xi)$$

is called an optimal policy for the initial law (π, ξ). An admissible policy is called optimal if it is optimal for any initial law. Our aim is to find an admissible optimal policy which is homogeneous Markov and non-randomized.

We now introduce the (discounted) occupation measures [12]. Let $v(\cdot)$ be an admissible policy and $\bigl(X(\cdot), S(\cdot)\bigr)$ the corresponding process with initial law (π, ξ). Define the occupation measure $\nu[\pi, \xi; v] \in \boldsymbol{\mathcal{P}}(\mathbb{R}^N \times \mathcal{S} \times U)$ by

$$\int f \, d\nu[\pi, \xi; v] = \alpha E\biggl[\int_0^\infty e^{-\alpha t} \int_U f(X(t), S(t), u)\, v(t)(du)\, dt\biggr] \tag{2.17}$$

for $f \in C_b(\mathbb{R}^N \times \mathcal{S} \times U)$. Also, we define

$$M_1[\pi, \xi] = \bigl\{\nu[\pi, \xi; v] \,:\, v(\cdot) \text{ is admissible}\bigr\} \tag{2.18}$$

$$M_2[\pi, \xi] = \bigl\{\nu[\pi, \xi; v] \,:\, v(\cdot) \text{ is homogeneous Markov}\bigr\} \tag{2.19}$$

$$M_3[\pi, \xi] = \bigl\{\nu[\pi, \xi; v] \,:\, v(\cdot) \text{ is homogeneous non-randomized Markov}\bigr\} \tag{2.20}$$

In terms of these occupation measures

$$J_v(\pi, \xi) = \alpha^{-1} \int \overline{c}\, d\nu[\pi, \xi; v]. \tag{2.21}$$

We will show in Section 4 that $M_1[\pi, \xi] = M_2[\pi, \xi]$ and that $M_2[\pi, \xi]$ is compact, convex and $M_2^e[\pi, \xi] \subset M_3[\pi, \xi]$, where $M_2^e[\pi, \xi]$ is the set of extreme points of $M_2[\pi, \xi]$. Thus, for a fixed initial law, the optimization problem (2.13) will reduce to a convex optimization problem in view of (2.21).

3. Compactness of Laws. We will establish the compactness of laws of the process $\bigl(X(\cdot), S(\cdot)\bigr)$ under various policies using the approach in [11, Chap. 2]. Let $\pi_0 \in \boldsymbol{\mathcal{P}}(\mathbb{R}^N)$, $\xi_0 \in \boldsymbol{\mathcal{P}}(\mathcal{S})$. Let $\mathcal{L}_i[\pi_0, \xi_0] \subset \boldsymbol{\mathcal{P}}\bigl(C(\mathbb{R}_+; \mathbb{R}^N) \times D(\mathbb{R}_+; \mathcal{S})\bigr)$, $i = 1, 2, 3$, denote the set of laws of $\bigl(X(\cdot), S(\cdot)\bigr)$ under all admissible/Markov/homogeneous Markov policies with fixed initial law (π_0, ξ_0).

THEOREM 3.1. *The set $\mathcal{L}_1[\pi_0, \xi_0]$ is compact in $\boldsymbol{\mathcal{P}}\bigl(C(\mathbb{R}_+; \mathbb{R}^N) \times D(\mathbb{R}_+; \mathcal{S})\bigr)$.*

Proof. It clearly suffices to replace $\mathbb{R}_+$ by $[0, T]$ for arbitrary $T > 0$. Fix $T > 0$. Let $\bigl(X^n(\cdot), S^n(\cdot), W^n(\cdot), \mathfrak{p}^n(\cdot, \cdot), v^n(\cdot), X_0^n, S_0^n\bigr)$, $n \geq 1$, satisfy (2.4) on probability spaces $(\Omega^n, \mathfrak{F}^n, P^n)$ respectively, the laws of X_0^n, S_0^n being π_0, ξ_0 respectively for all n. Let $\{f_i\}$ be a countable dense subset of the unit ball of $C(U)$. Define $\beta_j^n(t) = \int f_j \, dv^n(t)$, $t \in [0, T]$. Let B denote closed unit ball of $L^\infty[0, T]$ with the topology given by the weak topology of $L^2[0, T]$ relativized to B. Let E be a countable product of replicas of B. Since B is compact and metrizable and hence Polish, the same follows for E. Let $\beta^n(\cdot) = [\beta_1^n(\cdot), \beta_2^n(\cdot), \dots]$, $n \geq 1$, viewed as E-valued random variables. Using the assumed conditions on b, it can be easily shown that for $t_1, t_2 \in [0, T]$,

$$E\Bigl[\|X^n(t_2) - X^n(t_1)\|^4\Bigr] \leq K|t_2 - t_1|^2$$

for some T-dependent $K > 0$. It follows that the laws of the sequence $\{X^n(\cdot)\}$ are tight in $\boldsymbol{\mathcal{P}}\bigl(C(\mathbb{R}_+; \mathbb{R}^N)\bigr)$. Since $\mathcal{S}$ is finite and E is compact, it follows by Prohorov's theorem [24, Thm. 2.6, p. 7] that, for $A_1 \in \mathfrak{B}(\mathbb{R}_+)$, $A_2 \in \mathfrak{B}(\mathbb{R})$ fixed, the sequence $\bigl(X^n(\cdot), S^n(\cdot), \beta^n(\cdot), W^n(\cdot), \mathfrak{p}^n(A_1 \times A_2)\bigr)$ converges to a limit

$$\bigl(X(\cdot), S(\cdot), \beta(\cdot), W(\cdot), \mathfrak{p}(A_1 \times A_2)\bigr)\,.$$

Dropping to a subsequence if necessary and invoking Skorohod's theorem [24, p. 9], we may assume that all these random variables are defined on a common probability space and the convergence is a.s. on this probability space. By [11, Lemma II.1.2, p. 24] we can find a $\mathcal{V}$-valued process $v(\cdot)$ such that $\beta_i(t) = \int f_i\, dv(t)$, $i \geq 1$. Define $Z^n(\cdot) = [Z_1^n(\cdot), \ldots, Z_N^n(\cdot)]^T$, $Y^n(\cdot)$, $n \geq 1$, by

$$Z_i^n(t) = X_i^n(t) - \int_0^t b_i\big(X^n(s), S^n(s), v^n(s)\big)\, ds, \qquad t \geq 0$$

$$Y^n(t) = S^n(t) - \sum_{j=1}^M \int_0^t \lambda_{S^n(s-),j}\big(X^n(s), v^n(s)\big)\,\big(j - S^n(s-)\big)\, ds, \qquad t \geq 0$$

and $Z(\cdot) = [Z_1(\cdot), \ldots, Z_N(\cdot)]^T$, $Y(\cdot)$ by

$$Z_i(t) = X_i(t) - \int_0^t b_i\big(X(s), S(s), v(s)\big)\, ds, \qquad t \geq 0$$

$$Y(t) = S(t) - \sum_{j=1}^M \int_0^t \lambda_{S(s-),j}\big(X(s), v(s)\big)\,\big(j - S(s-)\big)\, ds, \qquad t \geq 0.$$

Then, by [11, Lemma II.1.3, p. 26] and standard representation theorems for semimartingales [25, pp. 172–178] applied to $Z_i(t)$ and $Y(t)$, it follows that on an augmented probability space $\big(X(\cdot), S(\cdot)\big)$ satisfies (2.4) for an admissible policy $v(\cdot)$ and driven by a Wiener process $\widetilde{W}(\cdot)$ and a Poisson random measure $\tilde{\mathfrak{p}}(\cdot,\cdot)$. □

We now state the next theorem without proof as it would be almost identical to the proof of [11, Thm. II.2.1, p. 29], in view of the estimates in [30, p. 582].

THEOREM 3.2. *The sets $\mathcal{L}_2[\pi_0, \xi_0]$, $\mathcal{L}_3[\pi_0, \xi_0]$ are compact.*

Let $\{v_n\}$ be a sequence of homogeneous Markov policies and $\big(X^n(\cdot), S^n(\cdot)\big)$ the corresponding solutions of (2.4) with $X^n(0) = x_0$, $S^n(0) = i_0$ for all $n \geq 0$. Let $p^n(t, x_0, i_0, y, j)$ be the fundamental solutions corresponding to the operators $\big(\frac{\partial}{\partial t} + L^{v_n}\big)$. Let $\big(X^n(\cdot), S^n(\cdot)\big) \longrightarrow \big(X^\infty(\cdot), S^\infty(\cdot)\big)$, where the latter is governed by a homogeneous Markov policy v_∞. Then, using the Hölder estimates on $p^n(t, x_0, i_0, y, j)$ [30, p. 582], one can show the following result as in [10, Thm. II.2.2, p. 33].

LEMMA 3.1. *For each $t > 0$, $p^n(t, x_0, i_0, \cdot, \cdot) \longrightarrow p^\infty(t, x_0, i_0, \cdot, \cdot)$ in $L^1(\mathbb{R}^N \times \mathcal{S})$. In other words, the laws of $\big(X^n(t), S^n(t)\big)$ converge to that of $\big(X^\infty(t), S^\infty(t)\big)$ in total variation.*

Following [11, p. 30], we topologize the space of all homogeneous Markov policies. Let

$$F = \big\{v : \mathbb{R}^N \times \mathcal{S} \longrightarrow \mathcal{V} \,:\, v \text{ is measurable}\big\}.$$

Topologize F as in [11, p. 30]. Then F is a compact metric space. Its topology is determined by the following convergence criterion [11, Lemma II.2.1, p. 32].

LEMMA 3.2. *Let $f \in L^2(\mathbb{R}^N \times \mathcal{S}) \cap L^1(\mathbb{R}^N \times \mathcal{S})$, $g \in C_b(\mathbb{R}^N \times \mathcal{S} \times U)$ and $v_n \to v$ in F. Then*

$$\int_{\mathbb{R}^N} f(x,i) \int_{\mathcal{S}} g(x,i,\cdot)\, dv_n(x,i)\, dx \underset{n\to\infty}{\longrightarrow} \int_{\mathbb{R}^N} f(x,i) \int_{\mathcal{S}} g(x,i,\cdot)\, dv(x,i)\, dx \tag{3.1}$$

for each $i \in \mathcal{S}$. Conversely, if (3.1) holds for all such f, g and $i \in \mathcal{S}$, then $v_n \to v$ in F.

Let $\mathcal{L}(v)$ denote the law of $\big(X(\cdot), S(\cdot)\big)$ when $X(0) = x_0$, $S(0) = i_0$ and the homogeneous Markov policy v is used. Using Lemma 3.1, the following theorem can be proved exactly the same way as in [11, Thm. II.2.3, p. 34].

THEOREM 3.3. *The map* $v \longmapsto \mathcal{L}(v)$ *from* F *into* $\boldsymbol{\mathcal{P}}\big(C(\mathbb{R}_+; \mathbb{R}^N) \times D(\mathbb{R}_+; \mathcal{S})\big)$ *is continuous.*

4. Convexity and Extremality of Occupation Measures. In this Section we will study the properties of the occupation measures $\nu[\pi, \xi; v]$ introduced in (2.17), following the approach in [12].

LEMMA 4.1. *The sets* $M_1[\pi, \xi]$, $M_2[\pi, \xi]$, $M_3[\pi, \xi]$ *as defined in (2.18)–(2.20) are compact.*

Proof. This follows from Theorems 3.1 and 3.2. ∎

LEMMA 4.2. *For each fixed initial law* (π, ξ), $M_1[\pi, \xi] = M_2[\pi, \xi]$.

Proof. Let $\nu[\pi, \xi; v] \in M_1$. Disintegrate it as

$$\nu[\pi, \xi; v](dx \times \{i\} \times du) = \overline{\nu}[\pi, \xi; v](dx \times \{i\})\, \overline{v}(x, i)(du), \tag{4.1}$$

where $\overline{\nu}[\pi, \xi; v]$ is the marginal of $\nu[\pi, \xi; v]$ on $\mathbb{R}^N \times \mathcal{S}$ and $\overline{v}(x, i)$ is a version of the regular conditional law defined $\overline{\nu}[\pi, \xi; v]$ a.s. Pick any version from this equivalence class and keep it fixed henceforth. The map $\overline{v}(\cdot, \cdot)$ obviously defines a homogeneous Markov policy. Let $\big(X'(\cdot), S'(\cdot)\big)$ be the solution of (2.4) with $v(\cdot)$ replaced by $v'(\cdot) = \overline{v}\big(X'(\cdot), S'(\cdot)\big)$ and with initial law (π, ξ). Let $f \in C_b(\mathbb{R}^N \times \mathcal{S} \times U)$ and let

$$\varphi(x, i) = E\bigg[\int_0^\infty e^{-\alpha t} \int_U f\big(X'(t), S'(t), u\big)\, v'(t)(du)\, dt \;\bigg|\; X'(0) = x, S'(0) = i\bigg]. \tag{4.2}$$

Using the strong Markov property of $\big(X'(\cdot), S'(\cdot)\big)$ (this follows from the Feller property) and the local solvability of weakly coupled systems of elliptic equations [31, Chap. 7, p. 388] it can be shown by employing standard arguments involving Ito's formula that $\varphi(x, i)$ is the unique solution in $W^{2,p}_{\ell oc}(\mathbb{R}^N \times \mathcal{S}) \cap C_b(\mathbb{R}^N \times \mathcal{S})$, $2 \le p < \infty$, to

$$L^{\overline{v}(x,i)} \varphi(x, i) - \alpha \varphi(x, i) + \int_U f(x, i, u)\, \overline{v}(x, i)(du) = 0. \tag{4.3}$$

Define a process $Y(\cdot)$ by

$$Y(t) = \int_0^t \int_U e^{-\alpha s} f(X(s), S(s), u)\, v(s)(du)\, ds + e^{-\alpha t} \varphi\big(X(t), S(t)\big).$$

Then

$$\begin{aligned} E[Y(t)] - E[Y(0)] &= E[Y(t)] - \sum_{j=1}^{M} \int_{R^N} \varphi(x, j)\, \pi(dx)\, \xi(j) \\ &= E\bigg[\int_0^t e^{-\alpha s} \Big[L^{v(s)} \varphi\big(X(s), S(s)\big) - \alpha \varphi\big(X(s), S(s)\big) \\ &\qquad + \int_U f\big(X(s), S(s), u\big)\, v(s)(du)\Big]\, ds\bigg]. \end{aligned} \tag{4.4}$$

Letting $t \to \infty$ and using the definition of $\overline{v}(\cdot,\cdot)$ and (4.3), it follows that the right-hand side in (4.4) tends to zero (cf. [11, Thm. 4.2, pp. 40–42]). Thus,

$$\begin{aligned}\lim_{t\to\infty} E[Y(t)] &= E[\varphi(X_0, S_0)] \\ &= E[\varphi(X_0', S_0')].\end{aligned}$$

Since $f \in C_b(\mathbb{R}^N \times \mathcal{S} \times U)$ was arbitrary, it follows that $\nu[\pi,\xi;v] = \nu[\pi,\xi;\overline{v}]$. ∎

Let $\nu[\pi,\xi;v] \in M_2[\pi,\xi]$. By a routine extension of the inequality [28, p. 66] it follows that $\overline{\nu}[\pi,\xi;v]$ (as in (4.1)) is absolutely continuous with respect to the product of the Lebesgue measure on $\mathbb{R}^N$ and the counting measure on $\mathcal{S}$ and therefore has a density $\varphi[\pi,\xi;v]$. Let $\tilde{\nu}[\pi,\xi;v]$ be the marginal of $\overline{\nu}[\pi,\xi;v]$ on S. With 'supp' denoting the support of a measure, let

$$\operatorname{supp}(\tilde{\nu}[\pi,\xi;v]) = S_1[\pi,\xi;v] \subset \mathcal{S}. \tag{4.5}$$

It is not difficult to see that $\varphi[\pi,\xi;v](x,i) > 0$ a.e. $x \in \mathbb{R}^N$, $i \in S_1[\pi,\xi;v]$ and $\varphi[\pi,\xi;v](x,i) = 0$ for $i \in \mathcal{S} \setminus S_1[\pi,\xi;v]$. For $f \in W^{2,p}_{\ell oc}(\mathbb{R}^N \times \mathcal{S})$ define

$$L^v_\alpha f(x,i) = L^{v(x,i)} f(x,i) - \alpha f(x,i)\,. \tag{4.6}$$

Then, $\varphi[\pi,\xi;v]$ is the unique solution in $L^1(\mathbb{R}^N \times \mathcal{S})$ to:

$$\begin{aligned}&\sum_{i=1}^M \int_{\mathbb{R}^N} L^{v(x,i)}_\alpha g(x,i)\varphi(x,i)\,dx = -\sum_{i=1}^M \int_{\mathbb{R}^N} g(x,i)\pi(dx)\xi(i)\\ &\sum_{i=1}^M \int_{\mathbb{R}^N} \varphi(x,i)\,dx = 1\,, \qquad \varphi(x,i) \ge 0\,,\end{aligned} \tag{4.7}$$

for every $g \in C^\infty_0(\mathbb{R}^N \times \mathcal{S})$. Using the above, we will show that $M_2[\pi,\xi]$ is convex.

LEMMA 4.3. *The set $M_2[\pi,\xi]$ is convex.*

Proof. Let v_1, v_2 be two homogeneous Markov policies and $0 \le a \le 1$. Define a homogeneous Markov policy by

$$v(x,i) = \frac{a\varphi[\pi,\xi;v_1](x,i)v_1(x,i) + (1-a)\varphi[\pi,\xi;v_2](x,i)v_2(x,i)}{a\varphi[\pi,\xi;v_1](x,i) + (1-a)\varphi[\pi,\xi;v_2](x,i)} \tag{4.8}$$

for $(x,i) \in \mathbb{R}^N \times \big\{S_1[\pi,\xi;v_1] \cup S_1[\pi,\xi;v_2]\big\}$ and arbitrary otherwise. Let $f \in C^\infty_0(\mathbb{R}^N \times \mathcal{S})$. It is easy to see that

$$L^{v(x,i)}_\alpha f(x,i) = \frac{a\varphi[\pi,\xi;v_1](x,i)L^{v_1(x,i)}_\alpha f(x,i) + (1-a)\varphi[\pi,\xi;v_2](x,i)L^{v_2(x,i)}_\alpha f(x,i)}{a\varphi[\pi,\xi;v_1](x,i) + (1-a)\varphi[\pi,\xi;v_2](x,i)}\,.$$

Let $\varphi(x,i) = a\varphi[\pi,\xi;v_1](x,i) + (1-a)\varphi[\pi,\xi;v_2](x,i)$. From (4.7) and (4.8) it follows that $\varphi = \varphi[\pi,\xi;v]$. Thus

$$\begin{aligned}\nu[\pi,\xi;v]&(dx \times \{i\} \times du)\\ &= \varphi[\pi,\xi;v](x,i)\,dx\,v(x,i)(du)\\ &= a\varphi[\pi,\xi;v_1)(x,i)\,dx\,v_1(x,i)(du) + (1-a)\varphi[\pi,\xi;v_2](x,i)\,dx\,v_2(x,i)(du)\\ &= \big(a\nu[\pi,\xi;v_1] + (1-a)\nu[\pi,\xi;v_2]\big)(dx \times \{i\} \times du).\end{aligned}$$

∎

Let

$$\mathcal{I}[\pi,\xi] = \{\overline{\nu}[\pi,\xi;v] \in \boldsymbol{\mathcal{P}}(\mathbb{R}^N \times \mathcal{S}) \,:\, \nu[\pi,\xi;v] \in M_2[\pi,\xi]\}, \tag{4.9}$$

where $\overline{\nu}[\pi,\xi;v]$ is as in (4.1).

The proof of the next lemma is analogous to that of [12, Lemma 3.2]. We present a brief sketch describing the essential ideas.

LEMMA 4.4. *The set $\mathcal{I}[\pi,\xi]$ is compact in $\boldsymbol{\mathcal{P}}(\mathbb{R}^N \times \mathcal{S})$ in total variation.*

Proof. By a routine extension of the inequality [28, p. 66] to the present case, $\varphi[\pi,\xi;v]$ will be uniformly bounded in $L^p(\mathbb{R}^N)$. For the sake of convenience, assume that the initial condition is $(x_0,i_0) \in \mathbb{R}^N \times \mathcal{S}$. As in [11, Lemma 5.2, p. 44] we can show by considering appropriate estimates on the weakly coupled systems of elliptic equations [31, Chap. 7] that for any bounded open set A such that $\overline{A} \subset \mathbb{R}^N \setminus \{x_0\}$ and $i \in \mathcal{S} \setminus \{i_0\}$ there exists a $\beta > 0$ and a $K \in (0,\infty)$ such that

$$\bigl|\varphi[x_0,i_0;v](y,i) - \varphi[x_0,i_0;v](z,i)\bigr| \le K\|y-z\|^\beta, \qquad y,z \in A, \tag{4.10}$$

under any choice of a homogeneous Markov policy v. By Theorem 3.2, $\mathcal{I}[x_0,i_0]$ is compact in the Prohorov topology of $\boldsymbol{\mathcal{P}}(\mathbb{R}^N \times \mathcal{S})$. Let $\{\overline{\nu}[x_0,i_0;v_n]\}$ be a sequence in $\mathcal{I}[x_0,i_0]$ and $\overline{\nu}[x_0,i_0;v_\infty]$ a weak limit point of $\{\overline{\nu}[x_0,i_0;v_n]\}$. We need to show that $\varphi[x_0,i_0;v_n] \longrightarrow \varphi[x_0,i_0;v_\infty]$ in $L^1(\mathbb{R}^N \times \mathcal{S})$. The equicontinuity of $\{\varphi[x_0,i_0;v_n]\}$ follows from (4.10). Also (4.10) together with the uniform L^p-estimates implies pointwise boundedness. Thus, by the Arzela-Ascoli theorem, we may drop to a subsequence, if necessary, to conclude that for each $i \in \mathcal{S}$,

$$\varphi[x_0,i_0,v_n](\cdot,i) \longrightarrow \psi(\cdot,i),$$

for some $\psi(\cdot,i)$, uniformly on compact subsets of $\mathbb{R}^N$. By the uniform L^p-estimates the convergence is also in $L^1(\mathbb{R}^N)$. Thus

$$\int_{\mathbb{R}^N} \varphi[x_0,i_0;v_n](y,i) f(y)\,dy \underset{n\to\infty}{\longrightarrow} \int_{\mathbb{R}^N} \psi(y,i) f(y)\,dy, \tag{4.11}$$

for all $f \in C_b(\mathbb{R}^N)$. But (4.11) certainly holds with $\varphi[x_0,i_0;v_\infty](\cdot,i)$ replacing $\psi(\cdot,i)$. Therefore, $\varphi[x_0,i_0;v_\infty] \equiv \psi$. □

We are now in a position to characterize the extreme points of $M_2[\pi,\xi]$. Let v be a homogeneous Markov policy such that, for each $x \in \mathbb{R}^N$ and $i \in \mathcal{S}$,

$$v(x,i) = av_1(x,i) + (1-a)v_2(x,i), \tag{4.12}$$

where $a \in (0,1)$ and v_1, v_2 are distinct homogeneous Markov policies, i.e., there exists at least one $i_0 \in \mathcal{S}$ such that $v_1(\cdot,i_0)$ and $v_2(\cdot,i_0)$ differ on a set of strictly positive measure. The proof of the next lemma closely follows that of [12, Lemma 3.3]; we therefore present only a brief sketch of the proof.

LEMMA 4.5. *Let v be as in (4.12). Then, $\nu[\pi,\xi;v]$ is not an extreme point of $M_2[\pi,\xi]$.*

Proof. We will show that if v satisfies (4.12), then there are homogeneous Markov policies $\tilde{v}_1, \tilde{v}_2$ and $b \in (0,1)$ such that

$$\nu[\pi,\xi;v] = b\nu[\pi,\xi;\tilde{v}_1] + (1-b)\nu[\pi,\xi;\tilde{v}_2]\,.$$

It suffices to find $b \in (0,1)$ and $\tilde{v}_1, \tilde{v}_2$ satisfying

$$v(x,i) = \frac{b\varphi[\pi,\xi;\tilde{v}_1](x,i)\tilde{v}_1(x,i) + (1-b)\varphi[\pi,\xi;\tilde{v}_2](x,i)\tilde{v}_2(x,i)}{b\varphi[\pi,\xi;\tilde{v}_1](x,i) + (1-b)\varphi[\pi,\xi;\tilde{v}_2](x,i)} \tag{4.13}$$

for $(x,i) \in \mathbb{R}^N \times \{S_1[\pi,\xi;\tilde{v}_1] \cup S_1[\pi,\xi;\tilde{v}_2]\}$ (see (4.5)). For $R > 0$, let v_1', v_2' be homogeneous Markov policies defined by

$$v_j'(x,i) = \begin{cases} v_j(x,i), & \|x\| \le R \\ v(x,i), & \|x\| > R \end{cases} \qquad i \in \mathcal{S}, \quad j = 1,2. \tag{4.14}$$

Let $\overline{v}(\cdot)$ be a given homogeneous Markov policy. Define a homogeneous Markov policy v_2'' via

$$\begin{aligned} v(x,i) &= av_1'(x,i) + (1-a)v_2'(x,i) \\ &= \frac{b\varphi[\pi,\xi;v_1'](x,i)v_1'(x,i) + (1-b)\varphi[\pi,\xi;\overline{v}](x,i)v_2''(x,i)}{b\varphi[\pi,\xi;v_1'](x,i) + (1-b)\varphi[\pi,\xi;\overline{v}](x,i)} \end{aligned} \tag{4.15}$$

for $(x,i) \in \mathbb{R}^N \times \{S_1[\pi,\xi;v_1') \cup S_1[\pi,\xi;\overline{v}]\}$ and arbitrary otherwise. The arguments used in the proof of [12, Lemma 3.3] mutalis mutandis will ensure a suitable choice of $b \in (0,1)$ such that v_2'' is a genuine homogeneous Markov policy. Fix a $b \in (0,1)$ as in (4.15). Given a homogeneous Markov policy $\overline{v}(\cdot)$, we obtain $v_2''(\cdot)$ via (4.15). Thus, we have a map $\overline{\nu}[\pi,\xi;\overline{v}] \longmapsto \overline{\nu}[\pi,\xi;v_2'']$ from $\mathcal{I}[\pi,\xi]$ to $\mathcal{I}[\pi,\xi]$. Using Lemma 4.6, it can be shown as in the proof of [11, Lemma 3.3] that this map is continuous in the total variation. By Schauder's fixed point theorem [26, p. 220], this map has a fixed point. In other words, there exists a homogeneous Markov policy v_2'' such that

$$v(x,i) = \frac{b\varphi[\pi,\xi;v_1'](x,i)v_1'(x,i) + (1-b)\varphi[\pi,\xi;v_2'']v_2''(x,i)}{b\varphi[\pi,\xi;v_1'](x,i) + (1-b)\varphi[\pi,\xi;v_2''](x,i)}$$

for $(x,i) \in \mathbb{R}^N \times \{S_1[\pi,\xi;v_1'] \cup S_1[\pi,\xi;v_2'']\}$. Since $v_1' \neq v$ on a set of strictly positive measure for sufficiently large R, $v_2'' \neq v_1'$ on this set. Thus

$$\nu[\pi,\xi;v] = b\nu[\pi,\xi';v_1'] + (1-b)\nu[\pi,\xi;v_2'']$$

as desired. ∎

The results in this Section are now summarized as follows.

THEOREM 4.1. *$M_1[\pi,\xi] = M_2[\pi,\xi]$, and $M_2[\pi,\xi]$ is compact and convex, and each of its extreme points corresponds to some $\nu[\pi,\xi;v]$, where v is a homogeneous Markov non-randomized policy.*

5. Existence of an Optimal Policy. Using the results of the previous Section, we will establish the existence of an optimal policy.

THEOREM 5.1. *There exists a homogeneous Markov optimal policy.*

Proof. Let $(\pi,\xi) \in \boldsymbol{\mathcal{P}}(\mathbb{R}^N) \times \boldsymbol{\mathcal{P}}(\mathcal{S})$ such that $\text{supp}(\pi) = \mathbb{R}^N$ and $\text{supp}(\xi) = S$. Since $\bar{c}$ is bounded and continuous the map $M_2[\pi,\xi] \ni \nu \longmapsto \int \bar{c}\, d\nu$ is continuous. Thus, there exists a homogeneous Markov policy v^* such that

$$J_{v^*}(\pi,\xi) = \min_v \{J_v(\pi,\xi) \,:\, v \text{ is homogeneous Markov}\}.$$

By Lemma 4.2, it follows that

$$J_{v^*}(\pi, \xi) = V(\pi, \xi)\,.$$

Therefore, v^* is optimal for the initial law (π, ξ). We will show that v^* is optimal for any initial law. It suffices to show that v^* is optimal for any initial condition $(x, i) \in \mathbb{R}^N \times \mathcal{S}$. Suppose there exist $(x_0, i_0) \in \mathbb{R}^N \times \mathcal{S}$ and a homogeneous Markov policy v such that

$$J_v(x_0, i_0) < J_{v^*}(x_0, i_0). \tag{5.1}$$

Using the fact that the solution of (2.4) under a Markov policy is a Feller process, it can be easily shown that the function $J_v(x, i)$ is continuous in x for each v. Thus, (5.1) holds in a neighborhood B of x_0. Define a policy v' by

$$v'(t) = v^*\bigl(X(t), S(t)\bigr) I\{X_0 \notin B\} + v'\bigl(X(t), S(t)\bigr) I\{X_0 \in B\},$$

where $\bigl(X(\cdot), S(\cdot)\bigr)$ is governed by $v'(\cdot)$. Then, it is easily shown that

$$J_{v'}(\pi, \xi) < J_{v^*}(\pi, \xi),$$

which is a contradiction. Thus, v^* is optimal. □

THEOREM 5.2. *There exists a homogeneous Markov non-randomized optimal policy.*

Proof. Let v^* be as in Theorem 5.1. Let $M_2^e[\pi, \xi]$ be the set of extreme points of $M_2[\pi, \xi]$. Since $M_2[\pi, \xi]$ is compact, by Choquet's theorem [34], $\nu[\pi, \xi; v^*]$ is the barycenter of a probability measure m supported on $M_2^e[\pi, \xi]$. Therefore,

$$\int \overline{c}\, d\nu[\pi, \xi; v^*] = \int_{M_2^e[\pi,\xi]} \Bigl(\int \overline{c}\, d\mu\Bigr) m(d\mu). \tag{5.2}$$

Since v^* is optimal, it follows from (5.2) that there exists a $\nu[\pi, \xi; v] \in M_2^e[\pi, \xi]$ such that

$$\int \overline{c}\, d\nu[\pi, \xi; v^*] = \int \overline{c}\, d\nu[\pi, \xi; v]\,.$$

Thus, v is also optimal. By Theorem 4.1 it is non-randomized. □

6. Dynamic Programming Equations. Using the existence results of the previous Section, we will now derive the dynamic programming or Hamilton-Jacobi-Bellman equations (HJB) which in our case will be a weakly coupled system of quasi-linear elliptic equations, and then characterize the optimal policy as a minimizing selector of an appropriate "Hamiltonian". The HJB equations for our problem are

$$\alpha\psi(x, i) = \inf_{u \in U} \bigl\{L^u \psi(x, i) + \overline{c}(x, i, u)\bigr\}. \tag{6.1}$$

THEOREM 6.1. *The value function $V(x, i)$ is the unique solution of (6.1) in the space $W^{2,p}_{\ell oc}(\mathbb{R}^N \times \mathcal{S}) \cap C_b(\mathbb{R}^N \times \mathcal{S})$ for any $2 \le p < \infty$.*

Proof. We have already seen in the proof of Theorem 5.1 that $V(x, i) \in C_b(\mathbb{R}^N \times \mathcal{S})$. Let v^* be a homogeneous Markov non-randomized optimal policy and $\bigl(X(\cdot), S(\cdot)\bigr)$ the corresponding solution of (2.4). Then, for $(x, i) \in \mathbb{R}^N \times \mathcal{S}$,

$$V(x, i) = E\biggl[\int_0^\infty e^{-\alpha t} \overline{c}\bigl(X(t), S(t), v^*(X(t), S(t))\bigr)\, dt \biggm| X(0) = x, S(0) = i\biggr]. \tag{6.2}$$

By standard arguments (see the arguments following (4.2)), $V(x,i)$ is the unique solution in $W^{2,p}_{\ell oc}(\mathbb{R}^N \times \mathcal{S}) \cap C_b(\mathbb{R}^N \times \mathcal{S})$, for any $2 \le p < \infty$, of

$$\alpha V(x,i) = L^{v^*(x,i)} V(x,i) + \overline{c}\big(x,i,v^*(x,i)\big). \tag{6.3}$$

Suppose there exist $x_0 \in \mathbb{R}^N$, $i_0 \in \mathcal{S}$, $u \in U$ and $\delta > 0$ such that

$$\alpha V(x_0,i_0) > L^u V(x_0,i_0) + \overline{c}(x_0,i_0,u) + \delta\,.$$

Then, by the continuity of $V(\cdot,i_0)$, the above will hold in a neighborhood $N(x_0)$ of x_0. Define a homogeneous Markov non-randomized policy $\tilde{v}$ as follows:

$$\tilde{v}(x,i) = \begin{cases} v^*(x,i) & \text{if } (x,i) \notin N(x_0) \times \mathcal{S} \\ u & \text{if } (x,i) \in N(x_0) \times \mathcal{S}. \end{cases}$$

Then

$$\alpha V(x,i_0) > L^{\tilde{v}(x,i_0)} V(x,i_0) + \overline{c}\big(x,i_0,\tilde{v}(x,i_0)\big) + \delta I\big\{x \in N(x_0)\big\}.$$

It is easily seen that

$$V(x,i_0) \ge J_{\tilde{v}}(x,i_0) + \delta'$$

for some $\delta' > 0$, which is a contradiction. Hence, $V(x,i)$ satisfies (6.1). Let V' be another solution of (6.1) in the desired class. Then it can be shown using standard arguments (cf. [11, Thm. III.2.4, pp. 69–70]) that

$$|V(x,i) - V'(x,i)| \le 2Ke^{-\alpha t},$$

where $K > 0$ is a constant. Letting $t \to \infty$, $V \equiv V'$. ∎

Corollary 6.1. *Assume that for each $i \in \mathcal{S}$, $\overline{c}(\cdot,i,\cdot)$ is Lipschitz in its first argument uniformly with respect to the third. Then $V(x,i)$ is the unique solution of (6.1) in $C^2(\mathbb{R}^N \times \mathcal{S}) \cap C_b(\mathbb{R}^N \times \mathcal{S})$.*

Proof. It suffices to show that V is C^2. Since $V(x,i) \in W^{2,p}_{\ell oc}(\mathbb{R}^N \times \mathcal{S})$ for any $2 \le p < \infty$, by Sobolev's imbedding theorem, $V(x,i) \in C^{1,\gamma}(\mathbb{R}^N \times \mathcal{S})$, for $0 < \gamma < 1$, γ arbitrarily close to 1, and hence by our assumptions on $\overline{b}$, $\overline{\lambda}$, $\overline{c}$, it is easy to see that

$$\alpha V(x,i) - \inf_{u\in U}\left\{\sum_{j=1}^{N} \overline{m}_j(x,i,u)\frac{\partial V(x,i)}{\partial x_j} + \sum_{j=1}^{M} \overline{\lambda}_{ij}(x,u)\big(V(x,j) - V(x,i)\big) + \overline{c}(x,i,u)\right\}$$

is in $C^{0,\gamma}$. By elliptic regularity [23, p. 287] applied to (6.1) (V replacing ψ), we conclude that $V \in C^{2,\gamma}$. ∎

Theorem 6.2. *A homogeneous Markov non-randomized policy v is optimal if and only if*

$$\begin{aligned} &\sum_{j=1}^{N} \overline{m}_j\big(x,i,v(x,i)\big)\frac{\partial V(x,i)}{\partial x_j} \\ &\qquad + \sum_{k=1}^{M} \overline{\lambda}_{ik}\big(x,v(x,i)\big)\big(V(x,k) - V(x,i)\big) + \overline{c}\big(x,i,v(x,i)\big) \\ &= \inf_{u\in U}\left\{\sum_{j=1}^{N} \overline{m}_j(x,i,u)\frac{\partial V(x,i)}{\partial x_j} + \sum_{k=1}^{M} \overline{\lambda}_{ik}(x,u)\big(V(x,k) - V(x,i)\big)\right. \\ &\qquad\qquad \left. + \overline{c}(x,i,u)\right\}, \quad \textit{a.e. } x \in \mathbb{R}^N,\ i \in \mathcal{S}\,. \end{aligned} \tag{6.4}$$

Proof. The 'necessity' part is contained in the proof of Theorem 6.1. We establish the sufficiency. Let $v(\cdot,\cdot)$ satisfy (6.4). The existence of such a v is guaranteed by a standard measurable selection theorem [4, Lemma 1]. Let v' be any other homogeneous Markov non-randomized policy. By standard arguments involving Ito's formula and the strong Markov property, we conclude that

$$J_v(x,i) \le J_{v'}(x,i), \qquad \text{a.e. } x \in \mathbb{R}^N,\ i \in \mathcal{S}\,.$$

Hence, by Lemma 4.2,

$$J_v(x,i) \le J_{\overline{v}}(x,i)$$

for any admissible policy $\overline{v}$. Thus, v is optimal. □

Remark 6.1. Thus far, we have assumed that the cost function $\overline{c}$ is bounded. However, this condition can be relaxed, as we show in the Appendix.

7. An Application to a Simplified Model. We consider a modified version of the model studied in [2]. Suppose there is one machine producing a single commodity. We assume that the demand rate is a constant $d > 0$. Let the machine state $S(t)$ take values in $\{0,1\}$, $S(t) = 0$ or 1, according as the machine is down or functional. Let $S(t)$ be a continuous time Markov chain with generator

$$\begin{bmatrix} -\lambda_0 & \lambda_0 \\ \lambda_1 & -\lambda_1 \end{bmatrix}.$$

The inventory $X(t)$ is governed by the Ito equation

$$dX(t) = (u(t) - d)\,dt + \sigma\,dW(t), \tag{7.1}$$

where $\sigma > 0$. The production rate $u(t)$ is constrained by

$$u(t) = \begin{cases} 0 & \text{if } S(t) = 0 \\ \in [0,R] & \text{if } S(t) = 1. \end{cases}$$

Let $c : \mathbb{R} \to \mathbb{R}_+$ be the cost function which is assumed to be convex and Lipschitz continuous. Let $\alpha > 0$ be the discount factor and let $V(x,i)$ denote the value function. In this case $V(x,i)$ is the minimal non-negative C^2 solution of the HJB equation

$$\begin{pmatrix} \frac{\sigma^2}{2}V''(x,0) - dV'(x,0) \\ \frac{\sigma^2}{2}V''(x,1) - \min\limits_{u\in[0,R]}\Big\{(u-d)V'(x,1)\Big\} \end{pmatrix} = \begin{bmatrix} \lambda_0 + \alpha & -\lambda_0 \\ \lambda_1 & \alpha - \lambda_1 \end{bmatrix} \begin{pmatrix} V(x,0) \\ V(x,1) \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} c(x)\,. \tag{7.2}$$

Using the convexity of $c(\cdot)$, it can be shown as in [2] that $V(\cdot,i)$ is convex for each i. Hence, there exists an x^* such that

$$\begin{aligned} V'(x,1) &\le 0 \quad \text{for } x \le x^* \\ &\ge 0 \quad \text{for } x \ge x^*. \end{aligned} \tag{7.3}$$

From (7.2), it follows that the value of u which minimizes $(u-d)V'(x,1)$ is

$$u = \begin{cases} R & \text{if } x \le x^* \\ 0 & \text{if } x \ge x^*. \end{cases}$$

At $x = x^*$, $V'(x^*, 1) = 0$ and therefore any $u \in [0, R]$ minimizes $(u-d)V'(x, 1)$. Thus, in view of Theorem 6.2, we can choose any $u \in [0, R]$ at $x = x^*$. To be specific, we choose $u = d$ at $x = x^*$. It follows that the following homogeneous Markov non-randomized policy is optimal

$$v(x, 0) = 0\,, \qquad v(x, 1) = \begin{cases} R & \text{if } x < x^* \\ d & \text{if } x = x^* \\ 0 & \text{if } x > x^*. \end{cases} \tag{7.4}$$

We note at this point that the piecewise deterministic model, in general, would lead to a singular control problem when $V'(x, 1) = 0$ [2], [27]. In [2] Akella and Kumar have obtained the solution of the HJB equation (this would be (7.2) without the second order term) in closed form and have computed an explicit expression for x^*. They have shown that a policy of the type (7.4) is optimal among all homogeneous Markov non-randomized policies. In our case, the additive noise in (7.1) induces a smoothing effect to remove the singular situation; in addition, our results imply that the policy (7.4) is optimal among *all admissible policies.* The only limitation of our model is that it would, in general, be very difficult to solve (7.2) analytically. Therefore, one must rely on numerical methods to compute an optimal policy of the type (7.4).

We now discuss the manufacturing model studied in [27] as described in the introduction. The machine state $S(t)$ is again a prescribed continuous time Markov chain taking values in $S = \{1, \ldots, M\}$. For each $i \in \mathcal{S}$, the production rate $u = (u_1, \ldots, u_N)$ takes values in U_i, a convex polyhedron in $\mathbb{R}^N$. The demand rate is $d = [d_1, \ldots, d_N]^T$. In this case, if the cost function $c : \mathbb{R}^N \to \mathbb{R}_+$ is Lipschitz continuous and convex, it can be shown that for each $i \in \mathcal{S}$, the value function $V(\cdot, i)$ is convex. But from this fact alone optimal policies of the type (7.4) cannot be obtained. However, since an optimal homogeneous Markov non-randomized policy $v(x, i)$ is determined by minimizing

$$\sum_{j=1}^{N} (u_j - d_j) \frac{\partial V(x, i)}{\partial x_j}$$

over U_i, $v(x, i)$ takes values at extreme points of U_i. Thus, for each machine state i, an optimal policy divides the buffer state space into a set of regions in which the production rate is constant. If the gradient $\nabla V(x, i)$ is zero or orthogonal to a face of U_i, a unique minimizing value does not exist. But again, in view of Theorem 6.2, we may prescribe arbitrary production rates at those points where $\nabla V(x, i) = 0$, and if $\nabla V(x, i)$ is orthogonal to a face of U_i, we can choose any corner of that face. Hence, once again, we can circumvent the singular situation.

8. Concluding Remarks. We have analyzed the optimal control of switching diffusions with a discounted criterion on the infinite horizon. The model allows a very general form of coupling between the continuous and the discrete components of the process. We have shown that there exists a homogeneous, non-randomized Markov policy which is optimal in the class of all admissible policies. Also, the existence of a unique solution in a certain class to the associated Hamilton-Jacobi-Bellman equations is established and the optimal policy is characterized as a minimizing selector of an appropriate Hamiltonian.

The primary motivation for this study is a class of control problems encountered in flexible manufacturing systems. By explicitly taking into account the noise present in the dynamics, we are able to remove singularities arising in the noiseless situation. In addition, we show that hedging type policies are optimal in a much wider class of non-anticipative policies than previously considered. We have confined our attention to the flow control level only. However, our results can be used to study control problems at other levels in hierarchical manufacturing systems [21], as well as control problems in other hybrid systems (see, e.g., [17], [38], [39]).

Here we have studied only the discounted criterion. Following [12], we can obtain similar results for the finite horizon and exit time criteria. However, the long-run average cost problem is more involved and is currently under study.

REFERENCES

[1] R. Akella, Y. Choong and S. B. Gershwin, *Performance of hierarchical production scheduling policy*, IEEE Trans. on Components, Hybrids and Manufacturing Technology, CHMT-7 (1984), pp. 225–240.

[2] R. Akella and P. R. Kumar, *Optimal control of production rate in a failure prone manufacturing system*, IEEE Trans. Automat. Control, AC-31 (1986), pp. 116–126.

[3] D. G. Aronson, *Bounds for the fundamental solution of a parabolic equation*, Bull. A.M.S., 73 (1967), pp. 890–896.

[4] V. E. Beneš, *Existence of optimal strategies based on specified information for a class of stochastic decision problems*, SIAM J. Control, 8 (1970), pp. 179–188.

[5] A. Bensoussan, *Stochastic Control by Functional Analysis Methods*, North-Holland, Amsterdam, 1982.

[6] A. Bensoussan and J. L. Lions, *Impulse Control and Quasi-Variational Inequalities*, Gauthier-Villars, Paris, 1984.

[7] A. Bensoussan and P. L. Lions, *Optimal control of random evolutions*, Stochastics, 5 (1981), pp. 169–199.

[8] A. Bensoussan, S. P. Sethi, R. Vickson and N. Derzko, *Stochastic production planning with production constraints*, SIAM J. Control Optim., 22 (1984), pp. 920–935.

[9] T. Bielecki and P. R. Kumar, *Optimality of zero-inventory policies for unreliable manufacturing systems*, Oper. Res., 36 (1988), pp. 532–546.

[10] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York, 1968.

[11] V. S. Borkar, *Optimal Control of Diffusion Processes*, Pitman Research Notes in Math. Series, No. 203, Longman, Harlow, UK, 1989.

[12] V. S. Borkar and M. K. Ghosh, *Controlled diffusions with constraints*, J. Math. Anal. Appl., 152 (1990), pp. 88–108.

[13] R. W. Brockett and G. L. Blankenship, *A representation theorem for linear differential equations with Markovian coefficients*, in Proc. of 1977 Allerton Conference on Circuits and Systems Theory, Urbana, Illinois, 1977.

[14] E. K. Boukas and A. Haurie, *Optimality conditions for continuous time systems with controlled jump Markov disturbances: application to an FMS planning problem*, in Analysis and Optimization of Systems (A. Bensoussan and J. L. Lions, eds.), Lecture Notes on Control and Information Sciences, Springer-Verlag, New York, 1988, pp. 633–676.

[15] ———, *Manufacturing flow control and preventive maintenance: a stochastic control approach*, IEEE Trans. Automat. Control, AC-35 (1990), pp. 1024–1031.

[16] E. K. Boukas, A. Haurie and P. Michel, *An optimal control problem with a random stopping time*, JOTA, 64 (1990), pp. 471–480.

[17] D. A. Castanon, M. Coderch, B. C. Levy, and A. S. Willsky, *Asymptotic analysis, approximation, and aggregation methods for stochastic hybrid systems*, in Proc. 1980 Joint Automatic Control Conference, San Francisco, CA, 1980.

[18] J. Chabrowski and N. A. Watson, *Properties of solutions of weakly coupled parabolic systems*, J. London Math. Soc., 23 (1981), pp. 475–495.

[19] M. H. A. Davis, *Stochastic Control and Nonlinear Filtering*, Tata Institute of Fundamental Research, Bombay, 1984.

[20] W. H. Fleming, S. P. Sethi and H. M. Soner, *An optimal stochastic production planning with randomly fluctuating demand*, SIAM J. Control Optim., 25 (1987), pp. 1494–1502.

[21] S. B. Gershwin, *Hierarchical flow control: a framework for scheduling and planning discrete events in manufacturing systems*, Proc. IEEE, 77 (1989), pp. 195–209.
[22] I. I. Gihman and A. V. Skorohod, *Controlled Stochastic Processes*, Springer-Verlag, New York, 1979.
[23] P. Grisvard, *Elliptic Problems in Non-Smooth Domains*, Pitman, Boston, 1965.
[24] N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes*, North-Holland, Amsterdam, 1981.
[25] J. Jacod and A. N. Shiryayev, *Limit Theorems for Stochastic Processes*, Springer-Verlag, New York, 1980.
[26] S. Kesavan, *Topics in Functional Analysis and Applications*, Wiley, New Delhi, 1989.
[27] J. Kimenia and S. B. Gershwin, *An algorithm for the computer control of a flexible manufacturing system*, IIE Trans., 15 (1983), pp. 353–362.
[28] N. V. Krylov, *Controlled Diffusion Processes*, Springer-Verlag, New York, 1980.
[29] H. J. Kushner, *Optimal discounted stochastic control for diffusion processes*, SIAM J. Control, 5 (1967), pp. 520–531.
[30] O. A. Ladyzhenskaya, V .A. Solonnikov and N. N. Ural'ceva, *Linear and Quasilinear Equations of Parabolic Type*, American Mathematical Society, Providence, RI, 1968.
[31] O. A. Ladyzhenskaya and N. N. Ural'ceva, *Linear and Quasilinear Elliptic Equations*, Academic Press, New York, 1968.
[32] J. Lehoczky, S. Sethi, H. M. Soner and M. Taksar, *An asymptotic analysis of hierarchical control of manufacturing systems under uncertainty*, Math. O.R., 16 (1991), pp. 596–608.
[33] G. J. Oldser and R. Suri, *Time optimal control of parts-routing in a manufacturing system with failure prone machines*, in Proc. 19th IEEE Conference on Decision and Control, Albuquerque, New Mexico, 1980, pp. 722–727.
[34] R. Phelps, *Lectures on Choquet's Theorem*, Van Nostrand, New York, 1966.
[35] R. Rishel, *Dynamic programming and minimum principles for systems with jump Markov disturbances*, SIAM J. Control, 13 (1975), pp. 338–371.
[36] S. Sethi and M. I. Taskar, *Deterministic equivalent for a continuous time linear-convex stochastic control problem*, JOTA, 64 (1990), pp. 169–181.
[37] A. Sharifnia, *Production control of a manufacturing system with multiple machine states*, IEEE Trans. Automat. Control, AC-33 (1988), pp. 620–625.
[38] D. D. Sworder, *Feedback control of a class of linear systems with jump parameters*, IEEE Trans. Automat. Control, AC-14 (1969), pp. 9–14.
[39] ———, *Control of systems subject to sudden change in character*, Proc. IEEE, 64 (1976), pp. 1219–1225.
[40] ———, *Control of a linear system with non-Markovian modal changes*, Journal of Economic Dynamics and Control, 2 (1980), pp. 233–240.
[41] A. Ju. Veretennikov, *On strong solutions of Ito stochastic equations with jumps*, Theory of Probability and Its Applications, 32 (1988), pp. 148–152.
[42] D. Vermes, *Optimal control of piecewise deterministic Markov processes*, Stochastics, 14 (1985), pp. 165–208.
[43] W. M. Wonham, *Random differential equations in control theory*, in Probabilistic Methods in Applied Mathematics (A. T. Bharucha-Reid, ed.), Academic Press, New York, 1970, pp. 131–212.

Appendix. Note that by the arguments of Section 5 we can establish the existence of a homogeneous Markov non-randomized policy for each fixed initial law. The independence of the optimal policy of the initial law results from the dynamic programming characterization of the optimal policy via Theorem 6.2. Using probabilistic arguments, dynamic programming equations can be derived by suitably adapting the approach in [11, Chap. 3]. However, we will present in a brief sketch an alternative analytical approach, which parallels that used for classical diffusions in [29].

We assume that for each $i \in \mathcal{S}$, $\bar{c}(\cdot, i, \cdot)$ is Lipschitz in its first argument uniformly with respect to the third. We further assume that for each $x \in \mathbb{R}^N$ and $i \in \mathcal{S}$, the value function $V(x, i) < \infty$. (This assumption may be replaced by some ergodicity hypotheses of the process under some homogeneous Markov policy.) Let $B_R = \{x \in \mathbb{R}^N : \|x\| < R\}$. Consider the Dirichlet problem on B_R

$$\begin{aligned} &\inf_{u\in U} L^u \varphi(x,i) = \alpha\varphi(x,i)\,, \qquad \text{in } B_R\times\mathcal{S} \\ &\varphi(x,i)\Big|_{\partial B_R} = 0\,. \end{aligned} \tag{A.1}$$

The existence of a unique solution $\varphi_R(x,i)$ of (A.1) in $W^{2,p}_{\ell oc}(\mathbb{R}^N\times\mathcal{S})$, $2\le p<\infty$, is guaranteed by [31, Thm. 5.1, p. 422]. Thus, to each $R>0$ there corresponds a solution φ_R to (A.1) belonging to $W^{2,p}_{\ell oc}(\mathbb{R}^N\times\mathcal{S})$, for $2\le p<\infty$. Using elliptic regularity results as in Corollary 6.1, it follows that $\varphi_R(x,i)\in C^{2,\gamma}(B_R\times\mathcal{S})$, $0<\gamma<1$, γ arbitrarily close to 1. Let v_R be a homogeneous Markov non-randomized policy which is a minimizing selector in (6.4). Standard arguments involving Ito's formula yield

$$\begin{aligned} \varphi_R(x,i) &= E\biggl[\int_0^{\tau_R} e^{-\alpha t}\bar{c}\bigl(X(t),S(t),v_R(X(t),S(t))\bigr)\,dt \,\Bigm|\, X(0)=x, S(0)=i\biggr] \\ &= \inf_{u(\cdot)} E\biggl[\int_0^{\tau_R} e^{-\alpha t}\bar{c}\bigl(X(t),S(t),u(t)\bigr)\,dt \,\Bigm|\, X(0)=x, S(0)=i\biggr], \end{aligned} \tag{A.2}$$

where τ_R is the hitting time of ∂B_R of the process $X(\cdot)$. Clearly $\varphi_R(x,i)\le V(x,i)$ and it can be easily seen from (A.2) that $\varphi_R(x,i)$ is increasing in R. Let $R'>R$. Then, by the interior estimates [31, pp. 398–402], $\{\varphi_{R'}\}_{R'>R}$ is bounded in B_R uniformly in R' and $\{\nabla\varphi_{R'}\}_{R'>R}$ is bounded in $W^{1,2}(B_R\times\mathcal{S})$ uniformly in R'. By Sobolev's imbedding theorem, $W^{1,2}(B_R\times\mathcal{S})\hookrightarrow L^{2+\varepsilon}(B_R\times\mathcal{S})$, for some $\varepsilon>0$. Then, by suitably modifying (4.10) of [31, p. 400], we obtain

$$\|\varphi_{R'}\|_{W^{2,2+\varepsilon}(B_R\times\mathcal{S})}\le K_R\,,$$

where K_R is a constant which does not depend on R'. (The modification is needed because of the factor $\varepsilon>0$, but it is routine.) Repeating the above procedure over and over again, we conclude that $\{\varphi_{R'}\}_{R'>R}$ is uniformly bounded in $W^{2,p}(B_R)$, for $2\le p<\infty$. Since $W^{2,p}(B_R)\hookrightarrow W^{1,p}(B_R)$ and the injection is compact, it follows that $\{\varphi_R\}$ converges strongly in $W^{1,p}(B_R)$. Thus, given any sequence $\{R_n\}$, $R_n\to\infty$, as $n\to\infty$ and for any fixed integer $N\ge2$, we can choose a subsequence $\{R_{n_i}\}$ such that $\{\varphi_{R_{n_i}}\}$ converges strongly in $W^{1,p}(B_{N-1})$. Using a suitable diagonalization, we may assume that $\{\varphi_{R_{n_i}}\}$ converges strongly in $W^{1,p}(B_{N-1})$ for each integer $N\ge2$. Let ψ be a limit point of $\{\varphi_{R_{n_i}}\}$. It can be shown as in [5, p. 148] (see also [31, p. 420]) that

$$\begin{aligned} &\inf_{u\in U}\Biggl\{\sum_{k=1}^N \bar{b}_k(x,j,u)\frac{\partial\varphi_{R_{n_i}}(x,j)}{\partial x_k} + \sum_{\ell=1}^M \bar{\lambda}_{j\ell}(x,u)\bigl(\varphi_{R_{n_i}}(x,\ell)-\varphi_{R_{n_i}}(x,j)\bigr) + \bar{c}(x,j,u)\Biggr\} \\ &\qquad \underset{n_i\to\infty}{\longrightarrow} \inf_{u\in U}\Biggl\{\sum_{k=1}^N \bar{b}_k(x,j,u)\frac{\partial\psi(x,j)}{\partial x_k} + \sum_{\ell=1}^M \bar{\lambda}_{j\ell}(x,u)\bigl(\psi(x,\ell)-\psi(x,j)\bigr) + \bar{c}(x,j,u)\Biggr\} \end{aligned}$$

strongly in $L^p(B_{N-1})$. Therefore, $\psi\in W^{1,p}_{\ell oc}(\mathbb{R}^N\times\mathcal{S})$ and ψ satisfies

$$\inf_{u\in U} L^u\psi(x,i) = \alpha\psi(x,i)$$

in $\mathcal{D}'(\mathbb{R}^N \times \mathcal{S})$, i.e., in the sense of distributions. By elliptic regularity, $\psi \in W^{2,p}_{\ell oc}(\mathbb{R}^N \times \mathcal{S})$, $2 \leq p < \infty$. Therefore, as in Corollary 6.1, it follows that $\psi \in C^{2,\gamma}(\mathbb{R}^N \times \mathcal{S})$, $0 < \gamma < 1$, γ arbitrarily close to 1. Let v be a minimizing selector corresponding to ψ. Then, by standard arguments involving Ito's formula, it can be shown that

$$\begin{aligned}\psi(x,i) &= E\left[\int_0^\infty e^{-\alpha t}\overline{c}\big(X(t), S(t), v(X(t), S(t))\big)\, dt \;\Big|\; X(0) = x, S(0) = i\right] \\ &= \inf_{u(\cdot)} E\left[\int_0^\infty e^{-\alpha t}\overline{c}\big(X(t), S(t), u(t)\big)\, dt \;\Big|\; X(0) = x, S(0) = i\right].\end{aligned}$$

Thus, $\psi(x,i) = V(x,i)$. In this situation, (6.1) does not have a unique solution in general, but $V(x,i)$ can be identified as a minimal nonnegative solution of (6.1) in $C^2(\mathbb{R}^N \times \mathcal{S})$. The assertion of Theorem 6.2 is also valid in this case.