DISCOUNTED STOCHASTIC FLUID PROGRAMS

Nicole Bäuerle
Department of Mathematics VII
University of Ulm
D-89069 Ulm, Germany
e-mail: baeuerle@mathematik.uni-ulm.de

Abstract

We consider optimal control problems for stochastic fluid models of the following type: suppose (Z_t) is a continuous-time Markov chain with finite state space. As long as $Z_t = z$, the dynamics of the system at time t are given by a function $b^z(u(\cdot))$, where u is a control we have to choose. A cost rate function c is given, depending on the state and the action. We want to control the system in such a way as to minimize the expected discounted cost over an infinite horizon. We will call a problem of this type a *Stochastic Fluid Program* (SFP). They typically appear in production and telecommunication systems. We formulate the optimization problem as a discrete time Markov decision process and give conditions under which an optimal stationary policy exists. Furthermore, we show how to solve SFPs numerically, using Kushner's approximating Markov chain approach. Last but not least we apply our results to a multi-product manufacturing system without backlog.

Key words: Stochastic fluid models; Piecewise deterministic Markov process; Markov decision process; Maximum principle; Approximating Markov chain approach

1 Introduction

In manufacturing and telecommunication systems we often encounter the situation that there are different timescales for the occurrence of events. For example, if we allow for random breakdowns of machines in manufacturing models, we typically assume that the production process itself is much faster than the breakdowns of machines (cf. Sethi/Zhang (1994)). Another example is the Anick/Mitra/Sondhi-model (1982). There, the authors suppose that the cell stream sources in ATM multiplexers are on-off sources. Thus, we have a certain cell transmission when the source is on and no transmission when the source is off. The durations of the state lengths are random. In both cases we obtain adequate models when we replace quantities that vary faster with their averages, whereas we keep the slower processes stochastic. Formulations of this type are commonly used and important in stochastic modeling. We now want to give a unified approach towards the optimal control of such systems which we will call *Stochastic Fluid Programs*. An informal description of the evolution of stochastic fluid programs is the following: Suppose $S \subset I\!R^N$ is the *state space* of the system and $y \in S$ the initial state. The local dynamics of the system are determined by an external *environment process* (Z_t) which we assume to be a continuous-time Markov chain with finite state space Z and generator Q (this assumption can be relaxed to (Z_t) being a semi-Markov process). Whenever $Z_t = z$, the system evolves according to $y_t = y + \int_0^t b^z(u(y,z,s))\, ds$, where $u : S \times Z \times I\!R_+ \to U \subset I\!R^K$ is a control and b^z is a given measurable function $b^z : U \to S$. U is our *action space*. Moreover, a *cost rate function* $c : S \times Z \times U \to I\!R_+$ and an *interest rate* $\beta \geq 0$ are given. The 6-tuple $(E = S \times Z, U, b, Q, c, \beta)$ will be called a Stochastic Fluid Program (SFP). We are interested in minimizing the β-discounted cost of the system over an infinite horizon for $\beta > 0$.

Let us first look at the following *example* of a multi-product manufacturing system without backlog. Each of N parallel machines produces a different item for consumption. The demand rate for item i is $\mu_i > 0$ for $i = 1, \dots, N$. Since the machines are subject to random breakdown and repair, the total production capacity $\lambda_j(z) \in I\!R_+$ for items of type j depends on the number $z = Z_t$ of working machines at time t. Z_t is our environment process. The vector $Y_t = (Y_1(t), \dots, Y_N(t))$ gives the inventory of each product at time

t and we assume $S = I\!R_+^N$. We obtain a reward rate r_j for each unit of item j which is produced, but incur an inventory holding cost at rate $\hat{c}(y)$ which depends on the joint inventory $y \in S$. We have to decide now upon the production rate of each item and to be able to get a non-negative inventory, we have to control the demand rate. Hence we define $(u, v) \in U = [0,1]^N \times [0,1]^N$, where $\lambda_j(z)u_j$ is the production rate and $v_j\mu_j$ is the demand rate for items of type j, $j = 1, \ldots, N$. For $u \in U, z \in Z$ the local dynamics of the system are given by $b_j^z(u) = \lambda(z)u_j - v_j\mu_j, j = 1, \ldots, N$. Hence,

$$\begin{aligned}
&E = I\!R_+^N \times Z \\
&U = [0,1]^N \times [0,1]^N \\
&b_j^z(u) = \lambda(z)u_j - \mu_j v_j,\ j = 1, \ldots, N \\
&c(x) = \hat{c}(y) - \sum_{j=1}^{N} r_j \lambda_j(z) u_j
\end{aligned}$$

together with an interest rate β and generator Q of the environment process specifies our problem. It is important to note that the control variable v_j is introduced only in order to be able to keep the inventory non-negative. For a realistic cost rate function which is increasing in the inventory we obtain for an optimal control $v_j = 1$, whenever $y_j > 0$.
By (Y_t) we denote the stochastic process of the buffer contents and by $(X_t) = (Y_t, Z_t)$ the joint state process. $x \in E$ should always be understood as $x = (y, z)$. At the jump times (T_n) of the environment process (Z_t), decisions have to be taken in form of a control $u : E \times [0, \infty) \to U$ and $\phi_t(x, u) := y + \int_0^t b^z(u(x, s))\, ds$ gives the state of the system at time t under control u, starting in x. According to Yushkevich (1980) we can w.l.o.g. restrict to decisions taken at jumps only (cf. also Remark 6a)). u is called admissible if $\phi_t(x, u) \in S$ for all $t \geq 0$ and a sequence $\pi = (u_n)$ of admissible u_n defines a policy. Hence we have $Y_t = \phi_{t-T_n}(X_{T_n}, u_n)$ for $T_n \leq t < T_{n+1}$ and $\pi_t = u_n(X_{T_n}, t - T_n)$. The optimization problem is

$$V(x) = \inf_\pi V_\pi(x) = \inf_\pi E_x^\pi \left[\int_0^\infty e^{-\beta t} c(X_t, \pi_t)\, dt \right],$$

where the infimum is taken over all policies. Thus SFPs are a special class of *piecewise deterministic Markov processes* (see e.g. Davis (1993)) with one exception: in our model we allow for constraints on the actions and the process can move along the boundary of the state space. In the classical theory for piecewise deterministic Markov processes the state process automatically jumps back into the interior of the state space as soon as it has reached the boundary. However, as far as applications in telecommunications and manufacturing systems are concerned, the state process naturally moves along boundaries such as non-negativity constraints. This boundary behavior in particular makes it difficult to determine an optimal control. In the literature one can find examples of SFP which have been solved explicitly, see e.g. Akella/Kumar (1986), Presman et al. (1995), Rajagopal et al. (1995). Related models are Markov decision drift processes (cf. Hordijk/Van der Duyn Schouten (1983)) and the more specific semi-Markov decision processes. In contrast to our model, Markov decision drift processes and semi-Markov decision processes do not control process movement between transitions while in SFPs the decision effects the deterministic behavior between transitions. The jumps themselves cannot be controlled in SFPs. Consequently we will use numerous results from piecewise deterministic Markov processes and accommodate them to our constrained problem. In particular we will exploit the fact that the optimization problem can be reduced to a discrete-time *Markov decision process*. To prevent the optimality of randomized controls, we will make several convexity assumptions. For our applications this is no crucial restriction. After defining the mathematical model and its discrete time reduction rigorously in section 2 - 4, we will prove in section 5 under some continuity and compactness assumptions that an optimal stationary policy exists which is the solution of a deterministic control problem (Theorem 4). Section 6 summarizes properties of the value function which are helpful in applications. In section 7 we comment shortly on how to solve SFPs numerically and in section 8 we apply our results to a multi-product manufacturing system.

2 Continuous-time Stochastic Fluid Program

We will first give a definition of a *Stochastic Fluid Program* in continuous time and make some basic assumptions about our model which will be valid throughout the paper without further mentioning them. Let Z be a finite set and Q a generator for a Markov chain on Z. We assume that $Q = (q_{zz'})$ defines an *irreducible* Markov chain. As usual denote $q_z :=$

$-q_{zz}$ for $z \in Z$. Let $S \subset I\!R^N$ and define by $\mathfrak{B}(S)$ the Borel-σ-algebra on S. $E := S \times Z$ is called *state space* of the system. A state $x \in E$ is denoted by $x = (y, z)$. $U \subset I\!R^K$ is the *action space* of the system. For all $z \in Z$, measurable functions $b^z : U \to I\!R^N$ are given, the so-called *dynamics* of the system. We will write $b : Z \times U \to I\!R^N$ to summarize all b^z. A measurable function $u : E \times [0, \infty) \to U$ is called an *open-loop control*. Define

$$\phi_t(x, u) := y + \int_0^t b^z(u(x, s))\, ds.$$

$\phi_t(x, u)$ gives the state of the system at time t under control u, starting in state x. u is called *admissible* if $\phi_t(x, u) \in S$ for all $t \geq 0$. Let $\pi = (u_n)$ be a sequence of controls, where all u_n are admissible. In this case we will call π a *policy*. When we denote by (T_n), $T_0 = 0$ the jump times of the environment process (Z_t), then $u_n(X_{T_n}, t - T_n)$ is the control which has to be applied for t in the time interval $[T_n, T_{n+1})$. Moreover, we are given a measurable *cost rate function* $c : E \times U \to I\!R_+$ and an interest rate $\beta > 0$. These objects together will define our program. Instead of $c \geq 0$ it is sufficient to assume that c is bounded below. By $1_B(\cdot)$ we denote the indicator function of set B.

Definition 1:

The 6-tuple (E, U, b, Q, c, β) is called a (discounted) *Stochastic Fluid Program* (SFP).

For a fixed policy π, there exists a family of probability measures $\{P_x^\pi \mid x \in E\}$ on a measurable space $(\Omega, \mathcal{F})$ and stochastic processes $(X_t) = (Y_t, Z_t)$ and (π_t) such that for $0 := T_0 < T_1 < T_2 < \dots$

$$Z_t = Z_{T_n}, \quad Y_t = \phi_{t-T_n}(X_{T_n}, u_n), \quad \pi_t = u_n(X_{T_n}, t - T_n) \ \text{ for } \ T_n \leq t < T_{n+1}$$

and

(i) $P_x^\pi(X_0 = x) = P_x^\pi(T_0 = 0) = 1$ for all $x \in E$.

(ii) $P_x^\pi(T_{n+1} - T_n > t \mid T_0, X_{T_0}, \dots, T_n, X_{T_n}) = e^{-q_{Z_{T_n}} t}$.

(iii) $P_x^\pi(X_{T_{n+1}} \in B \times \{z'\} \mid T_0, X_{T_0}, \dots, X_{T_n}, T_{n+1}) = \frac{q_{Z_{T_n} z'}}{q_{Z_{T_n}}} 1_B\left(\phi_{T_{n+1}-T_n}(X_{T_n}, u_n)\right)$ for $z' \in Z$, $z' \neq Z_{T_n}$ and $B \in \mathfrak{B}(S)$ and zero, if $z' = Z_{T_n}$.

The process $(X_t) = (Y_t, Z_t)$ will be called *state process*. Obviously (Z_t) is a continuous-time Markov chain with generator Q and jump times (T_n). The optimization problem we are interested in is the following:

Definition 2:

Let π be a policy. For $x \in E$ define

a) the *expected discounted cost over an infinite horizon under policy* π*, starting in* x by

$$V_\pi(x) := E_x^\pi \left[\int_0^\infty e^{-\beta t} c(X_t, \pi_t)\, dt \right]$$

b) the *minimal expected discounted cost over an infinite horizon, starting in* x by

$$V(x) := \inf_\pi V_\pi(x).$$

c) π is called *optimal*, if it attains the infimum in b) for all $x \in E$.

Remark 1:

a) Since the jump times of (Z_t) cannot be controlled, it is easily possible to define for fixed $x \in E$ a common probability measure P_x on a measurable space $(\Omega', \mathcal{F}')$ such that for all policies π there exist processes $(X_t^\pi) = (Y_t^\pi, Z_t)$ such that $P_x(X_t^\pi \in \cdot) = P_x^\pi(X_t \in \cdot)$. This observation is useful for sample path arguments.

b) When we have only one environment state, i.e. $|Z| = 1$, then the problem reduces to a purely deterministic control problem.

3 Discrete-time Stochastic Fluid Program

We will now show that the optimization problem in Definition 1 can be transferred into an equivalent *discrete-time dynamic program* with substochastic transition kernel. Exploiting this fact, it is (in principle) possible to apply the theory of Markov decision processes.

The point is that the evolution between jumps of the environment process is purely deterministic. This enables us to choose at the jump time points of the environment process a control which is a function of the time only and which is applied until the next jump occurs.

Suppose a SFP (E, U, b, Q, c, β) as defined in the previous section is given. On U we assume to have the usual Borel σ-algebra. Denote by $A := \{a : I\!\!R_+ \to U \mid a \text{ measurable}\}$ the *action space* and for $x \in E$ by

$$D(x) := \{a \in A \mid \phi_t(x,a) = y + \int_0^t b^z(a_s)\, ds \in S, \forall t \geq 0\}$$

the set of *admissible actions*. If no further jump has occurred, a_t is the action which is applied t time units after the last jump. Note that the action is now a function of the time only. We assume that $D(x) \neq \emptyset$ for all $x \in E$ and define $D := \{(x,a) \mid a \in D(x)\}$. Furthermore, let the transition kernel $p : D \times \mathfrak{B}(S) \times Z \to [0,1]$ be defined by

$$p(x,a;B \times \{z'\}) := \begin{cases} q_{zz'} \int\limits_0^\infty e^{-(\beta+q_z)t} 1_B\left(\phi_t(x,a)\right)\, dt, & \text{if } z' \neq z \\ 0 & \text{if } z' = z \end{cases}$$

and the *one-step cost function* $C : D \to \overline{I\!\!R}_+$ by

$$C(x,a) := \int_0^\infty e^{-(\beta+q_z)t} c\left(\phi_t(x,a), z, a_t\right)\, dt.$$

p is obviously a substochastic transition kernel. Note that $p(x,a;B \times \{z'\})$ is exactly the discounted probability of getting from state x at jump time T_n to a state in the set $B \times \{z'\}$ at the next jump time point T_{n+1} under the control a. C is the cost that is incurred during such a period from one jump to another. Theorem 1 below gives the justification for these definitions. A σ-algebra on A will be defined in section 4. $F := \{f : E \to A \mid f \text{ measurable, } f(x) \in D(x)\}$ is called the set of *decision rules* and $\sigma = (f_n)$, where $f_n \in F$ is called a *policy* in the discrete case. After adding an absorbing state to make the transition kernel stochastic, we obtain for a fixed policy σ that there exists a family of probability measures $\{\hat{P}_x^\sigma \mid x \in E\}$ on a measurable space $(\hat{\Omega}, \hat{\mathcal{F}})$ and a discrete-time stochastic process $(X_n) = (Y_n, Z_n)$ on $(\hat{\Omega}, \hat{\mathcal{F}})$ such that

(i) $\hat{P}_x^\sigma(X_0 = x) = 1$ for all $x \in E$.

(ii) $\hat{P}_x^\sigma(X_{n+1} \in B \times \{z'\} \mid X_0, \dots, X_n) = p(X_n, f_n(X_n); B \times \{z'\})$ for all $z' \in Z$ and $B \in \mathfrak{B}(S)$.

Remark 2:

a) It is important to point out that the Markov chain (X_n) as previously defined and the process (X_t) as defined in Section 2 are two different objects, as well as the corresponding policies. It should always be clear from the context, whether the continuous or the discrete version is considered and the notation should not lead to any confusion.

b) By $\hat{E}_x^\sigma$ we denote the expectation w.r.t. the probability measure $\hat{P}_x^\pi$.

Definition 3:

The 6-tuple (E, A, D, p, C, β) is called the *Discretized Stochastic Fluid Program* (DSFP).

Remark 3:

To obtain the connection with the continuous-time definition it is important to note that whenever $\pi = (u_n)$ is a policy for the SFP, $\sigma = (f_n)$, where $f_n(x)(t) = u_n(x,t)$ is a policy for the DSFP and vice versa. This result is not trivial since f_n and u_n have different measurability requirements.

Theorem 1:

Let π be a policy for the SFP and σ the corresponding policy for the DSFP. Then we obtain

a) $V_\pi(x) = \hat{E}_x^\sigma \left[\sum_{n=0}^{\infty} C\left(X_n, f_n(X_n)\right) \right]$

b) $V(x) = \inf_\sigma \hat{E}_x^\sigma \left[\sum_{n=0}^{\infty} C\left(X_n, f_n(X_n)\right) \right]$

A proof of Theorem 1 can be found in the appendix. For further investigations it is convenient to define the following operators. If $v : E \to I\!R_+$ we denote the operator $\mathcal{U}$ by

$$\mathcal{U}v(x) := \inf_{a \in D(x)} \left[C(x,a) + \int_0^\infty e^{-(\beta+q_z)t} \sum_{z' \neq z} q_{zz'} v\left(\phi_t(x,a), z'\right) \, dt \right].$$

For $f \in F$ we will use the following notation

$$\mathcal{U}_f v(x) := C(x, f(x)) + \int_0^\infty e^{-(\beta+q_z)t} \sum_{z' \neq z} q_{zz'} v\left(\phi_t(x, f(x)), z'\right) \, dt.$$

$f \in F$ will be called *minimizer* of v if f attains the infimum in $\mathcal{U}v$. Note that the optimization problem given by operator $\mathcal{U}$ is a deterministic control problem.

Remark 4:

a) Let $\sigma = (f_n)$ be a policy for the DSFP. Then we have $V_\sigma = \lim_{n\to\infty} U_{f_0} \ldots U_{f_n} 0$. The proof is similar to the one for Theorem 1.

b) It is easily seen that both operators $\mathcal{U}_f$ and $\mathcal{U}$ are monotone, i.e. if we have $v, w : E \to I\!R_+$ with $v \leq w$ then $\mathcal{U}_f v \leq \mathcal{U}_f w$ and $\mathcal{U}v \leq \mathcal{U}w$.

c) Since $\mathcal{U}$ is a mapping from E to $I\!R_+$ we can again apply the operator $\mathcal{U}$ to the result $\mathcal{U}v$. An n-times iterated application of $\mathcal{U}$ is denoted by $\mathcal{U}^n v$.

The next aim will be to show the existence of optimal policies for the DSFP. In order to do this, we have to establish compactness and continuity properties. However, this causes some difficulties since we have to find a topology on A which guarantees that A is compact and that the right-hand side of the operator $\mathcal{U}$ is lower semicontinuous. Moreover, we have not even yet defined a σ-algebra on A. The usual way to cope with this problem is to pass over to randomized actions or so-called relaxed controls. Relaxed controls have first been introduced by Young for problems from the calculus of variations (cf. Kushner/Dupuis (1992) chapter 9.5). The action space can then be shown to be compact w.r.t. the *Young topology* (cf. Davis (1993)). This procedure will be explained briefly in the next section. However, from a practical point of view we do not want to deal with randomized actions. Thus, we will impose certain convexity assumptions which are quite natural and which allow for the minimum to be taken in the smaller set A of deterministic actions. The assumptions we need now are the following.

Assumption 1:

(i) S is closed and U is convex and compact w.r.t. the usual Euclidian norm.

(ii) $u \mapsto b^z(u)$ is linear for all $z \in Z$.

(iii) c is lower semicontinuous on $E \times U$ and $u \mapsto c(x,u)$ is convex for all $x \in E$.

4 A Relaxed Problem

As indicated in the last section we will relax our DSFP by considering randomized actions. Denote by $\mathbb{P}(U)$ the set of all probability measures on U. Then we denote

$$\mathcal{R} := \{r : \mathbb{R}_+ \to \mathbb{P}(U) \mid r \text{ measurable}\}.$$

Thus r_t is now a probability measure which gives the probability with which actions are taken at time t. Let a DSFP be given. For $r \in \mathcal{R},\ x \in E,\ B \in \mathfrak{B}(S),\ z' \in Z$ we define

$$\begin{aligned}
\tilde{\phi}_t(x,r) &:= y + \int_0^t \int_U b^z(u) r_s(du)\, ds \\
\tilde{C}(x,r) &:= \int_0^\infty e^{-(\beta+q_z)t} \int_U c(\tilde{\phi}_t(x,r), z, u) r_t(du)\, dt \\
\tilde{p}(x,r; B \times \{z'\}) &:= \begin{cases} q_{zz'} \int\limits_0^\infty e^{-(\beta+q_z)t} 1_B(\tilde{\phi}_t(x,r))\, dt, & \text{if } z \neq z' \\ 0 & \text{if } z = z' \end{cases} \\
\tilde{D}(x) &:= \{r \in \mathcal{R} \mid \tilde{\phi}_t(x,r) \in S,\ \forall t \geq 0\} \\
\tilde{D} &:= \{(x,r) \mid r \in \tilde{D}(x)\}
\end{aligned}$$

The relaxed DSFP is given by the previously defined quantities $(E, \mathcal{R}, \tilde{D}, \tilde{p}, \tilde{C}, \beta)$.

Remark 5:

a) As usual in $\mathcal{L}^p$-spaces, r should be thought of as an element of the λ^1-equivalence class.

b) $\mathbb{P}(U)$ is endowed with the Borel-σ-algebra which is induced by the weak topology.

c) $A \subset \mathcal{R}$ since the elements of A can be interpreted as the one-point measures in $\mathcal{R}$. Thus, we have in particular if $r = \delta_a$, i.e. the one-point measure on action $a \in A$, then $\tilde{\phi}_t(x, r) = \phi_t(x, a), \tilde{c}(x, r) = c(x, a)$ and $\tilde{p}(x, r; B \times \{z'\}) = p(x, a; B \times \{z'\})$.

It is possible to show that $\mathcal{R}$ is compact w.r.t. the Young-topology and $\mathcal{R}$ is metrizable. For a definition of the Young-topology and a proof of these results we refer the reader to Davis (1993) Section 4.3. The following Lemma will now be crucial.

Lemma 2:

Let a relaxed DSFP $(E, \mathcal{R}, \tilde{D}, \tilde{p}, \tilde{C}, \beta)$ be given. Under Assumption 1 it holds that

a) The mapping $(x, r) \mapsto \tilde{\phi}_t(x, r)$ is continuous for all $t \geq 0$.

b) $\tilde{D}(x)$ is compact for all $x \in E$ and $\tilde{D}$ is closed.

c) The mapping $(x, r) \mapsto \tilde{C}(x, r)$ is lower semicontinuous and $\tilde{C} \geq 0$.

d) $\tilde{p}$ is weakly continuous, i.e. $(x, r) \mapsto \int v(x')\tilde{p}(x, r; dx')$ is continuous and bounded for every continuous, bounded function $v : E \to I\!R$.

e) The set-valued mapping $x \mapsto \tilde{D}(x)$ is upper semicontinuous.

Proof:

a) See e.g. Davis (1993) Theorem 43.5.

b) Fix $x \in E$. We have

$$\tilde{D}(x) = \{r \in \mathcal{R} \mid \tilde{\phi}_t(x, r) \in S \ \forall t \geq 0\} = \cap_{t \geq 0}\{r \in \mathcal{R} \mid \tilde{\phi}_t(x, r) \in S\}.$$

Since S is closed and $\tilde{\phi}_t(x, r)$ is continuous in r for all x and t, $\{r \in \mathcal{R} \mid \tilde{\phi}_t(x, r) \in S\}$ is closed. Hence $\tilde{D}(x)$ is closed as the intersection of closed sets and since $\tilde{D}(x) \subset \mathcal{R}$ it is compact. Analogously we can write $\tilde{D} = \cap_{t \geq 0}\{(x, r) \mid \tilde{\phi}_t(x, r) \in S\}$ and since $(x, r) \mapsto \tilde{\phi}_t(x, r)$ is continuous for all $t \geq 0$ we obtain that D is closed.

c) and d) see e.g. Davis (1993) Theorem 44.11.

e) Define the mapping $\psi : E \to \tilde{D}$ by $\psi(x) = \tilde{D}(x)$. Let $B \subset \mathcal{R}$ be closed (since $\mathcal{R}$ is compact, B is also compact). We have to show that

$$\psi^{-1}[B] := \{x \in E \mid \tilde{D}(x) \cap B \neq \emptyset\}$$

is again closed. Let $x_n \in \psi^{-1}[B]$ with $x_n \to x$. Choose $r_n \in \mathcal{R}, n \in I\!N$ such that $r_n \in \tilde{D}(x_n) \cap B \subset B$. Since B is compact there exists a convergent subsequence $r_{n_k} \to r \in B$ for $k \to \infty$. Because of the closedness of $\tilde{D}$ it holds that $(x_{n_k}, r_{n_k}) \to (x, r) \in \tilde{D}$. This implies $x \in \psi^{-1}[B]$. □

For $v \in \mathfrak{C}_{lsc} := \{v : E \to I\!R_+ \mid v \text{ is lower semicontinuous}\}$ define the operator $\mathcal{T}$ for the relaxed problem as

$$\mathcal{T}v(x) = \inf_{r \in \tilde{D}(x)} \left[\tilde{C}(x,r) + \int_0^\infty e^{-(\beta+q_z)t} \sum_{z' \neq z} q_{zz'} v\left(\tilde{\phi}_t(x,r), z'\right) \, dt \right].$$

Theorem 3:
Let a DSFP be given and $v \in \mathfrak{C}_{lsc}$. Under Assumption 1 we have $\mathcal{U}v \in \mathfrak{C}_{lsc}$ and there exists an $f^* \in F$ such that

$$\mathcal{U}_{f^*}v = \mathcal{U}v = \mathcal{T}v.$$

Proof: Consider the relaxed DSFP. Due to our assumptions and using Proposition 7.31 in Bertsekas/Shreve (1978) (which also holds for substochastic transition kernels) we can apply the measurable selection Theorem given in Hernández-Lerma/Lasserre (1996) (Proposition D.5) to show that there exists a measurable $g : E \to \mathcal{R}$ with $g(x) \in \tilde{D}(x)$ for all $x \in E$ which attains the infimum in $\mathcal{T}v$ and $\mathcal{T}v \in \mathfrak{C}_{lsc}$. Since $A \subset \mathcal{R}$ implies $\mathcal{U}v \geq \mathcal{T}v$, it is now enough to show that there exists an $f^* \in F$ with $\mathcal{U}_{f^*}v = \mathcal{U}v \leq \mathcal{T}v$.
For $r \in \mathcal{R}$ define $a_t = \int_U u r_t(du)$, $t \geq 0$. Since U is convex, $a_t \in U$ for all $t \geq 0$ (see e.g. Hinderer (1984) Theorem 25.10) and it is measurable, hence $a \in A$. Moreover, since b^z is linear

$$\tilde{\phi}_t(x,r) = y + \int_0^t \int_U b^z(u) r_s(du) = y + \int_0^t b^z\Big(\int_U u r_s(du)\Big)\, ds = \phi_t(x,a)$$

which implies in particular that $a \in D(x)$. Using the convexity of c in the last component we obtain with the Jensen inequality

$$\begin{aligned}\tilde{C}(x,r) &= \int_0^\infty e^{-(\beta+q_z)t} \int_U c(\tilde{\phi}_t(x,r), z, u) r_t(du)\, dt \\ &\geq \int_0^\infty e^{-(\beta+q_z)t} c(\phi_t(x,a), z, \int_U u r_t(du))\, dt = C(x,a).\end{aligned}$$

Now we define for all $x \in E$ and $t \geq 0$

$$f^*(x)(t) = \int_U u g(x)(t, du).$$

Then $f^* : E \to A$ is measurable and $f^*(x) \in D(x)$. Moreover, for fixed $x \in E$ we obtain $\mathcal{T}v = \mathcal{T}_g v \geq U_{f^*} v \geq \mathcal{U}v$ which implies $\mathcal{T}v = \mathcal{U}v$ and the proof is complete. □

5 β-Discounted Cost Optimality Equation

The following assumption is needed to state our main theorem.

Assumption 2: There exists a policy π such that $V_\pi(x) < \infty$ for all $x \in E$.

Theorem 4: *(β-Discounted cost optimality equation)*
Suppose that Assumptions 1 and 2 hold. Then

a) V is the minimal solution of the *β-discounted cost optimality equation* $V = \mathcal{U}V$, i.e. for all $x \in E$

$$V(x) = \min_{a \in D(x)} \left[C(x,a) + \int_0^\infty e^{-(\beta+q_z)t} \sum_{z' \neq z} q_{zz'} V\left(\phi_t(x,a), z'\right)\, dt \right]. \qquad (1)$$

b) There exists a minimizer $f^* \in F$ of V in (1) and the stationary policy $(f^*, f^*, \ldots)$ is optimal.

The proof of part a) and b) follows essentially as in Hernández-Lerma/Lasserre (1996).

Proof: a),b) Since $0 \le C$ we obtain immediately for all $x \in E$

$$0 \le V_n := \mathcal{U}^n 0 \le V$$

and since the operator $\mathcal{U}$ is monotone we have $V_n \uparrow \hat{V} \le V$. From Lemma 4.2.4 in Hernández-Lerma/Lasserre (1996) (interchange of min and lim) together with Theorem 3 and the monotone convergence Theorem it follows that

$$\hat{V} := \lim_{n\to\infty} V_n = \lim_{n\to\infty} \mathcal{U} V_{n-1} = \lim_{n\to\infty} \mathcal{T} V_{n-1} = \mathcal{T} \lim_{n\to\infty} V_{n-1} = \mathcal{T}\hat{V} = \mathcal{U}\hat{V}$$

i.e. $\hat{V}$ is a solution of the optimality equation and $\hat{V}$ is lower semicontinuous. On the other hand we know from Theorem 3 that there exists a decision rule f^* which attains the infimum in $\hat{V} = \mathcal{U}\hat{V}$. Thus we obtain

$$\hat{V} = \mathcal{U}^n_{f^*} \hat{V} \ge \mathcal{U}^n_{f^*} 0$$

for all $n \in I\!N$ which implies $\hat{V} \ge V_{(f^*,f^*,\ldots)} \ge \inf_\pi V_\pi = V$. Therefore, $\hat{V} = V$. Moreover, if W is an arbitrary solution of the optimality equation we can repeat the arguments and obtain $W \ge V$. This completes the proof of a) and b). □

Remark 6:

a) A natural question that arises is why the policies have been defined in a discrete way in section 2. A natural candidate for a policy would be a measurable mapping $\pi_t : H_t \to U,\ t \ge 0$, where H_t gives the history of the process (X_t) up to time t and the corresponding state process satisfies $Y_t^\pi \in S$ for all $t \ge 0$. However, it is known that Theorem 4 remains valid if we would minimize over all policies $\pi = (f_n)$ such that f_{n+1} depends on the history $h_n = 0x_0f_1t_1x_1 \ldots f_nt_nx_n,\ n \in I\!N$. Thus in terms of Yushkevich (1980) Theorem 4 states that the optimal policy can be found among the simple strategies and applying Theorem 2 of Yushkevich (1980), we obtain under our assumption that minimizing over policies π_t gives the same value function.

b) All the previous Lemmas and Theorems remain valid, when we allow the environment process (Z_t) to be a more general semi-Markov process, i.e. if for $x \in E$ and policy π

$$P_x^\pi(T_{n+1} - T_n \leq t,\ Z_{T_{n+1}} = z' \mid T_0, X_{T_0}, \dots, T_n, Y_{T_n}, Z_{T_n} = z) = F_{zz'}(t) p_{zz'}.$$

If we denote by $\bar{F}_{zz'}(t) := 1 - F_{zz'}(t)$ the survival function, by $\bar{F}_z(t) := \sum_{z'} p_{zz'} \bar{F}_{zz'}(t)$ and by $f_{zz'}$ the density of $F_{zz'}$, then we obtain for the DSFP

$$p^{SM}(x, a; B \times \{z'\}) = p_{zz'} \int\limits_0^\infty e^{-\beta t} f_{zz'}(t) 1_B(\phi_t(x,a))\ dt$$

$$C^{SM}(x,a) := \int_0^\infty e^{-\beta t} \bar{F}_z(t) c(\phi_t(x,a), z, a_t)\ dt.$$

All other data remains the same. In particular the optimality equation (1) is now of the form

$$V(x) = \min_{a \in D(x)} \left[C^{SM}(x,a) + \int_0^\infty e^{-\beta t} \sum_{z'} p_{zz'} f_{zz'}(t) V(\phi_t(x,a), z')\ dt \right].$$

6 Properties of the Value Function

Suppose a SFP as defined in Section 2 is given and Assumptions 1 and 2 hold. We will prove several properties of the value function which will be important in obtaining structural results for the optimal control. In the following, we fix $z \in Z$.

Lemma 5:
If S is convex and $y \mapsto c(y,z,u)$ is convex for all $u \in U, z \in Z$ then $V(y,z)$ is convex in y.

Proof: The proof is by means of a sample path argument. The underlying probability measure is here the one of Remark 1 b). Let $y, y' \in S,\ \alpha \in [0,1]$. Moreover, denote by (π_t) and (π_t') the processes of the optimal policies for start in y and y' respectively. Define $\hat{\pi}_t = \alpha\pi_t + (1-\alpha)\pi_t'$. $\hat{\pi}_t \in U$ for all $t \geq 0$ since U is convex. Obviously $(\hat{\pi}_t)$ defines a policy. Take $(\hat{\pi}_t)$ as a control for start in $\alpha y + (1-\alpha)y'$. Hence

$$Y_t^{\hat{\pi}} = \alpha y + (1-\alpha)y' + \int_0^t b^{Z_t}(\alpha\pi_s + (1-\alpha)\pi_s')\ ds = \alpha Y_t^\pi + (1-\alpha) Y_t^{\pi'} \in S$$

since S is convex which yields that $\hat{\pi}$ is admissible. Therefore, we obtain

$$\begin{aligned} V(\alpha y + (1-\alpha)y') &\leq V_{\hat{\pi}}(\alpha y + (1-\alpha)y') = E_x\Big[\int_0^\infty e^{-\beta t} c(Y_t^{\hat{\pi}}, Z_t, \hat{\pi}_t)\, dt\Big] \\ &\leq \alpha V_\pi(y,z) + (1-\alpha)V_{\pi'}(y',z) = \alpha V(y,z) + (1-\alpha)V(y',z) \end{aligned}$$

and the proof is complete. □

For the next Lemma, we will need the following growth assumption on the cost rate function c

Assumption 3:

There exist constants $k \in I\!N$ and $C_0 \in I\!R_+$ such that for all $z \in Z, u, u' \in U$ and $y, y' \in S$

$$|c(y,z,u) - c(y',z,u')| \leq C_0\Big(1 + \|y\|^k + \|y'\|^k\Big)\Big(\|y-y'\| + \|u-u'\|\Big)$$

Lemma 6:

If $S = I\!R^N$ and $y \mapsto c(y,z,u)$ is continuously differentiable and convex for all $u \in U, z \in Z$ and fulfills Assumption 3 then $V(y,z)$ is continuously differentiable w.r.t. y.

Proof: Since V is convex due to Lemma 5 it suffices to show that the partial derivatives exist (cf. Rockafellar (1970)). Let $y, y' \in I\!R^N$ and $h > 0$. By e_ν we denote the ν-th unit vector. The convexity of c implies the convexity of V (Lemma 5), hence

$$D_1(y,h) := V(y,z) - V(y - he_\nu, z) \leq V(y + he_\nu, z) - V(y,z) =: D_2(y,h).$$

Let (π_t) be the process of the optimal policy for start in y. Due to our assumptions, (π_t) is also admissible for start in $y + he_\nu$ and $y - he_\nu$. Therefore, we obtain for the two differences above

$$D_2(y,h) \leq E_y^\pi\left[\int_0^\infty e^{-\beta t}\Big(c(Y_t + he_\nu, Z_t, \pi_t) - c(Y_t, Z_t, \pi_t)\Big)\, dt\right]$$

$$D_1(y,h) \geq E_y^\pi\left[\int_0^\infty e^{-\beta t}\Big(c(Y_t, Z_t, \pi_t) - c(Y_t - he_\nu, Z_t, \pi_t)\Big)\, dt\right].$$

If we now define

$$f(h) := \int_0^\infty e^{-\beta t} \frac{1}{h}\Big(c(Y_t + he_\nu, Z_t, \pi_t) - c(Y_t, Z_t, \pi_t)\Big)\, dt$$

then we have with Assumption 3 for $|h|$ small enough

$$|f(h)| \le C_0 \int_0^\infty e^{-\beta t}\Big(1 + \|Y_t + he_\nu\|^k + \|Y_t\|^k\Big)\, dt \le C_0'(y),$$

since the trajectories can grow at most linear. An analogous bound can be derived for the second difference. Thus, dividing both sides by h and letting $h \to 0$ we obtain with bounded convergence

$$E_y^\pi \left[\int_0^\infty e^{-\beta t} \frac{\partial}{\partial y} c(Y_t, Z_t, \pi_t)\, dt\right] \le \lim_{h\downarrow 0} \frac{D_1(y,h)}{h}$$
$$\le \lim_{h\downarrow 0} \frac{D_2(y,h)}{h} \le E_y^\pi \left[\int_0^\infty e^{-\beta t} \frac{\partial}{\partial y} c(Y_t, Z_t, \pi_t)\, dt\right] < \infty$$

which implies the statement. □

Lemma 7:

If $N = 2$ and $y \mapsto c(y, z, u)$ is supermodular for all $u \in U, z \in Z$ then $V(y, z)$ is supermodular in y.

Proof: Let $y, y' \in S$. W.l.o.g. $y_1 < y_1'$, $y_2 > y_2'$. Denote by $(\pi'{}_t)$ and (σ_t') the processes of the optimal policies for start in $(y \wedge y')$ and $(y \vee y')$ respectively ($\wedge$ denotes the componentwise minimum and $\vee$ the componentwise maximum). Define $t_1 := \inf\{t \ge 0 \mid Y_1^{\pi'}(t) = Y_1^{\sigma'}(t)\}$ and $t_2 = \inf\{t \ge 0 \mid Y_2^{\pi'}(t) = Y_2^{\sigma'}(t)\}$. Suppose that for fixed $\omega \in \Omega$ w.l.o.g. $t_1 \le t_2$. Define

$$\pi_1(t) = \begin{cases} \pi_1'(t) & , t \le t_1 \\ \sigma_1'(t) & , t > t_1 \end{cases} \qquad \sigma_1(t) = \begin{cases} \sigma_1'(t) & , t \le t_1 \\ \pi_1'(t) & , t > t_1 \end{cases}$$

$$\pi_2(t) = \sigma_2'(t),\ t \ge 0, \quad \sigma_2(t) = \pi_2'(t),\ t \ge 0.$$

Hence π and σ define admissible policies and

$$Y_1^\pi(t) = \begin{cases} Y_1^{\pi'}(t) & , t \le t_1 \\ Y_1^{\sigma'}(t) & , t > t_1 \end{cases} \qquad Y_1^\sigma(t) = \begin{cases} Y_1^{\sigma'}(t) & , t \le t_1 \\ Y_1^{\pi'}(t) & , t > t_1 \end{cases}$$

$$Y_2^{\pi}(t) = Y_2^{\sigma'}(t),\ t \geq 0, \quad Y_2^{\sigma}(t) = Y_2^{\pi'},\ t \geq 0.$$

Thus, the assertion follows with our assumption on c since

$$V(y \wedge y', z) + V(y \vee y', z) = V_{\pi'}(y \wedge y', z) + V_{\sigma'}(y \vee y', z)$$

$$= E\left[\int_0^\infty e^{-\beta t} c(Y_t^{\pi'}, Z_t, \pi'_t) + c(Y_t^{\sigma'}, Z_t, \sigma'_t)\ dt\right] \geq E\left[\int_0^\infty e^{-\beta t} c(Y_t^{\pi}, Z_t, \pi_t) + c(Y_t^{\sigma}, Z_t, \sigma_t)\ dt\right]$$

$$= V_\pi(y, z) + V_\sigma(y', z) \geq V(y, z) + V(y', z). \ \square$$

It is not clear whether Lemma 7 holds for $N > 2$. The construction of the policies π and σ we have used does not generalize to higher dimensions.

7 Numerical Methods for Stochastic Fluid Programs

When we have the special case of one environment state and a linear cost rate function then the optimization problem reduces to a so-called separated continuous linear program (SCLP) which can be solved quite efficiently (see e.g. Pullan (1993, 1995)). In general, we can use the *Approximating Markov chain approach* (see Kushner/Dupuis (1992)) to solve SFPs numerically. In what follows, we give a short outline of how to apply it to our SFP as defined in Section 2. First we look at a time discretization of our process. Let $h > 0$ be small and define $\Delta t^h = h\ (\max_{u \in U, z \in Z}\{\sum_{j=1}^N |b_j^z(u)|\})^{-1}$. Denote $D(x) = \{u \in U \mid y + b^z(u)\Delta t^h \in S\}$, $x \in E$ the set of admissible actions in state x. The discrete time optimality equation then reads

$$V^h(x) = \inf_{u \in D(x)} \left\{ \Delta t^h c(x, u) + e^{-\beta \Delta t^h} \left[\Delta t^h \sum_{z' \neq z} q_{zz'} V^h(y, z') + (1 - \Delta t^h q_z) V^h\left(y + b^z(u)\Delta t^h, z\right)\right]\right\}$$

In a next step we restrict the state space to a grid with distance $h > 0$. This can be done by applying a finite-element method. The crucial point is that the new state $y + b^z(u)\Delta t^h$ can be written as a convex combination of grid points:

$$b^z(u)\Delta t^h \quad = \quad \sum_{j=1}^N h e_j \frac{b_j^z(u)^+ \Delta t^h}{h} + \sum_{j=1}^N (-h e_j) \frac{b_j^z(u)^- \Delta t^h}{h}$$

where x^+ and x^- denote the positive and negative part of x respectively. Notice that the sum of the weights is less than 1 due to the definition of Δt^h. Approximating the value function by a linearization over the grid, we obtain the following optimality equation

$$V^h(x) = \min_{u \in D(x)} \Big\{ \quad \Delta t^h c(x,u) + e^{-\beta \Delta t^h} \Big(\Delta t^h \sum_{z' \neq z} q_{zz'} V^h(y,z') + (1 - \Delta t^h q_z)$$

$$\Big[\sum_{j=1}^N \frac{b_j^z(u)^+ \Delta t^h}{h} V^h(y + he_j, z) + \sum_{j=1}^N \frac{b_j^z(u)^- \Delta t^h}{h} V^h(y - he_j, z) +$$

$$\Big[1 - \frac{\Delta t^h}{h} \sum_{j=1}^N |b_j^z(u)| \Big] V^h(x) \Big] \Big) \Big\}$$

Under Assumptions 1 and 2 there exists a minimizer f of V^h and the stationary policy $\pi = (f, f, \ldots)$ is optimal. From Kushner/Dupuis (1992) we know that for $h \to 0$ the value functions $V^h(x)$ converge to $V(x)$ for every environment state z.

8 Application

In this section we apply our general results to the *multi-product manufacturing system* described in the introduction.

A controller has to decide upon the production rates for N items in parallel. If the environment process is in state z at time t the maximal production rate for items of type j is $\lambda_j(z) > 0$. The demand rate for item j is $\mu_j > 0$, $j = 1, \ldots, N$. The controller obtains a reward r_j for each unit of item j which is produced, but has to pay inventory costs $\hat{c}(y)$ which depend on the joint inventory y. We formulate the control as a vector $(u, v) \in [0,1]^N \times [0,1]^N$, where $\lambda_j(z) u_j$ is the production rate for items of type j and $v_j \mu_j$ is the demand rate for items of type j. If we would fix the demand rate μ_j, then it is inconvenient to define the control set, since it must be possible to hold the inventory non-negative, which forces an artificial definition of u_j if $\lambda_j(z) < \mu_j$ and $y_j = 0$. It will turn out in our formulation that $v_j = 1$, whenever $y_j > 0$ (see Theorem 9b)). In terms of our SFP, the data is given by

$$E = \mathbb{R}_+^N \times Z$$

$$U = [0,1]^N \times [0,1]^N$$
$$b_j^z(u) = \lambda_j(z)u_j - \mu_j v_j,\ j = 1,\ldots,N$$
$$c(x) = \hat{c}(y) - \sum_{j=1}^{N} r_j \lambda_j(z) u_j$$

To obtain a reasonable model we will assume that the mapping $y \mapsto \hat{c}(y)$ is non-negative, strictly increasing and convex ($\leq$ is interpreted componentwise). Note that the problem does not decouple into N independent problems since there is a common cost function $\hat{c}$ for the inventory and we do not assume that $\hat{c}$ is additive, i.e. $\hat{c}(y) = \sum_{j=1}^N \hat{c}_j(y_j)$ (this of course would give us N independent problems). This is a realistic assumption when we suppose that the N type of items share a common inventory. Moreover, in Theorem 9 we will assume that $\hat{c}$ is supermodular which means that increases in the inventory of one type of item will make the storage of additional items of other types more expensive. This situation is often encountered in practice. Note also that we assume an infinite capacity for the inventory and that the only stochastic part in the model is the random variation of the production capacity of the machines (for a motivation see Sethi/Zhang (1994)). Obviously Assumptions 1 and 2 of sections 3 and 5 are satisfied (for the control which holds the inventory process in the initial state we have $V_\pi(x) < \infty$) and we obtain with Theorem 4:

Theorem 8:
In the parallel machine production problem without backlog there exists an optimal stationary policy $(f^*, f^*, \ldots)$ where f^* is a minimizer of V in (1).

In the cases of one or two items ($N \leq 2$) we can further show that the optimal policy has a certain structure. From now on we focus on determining the optimal production policy which is given by the first N components of f^*. We will call a feedback control $g : E \to [0,1]^N$, $g(y,z) = (g_1(y,z), \ldots, g_N(y,z))$ of switching-type, if it has the following properties for all $z \in Z, j = 1, \ldots, N$.

(i) $g_j(y,z) = 1,\ y' \leq y\ \Rightarrow g_j(y',z) = 1.$

(ii) $g_j(y,z) = 0,\ y' \geq y\ \Rightarrow g_j(y',z) = 0.$

When the production control is of switching-type this means that if it is optimal to produce items of type j at maximum rate when the inventory is y, then also when the inventory is $y' \leq y$. Vice versa, if it is optimal to stop production of items of type j when the inventory is y, then also when $y' \geq y$.

Theorem 9:

Suppose $N \leq 2$ and $y \mapsto \hat{c}(y)$ is strictly increasing, convex, supermodular, continuously differentiable and satisfies the growth condition of Assumption 3. Then we obtain

a) The value function $y \mapsto V(y,z)$ is strictly increasing, convex, supermodular and continuously differentiable for all $z \in Z$.

b) The optimal production policy is a feedback-control g of switching type and $v_j = 1$, if $y_j > 0$, $j = 1, \ldots, N$.

c) Further, suppose V is twice continuously differentiable and strictly convex. If $N = 2$, then there exists for every z and j a so-called *switching-curve* $y_1 \mapsto S_j(y_1, z)$ which is continuous and decreasing, such that for $j = 1, 2$

$$g_j(y,z) = \begin{cases} 1, & \text{if } y_2 < S_j(y_1,z) \\ 0, & \text{if } y_2 > S_j(y_1,z). \end{cases}$$

Proof:

a) See the appendix for this part.

b) We have to solve now the optimality equation (1). This is a deterministic control problem and the Hamiltonian of the optimization problem (1) is given by

$$\begin{aligned} H(y,u,v,p) &= \sum_{j=1}^{N} [p_j(\lambda_j(z)u_j - \mu_j v_j) - r_j \lambda_j(z) u_j] + \sum_{z' \neq z} q_{zz'} V(y,z') + \hat{c}(y) \\ &= \sum_{j=1}^{N} [u_j \lambda_j(z)(p_j - r_j) - p_j \mu_j v_j] + \sum_{z' \neq z} q_{zz'} V(y,z') + \hat{c}(y). \end{aligned}$$

Since V is continuously differentiable, it is well-known (cf. Seierstad/Sydsæter (1987) p. 212) that $p_t = \frac{\partial V}{\partial y}(y_t^*, z)$, where (y_t^*) is the optimal trajectory. According to Pontryagins maximum principle the optimal control $u_j^*(t)$ at time t minimizes the function

$u_j \mapsto u_j \left(\frac{\partial V}{\partial y_j}(y_t^*, z) - r_j \right)$. Thus we obtain

$$u_j^*(t) = \left\{ \begin{array}{ll} 1 & \text{, if } r_j > \frac{\partial V}{\partial y_j}(y_t^*, z) \\ 0 & \text{, if } r_j < \frac{\partial V}{\partial y_j}(y_t^*, z) \end{array} \right\} =: g_j(y_t^*, z), \quad 1 \le j \le N$$

and $u_j^*(t)$ holds the inventory on the line $r_j = \frac{\partial V}{\partial y_j}(y_t^*, z)$ whenever this is possible. This means that the optimal production control is a feedback control. Similar $v_j^*(t)$ at time t maximizes $v_j \mapsto v_j \frac{\partial V}{\partial y_j}(y_t^*, z)$. Since V is strictly increasing we have $\frac{\partial V}{\partial y} > 0$ and $v_j^*(t) = 1$ if $y_j^* > 0$ (if $y_j^* = 0$, $v_j^*(t)$ is not necessary equal to 1, because of the boundary condition). It remains to show that g_j is of switching-type. Therefore, let $(y, z) \in E$ and $g_j(y, z) = 1$, hence $r_j > \frac{\partial V}{\partial y_j}(y, z)$. Let $y' = y - \delta e_j$, $\delta \ge 0$, where e_j is the $j-$th unit vector in $I\!R^N$. Since V is convex we obtain that $\frac{\partial V}{\partial y_j}(y, z)$ is increasing in y_j, hence $\frac{\partial V}{\partial y_j}(y', z) < r_j$ which implies that $g_j(y', z) = 1$. Similar if $y' = y - \delta e_k$, $k \neq j$, $\delta \ge 0$ then, since V is supermodular, $\frac{\partial V}{\partial y_j}(y, z)$ is increasing in y_k, hence $\frac{\partial V}{\partial y_j}(y', z) < r_j$ which implies that $g_j(y', z) = 1$. The assertion for arbitrary $y \in I\!R_+^N$ follows by induction over the components. The second property of the switching-type control can be shown analogously.

c) If V is strictly convex then $\nabla^2 V$ is regular. Using the Theorem for implicit functions we obtain that there exists a continuous function $S_j : I\!R_+ \times Z \to \bar{I\!R}$ such that $r_j = \frac{\partial V}{\partial y_j}(y_1, S_j(y_1, z), z)$. The monotonicity of S_j follows from the proof of part b).

□

Remark 7:

a) The control on the switching curve is such that the inventory is kept on the switching curve if this is possible, until the next environment change occurs.

b) If $N = 1$, the optimal control is of threshold-type. That means, there exists a constant y_z^*, the so-called turnpike level, for every environment state z such that it is optimal to produce nothing if $y > y_z^*$ and to produce at maximal rate if $y < y_z^*$. The production rate at y_z^* is $\lambda(z)u$ with $u = \frac{\mu}{\lambda(z)}$ in the case that $\lambda(z) > \mu$. This is shown in Rajagopal et al. (1995).

c) When we look at the same model with backlog, i.e. $S = I\!\!R^N$, we obtain the same structure for the production policy. Indeed, the analysis is easier here, since we do not need the control variables $v_j,\ j = 1, \ldots, N$ and the differentiability of the value function follows directly from Lemma 6. Instead of $\hat{c}$ increasing we have of course to assume that $\hat{c}(y) \to \infty$ for $|y| \to \infty$.

The following numerical computation of the optimal policy has been done for the one- and two-item case with two environment states using the approximating Markov chain approach. In the finest grid we have used for our computation, the points had distance $h = 0.033$. This leads to 30,000 states in the discrete value iteration which terminates after 50,000 iterations. Obviously, for small h it is not practical to do the computation for more than a couple of environment states. The aim of the computation was to get some conjectures about how the switching curves depend on the production capacity and on the intensity with which the environment changes. For these questions we got some nice results.

Figure 1 and 2 refer to the **one-item case** with $\hat{c}(y) = (y+0.5)^2,\ \beta = 0.9,\ r = \frac{20}{9},\ \mu = 2$. In figure 1 we have fixed $q_0 = q_1 = 2, \lambda(0) = 4$ and have varied the maximal production rate in environment state 1, $\lambda(1)$ from 0 to 2.5. The curve consisting of circles represents the optimal threshold y_0^* in environment state 0 and the other curve, the optimal threshold y_1^* in environment state 1. In Sethi et al. (1992) it has been shown that if $\lambda(0), \lambda(1) \geq \mu$, which is the case if $\lambda(1) \geq 2$, the optimal thresholds are independent of the environment state and can be computed from $\frac{\partial}{\partial y}\hat{c}(y^*) = \beta r$ which gives $y^* = 0.5$ in our case. From Rajagopal et al. (1995) we know that $\lambda(1) \leq \lambda(0)$ implies that $y_1^* \geq y_0^*$. Moreover the numerical computations allow to conjecture that the optimal thresholds are decreasing in the production rate $\lambda(1)$. In terms of our model this means that a lower maximal production rate forces us to keep a higher threshold level which can be seen as a benchmark inventory we should try to keep. Since both environment states are coupled by a stochastic mechanism, this observation is true for both states, whereas of course state 1 with the lower production rate is more affected. The influence on state 0 is regulated by the intensities q_0, q_1 with which the environment changes as we will see in figure 2. The higher target inventory is the price we have to pay for being less flexible with a lower production rate. Also it is important to note from an economical point of view that it is sufficient to keep the production rate in any state slightly above the demand rate. Any further additional production rate does not lead to a lower target inventory. In figure 2 we have fixed the two maximal production rates $\lambda(0) = 4$ and $\lambda(1) = 1$ and have varied the intensity $q_0 = q_1$ with which the environment process changes. For $q_0 \to 0$ the system decouples into two deterministic systems with thresholds $y_0^* = 0.5$ and $y_1^* = 1.446$. For $q_0 \to \infty$ the environment process converges uniformly on compact sets to a constant production rate $\bar{\lambda} = \frac{1}{2}(\lambda(0) + \lambda(1)) = 2.5$. Hence we would expect that both $y_0^* = y_0^*(q_0)$ and $y_1^* = y_1^*(q_0)$ converge to 0.5 which is the optimal threshold in the deterministic case with production rate $\bar{\lambda}$. Indeed, for our simple example, this statement follows from Remark 7.3 in Chapter 5 of Sethi/Zhang (1994), in fact, numerical computation reveals that when $q_0 = 100$, $y_0^*(q_0) = 0.54$. Remark 5.7.3 in Sethi/Zhang (1994) suggests that this convergence is very slow, presumably of order of the fourth root of $\frac{1}{q_0}$. Moreover, $y_0^*(q_0)$ is decreasing and $y_1^*(q_0)$ has a unique maximum point which can be interpreted as the parameter setting possessing the most randomness. As far as the model is concerned it is interesting to note that the threshold levels are much higher in a setting with an unequal production rate of 1 and 4 compared to the situation with a constant average production rate of 2.5. Indeed, this effect seems to be quite resistant even when we considerably increase the speed of environment changes. Thus, we can conclude that it should be the highest priority of a manufacturing system to try and keep the production capacity as constant as possible.

The figures 3-6 for the **two-item case** show the same behavior. Here we have chosen the following data: $c(y_1, y_2) = e^{y_1+y_2}, \beta = 0.9, r_1 = r_2 = \frac{40}{9}, \mu_1 = \mu_2 = 2$. From Theorem 9 we know that the optimal production policy is characterized by 4 switching-curves $S_1(y_1, 0), S_2(y_1, 0), S_1(y_1, 1)$ and $S_2(y_1, 1)$ where $y_2 \leq S_j(y_1, z)$ if and only if the maximal production rate is used for item j in environment state z, when the inventories are y_1 and y_2 respectively. Since the data is symmetric in item 1 and 2, the optimal policy is also symmetric. Hence we can restrict w.l.o.g. to the policy for the first item. In figure 3 and 4 we see the optimal policy for item 1 in environment states 0 and 1 respectively, with $q_0 = q_1 = 2, \lambda_1(0) = \lambda_2(0) = 4$ where we have varied $\lambda_1(1) = \lambda_2(1)$ from 0.4 to 2.3. The region below the curve is the maximal production region. It seems that the optimal policy in the two-item case has the same properties as in the one-item case, that is: as soon as $\lambda_1(1) > \mu_1$, the policy does not change; both maximal production regions increase when $\lambda_1(1)$ decreases and the maximal production region in environment state 1 is always greater than the one in environment state 0. This implies that we can draw the same conclusions for the model in the two-buffer case, namely that a decrease of the maximal production rate below the demand rate leads to a sort of unflexibility which forces us to keep higher inventories. In figure 5 and 6 we have fixed $\lambda_1(0) = \lambda_2(0) = 4, \lambda_1(1) = \lambda_2(1) = 1$ and varied the intensity with which the environment process changes, where $q_0 = q_1$. Figure 5 refers to the optimal policy in environment state 0, figure 6 to the one in environment state 1. Again, for $q_0 \to 0$ and $q_0 \to \infty$ we are in completely deterministic settings and the maximal production region in environment state 1 is decreasing in q_0. Moreover, the acceptance region in environment state 1 is always greater than the one in environment state 0. Again we have the situation that the stochastic changes between two states, with one state having production capacity below demand rate, puts us in a worse situation than in the deterministic setting with average production rate.

Appendix

Proof of Theorem 1: Part b) follows directly from a). For a) let π be fixed. If we denote by $\{\mathcal{F}_t\}$ the natural filtration of the state process (X_t) we obtain by conditioning on $\{\mathcal{F}_{T_n}\}$

$$\begin{aligned} V_\pi(x) &= E_x^\pi\left[\int_0^\infty e^{-\beta t} c(X_t, \pi_t)\, dt\right] = E_x^\pi\left[\sum_{n=0}^\infty \int_{T_n}^{T_{n+1}} e^{-\beta t} c(X_t, \pi_t)\, dt\right] \\ &= E_x^\pi\left[\sum_{n=0}^\infty E_x^\pi\left\{\int_{T_n}^{T_{n+1}} e^{-\beta t} c(X_t, \pi_t)\, dt \Big| \mathcal{F}_{T_n}\right\}\right] \\ &= E_x^\pi\left[\sum_{n=0}^\infty e^{-\beta T_n} E_x^\pi\left\{\int_0^{T_{n+1}-T_n} e^{-\beta t} c(Y_{T_n+t}, Z_{T_n}, f_n(X_{T_n})(t-T_n))\, dt \;\Big|\; \mathcal{F}_{T_n}\right\}\right] \\ &= E_x^\pi\left[\sum_{n=0}^\infty e^{-\beta T_n} C\left(X_{T_n}, f_n(X_{T_n})\right)\right], \end{aligned}$$

c.f. also Davis (1993). Now we will show by induction on $m \in \mathbb{N}$ that for all $x \in E,\ m \in \mathbb{N}$

$$E_x^\pi\left[\sum_{n=0}^m e^{-\beta T_n} C\left(X_{T_n}, f_n(X_{T_n})\right)\right] = \hat{E}_x^\sigma\left[\sum_{n=0}^m C\left(X_n, f_n(X_n)\right)\right]$$

which yields the result. $m = 0$ is obvious. Suppose the assertion is valid for $k = 0, \dots, m-1$. Then we obtain by applying the induction hypothesis

$$\begin{aligned}
& E_x^\pi \left[\sum_{n=0}^{m} e^{-\beta T_n} C\left(X_{T_n}, f_n(X_{T_n})\right) \right] \\
= \; & C(x, f_0(x)) + E_x^\pi \left[e^{-\beta T_1} E_{X_{T_1}}^\pi \left\{ \sum_{n=1}^{m} e^{-\beta(T_n - T_1)} C(X_{T_n}, f_n(X_{T_n})) \;\Big|\; \mathcal{F}_{T_1} \right\} \right] \\
= \; & C(x, f_0(x)) + \sum_{z' \neq z} \frac{q_{zz'}}{q_z} \int_0^\infty e^{-\beta t} E_{(\phi_t(x, f_0), z')}^\pi \left[\sum_{n=0}^{m-1} e^{-\beta T_n} C(X_{T_n}, f_{n+1}(X_{T_n})) \right] q_z e^{-q_z t} \, dt \\
= \; & C(x, f_0(x)) + \sum_{z' \neq z} q_{zz'} \int_0^\infty e^{-(\beta + q_z)t} \hat{E}_{(\phi_t(x, f_0), z')}^\sigma \left[\sum_{n=0}^{m-1} C(X_n, f_{n+1}(X_n)) \right] dt \\
= \; & \hat{E}_x^\sigma \left[\sum_{n=0}^{m} C\left(X_n, f_n(X_n)\right) \right]. \; \square
\end{aligned}$$

Proof of Theorem 9 a): The convexity and supermodularity of $y \mapsto V(y, z)$ follow directly from Lemma 5 and Lemma 7 respectively. Let us next show that $y \mapsto V(y, z)$ is continuously differentiable. Since $S = I\!R_+^N$ we have to modify the proof of Lemma 6. We have to show that the right and left partial derivatives are the same in the interior of the state space (the single-sided derivatives at the boundary exist since V is convex and V is also continuous at the boundary). As in the proof of Lemma 6 we will do this for every sample path $\omega \in \Omega$. Moreover, since the control of one component does not influence the dynamics of the other components, it suffices to restrict to the case $N = 1$. Now let $y > h > 0$ and fix $\omega \in \Omega$. Denote by (π_t) the process of the optimal policy starting in y. Let $\tau := \inf\{t \geq 0 \mid Y_t^\pi = 0\}$ and suppose $Y_t^\pi = 0$ on a positive time interval $[\tau, \tau + \Delta]$ (it takes a small thought to see that the optimal trajectory cannot simply touch zero and then get positive again). If $\tau = \infty$, then we can proceed as in Lemma 6. Suppose $\tau < \infty$. The probability is $O(h)$ that a jump in the environment occurs during the time interval $[\tau - \Delta h, \tau + \Delta h]$ for fixed $\Delta > 0$. Therefore, suppose that no jump occurs. Denote by $\tau^l := \inf\{t \geq 0 \mid Y_t^\pi - h = 0\}$. Obviously, $\tau^l < \tau$ and let $h > 0$ be small enough such that $Y_t^\pi - h \leq 0$ for $t \in [\tau^l, \tau]$. Define $\tau^r := \tau + (\tau - \tau^l)$. We will construct the policies (π_t^r), (π_t^l) for start in state $y + h$ and $y - h$ respectively such that

$$\pi_t^l = \pi_t^r = \pi_t \quad \text{for} \quad 0 \leq t \leq \tau^l, \; t \geq \tau^r$$

$$\pi^l_t = \pi_{\tau+t-\tau^l} \quad \text{and} \quad \pi^r_t = \pi_t \quad \text{for} \quad \tau^l \le t \le \tau$$

$$\pi^l_t = \pi_t \quad \text{and} \quad \pi^r_t = \pi_{\tau_l+t-\tau} \quad \text{for} \quad \tau \le t \le \tau^r.$$

Hence (π^l_t) and (π^r_t) give admissible policies and we have $Y^{\pi^l}_{\tau^r} = Y^{\pi^r}_{\tau^r} = Y^{\pi}_{\tau^r} = 0$ and

$$\lim_{h\to 0} \int_0^\infty e^{-\beta t} \frac{1}{h} \Big(c(Y^{\pi^r}_t, Z_t, \pi^r_t) - c(Y^{\pi}_t, Z_t, \pi_t) \Big)\, dt$$

$$= \lim_{h\to 0} \int_0^{\tau^l} \ldots\, dt + \lim_{h\to 0} \int_{\tau^l}^{\tau} \ldots\, dt + \lim_{h\to 0} \int_{\tau}^{\tau^r} \ldots\, dt = \lim_{h\to 0} I^r_1(h) + \lim_{h\to 0} I^r_2(h) + \lim_{h\to 0} I^r_3(h).$$

And

$$\lim_{h\to 0} \int_0^\infty e^{-\beta t} \frac{1}{h} \Big(c(Y^{\pi}_t, Z_t, \pi_t) - c(Y^{\pi^l}_t, Z_t, \pi^l_t) \Big)\, dt = \lim_{h\to 0} \int_0^{\tau^l} \ldots\, dt + \lim_{h\to 0} \int_{\tau^l}^{\tau} \ldots\, dt$$

$$= \lim_{h\to 0} I^l_1(h) + \lim_{h\to 0} I^l_2(h).$$

Due to the construction of our policies it holds that $\lim_{h\to 0} I^r_2(h) = 0$ since on $[\tau^l, \tau]$ the rewards are equal for π and π^r and the derivative of $\hat{c}$ is bounded due to the growth condition of Assumption 3. Moreover, we have $\lim_{h\to 0} I^r_1(h) = \lim_{h\to 0} I^l_1(h)$ and $\lim_{h\to 0} I^r_3(h) = \lim_{h\to 0} I^l_2(h)$. The interchange of expectation and limit can be proved in the same way as in Lemma 6. Hence the statement follows.

The monotonicity of V can be shown, using a similar construction of policies. □

Acknowledgment

I am grateful to Ulrich Rieder for some deep discussions and to two referees for their valuable comments.

References

Akella RA and PR Kumar (1986) Optimal control of production rate in a failure prone manufacturing system. *IEEE Trans. Automa. Control AC* **31** 116-126.

Anick D, Mitra D and MM Sondhi (1982) Stochastic theory of a datahandling system with multiple sources. In *The Bell System Technical Journal* **61** 1871-1894.

Bertsekas DP and SE Shreve (1978) *Stochastic optimal control: the discrete time case.* Academic Press, New York.

Davis MHA (1993) *Markov models and optimization.* Chapman & Hall, London.

HERNÁNDEZ-LERMA O AND JB LASSERRE (1996) *Discrete-time Markov control processes.* Springer-Verlag, New York.

HINDERER K (1984) *Grundbegriffe der Wahrscheinlichkeitstheorie.* Springer-Verlag, Berlin Heidelberg.

HORDIJK A AND FA VAN DER DUYN SCHOUTEN (1983) Average optimal policies in Markov decision drift processes with applications to a queueing and a replacement model. *Adv. Appl. Probab.* **15** 274-303.

KUSHNER HJ AND PG DUPUIS (1992) *Numerical methods for stochastic control problems in continuous time.* Springer-Verlag, New York.

KUSHNER HJ (1990) Numerical methods for stochastic control problems in continuous time. *SIAM J. Contr. Optim.* 28, 999-1048.

PRESMAN E, SETHI SP AND Q ZHANG (1995) Optimal feedback production planning in a stochastic N-machine flowshop. *Automatica* **31** 1325-1332.

PULLAN MC (1993) An algorithm for a class of continuous linear programs. *SIAM J. Control Optim.* **31** 1558-1577.

PULLAN MC (1995) Forms of optimal solutions for separated continuous linear programs. *SIAM J. Control Optim.* **33** 1952-1977.

RAJAGOPAL S (1995) *Optimal control of stochastic fluid-flow systems with applications to telecommunication and manufacturing systems.* PhD Dissertation at the University of North Carolina, Chapel Hill.

RAJAGOPAL S, KULKARNI VG AND S STIDHAM (1995) Optimal flow control of stochastic fluid-flow systems. *IEEE Journal on selected areas in Communications* **13** 1219-1228.

ROCKAFELLAR RT (1970) *Convex analysis.* Princeton University Press, Princeton.

SEIERSTAD A AND K SYDSÆTER (1987) *Optimal control theory with economic applications*, North-Holland, Amsterdam.

SETHI SP, SONER HM, ZHANG Q AND J JIANG (1992) Turnpike sets and their analysis in stochastic production planning problems. *Math. Operations Res.* **17** 932-950.

SETHI SP AND Q ZHANG (1994) *Hierarchical decision making in stochastic manufacturing systems.* Birkhäuser, Boston.

YUSHKEVICH AA (1980) On reducing a jump controllable Markov model to a model with discrete time. *Theory Probab. and Appl.* **25** 58-69.