Optimal Control of Stochastic Fluid Programs

Habilitationsschrift
an der Fakultät für Mathematik und
Wirtschaftswissenschaften
der Universität Ulm

vorgelegt
von
Nicole Bäuerle

Ulm
1999

...meinem Mann Rolf,
für seine Liebe und Geduld.

List of Symbols

Commonly used Symbols

$I\!N$	set of positive integers
$I\!N_0$	$I\!N \cup \{0\}$
$I\!R$	set of real numbers
$I\!R_+$	set of nonnegative real numbers
$\overline{I\!R}_+$	$I\!R_+ + \{\infty\}$
$\mathfrak{B}(S)$	Borel-σ-algebra on S
$\mathring{S}$	interior of S
$1_S(\cdot)$	indicator function of set S
e_i	i-th unit vector
$1\!\!1_k$	vector of 1's with dimension k
$\|h\|$	$\max\{h, -h\}$.
$\\|\cdot\\|$	vector norm.
$x \wedge y$	componentwise minimum of vectors x and y.
$x \vee y$	componentwise maximum of vectors x and y.
$\frac{\partial}{\partial y} V(y, z)$	derivative w.r.t. y.
$\dot{p}_t$	derivative w.r.t. time t.
I	identity matrix
δ_x	Dirac measure.
$\Rightarrow$	weak convergence.
$< \cdot >$	quadratic variation.
$D^N[0, \infty)$	set of functions $f : [0, \infty) \to I\!R^N$ which are right continuous and have left-hand limits.

Abbreviations

a.s.	almost sure.
DSFP	Discretized Stochastic Fluid Program.
i.i.d.	independent and identically distributed.
SFP	Stochastic Fluid Program.
w.l.o.g.	without loss of generality.
w.r.t.	with respect to.

Contents

1 Introduction

In manufacturing and telecommunication systems we often encounter the situation that there are different timescales for the occurence 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)). In the celebrated Anick/Mitra/Sondhi-model (1982), 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 (talkspurt state) and no transmission when the source is off (silent state). 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 stochastics of the slower process. 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 linear 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 $\beta > 0$ as well as minimizing the average cost for $\beta = 0$.
Let us first look at the following *example* of a multi-product manufacturing system with backlog. We have a number of machines in parallel which can produce N different items and certain demand rates $\mu_1, \ldots, \mu_N \geq 0$ for the items. Denote $\mu := (\mu_1, \ldots, \mu_N)$. Since the machines are subject to random breakdown and repair, the total production capacity $\lambda(z) \in I\!R_+$ depends on the number $z = Z_t$ of working machines at time t. Z_t is our environment process. The vector $Y_t = (y_1(t), \ldots, y_N(t))$ gives the inventory/backlog of each product at time t and we assume $S = I\!R^N$. We have to decide now upon the partition of the production capacity, hence we define $U = \{u \in [0,1]^N \mid \sum_{j=1}^N u_j \leq 1\}$, where u_j is the percentage of the production capacity that is assigned to product j, $j = 1\ldots, N$. For $u \in U, z \in Z$ the local dynamics of the system are given by $b^z(u) = \lambda(z)u - \mu$. Hence, the data

$$E = I\!R^N \times Z$$

$$U = \{u \in [0,1]^N \mid \sum_{j=1}^N u_j \leq 1\}$$

$$b^z(u) = \lambda(z)u - \mu$$

together with a cost rate function c, interest rate β and generator Q of the environment process specifies our problem.
In Section 2 we will consider the β-discounted optimization problem. 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. 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 Davis (1993), Forwick (1998)) 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 literature one can find examples of SFP which have been solved explicitely, see e.g. Akella/Kumar (1986), Presman et al. (1995), Rajagopal et al. (1995), Bäuerle (1998b). 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, one is here allowed to control the jumps of the process and not the deterministic behaviour between jumps. 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 use of relaxed controls, we will make several convexity assumptions. For our applications this is no crucial restriction. We will prove under some continuity and compactness assumptions that an optimal stationary policy exists which is the solution of a deterministic control problem (Theorem 2.5). Moreover, we show under certain conditions that the value function V is a constrained viscosity solution of a *Hamilton-Jacobi-Bellman* (HJB) equation and derive a verification Theorem (Theorem 4.3).
Beyond the discounted cost, we will consider in Section 3 the minimization of the average cost, i.e. we are interested in finding

$$G(x) = \inf_\pi G_\pi(x) = \limsup_{t\to\infty} \frac{1}{t} E_x^\pi \left[\int_0^t c(X_s, \pi_s)\, ds \right].$$

Due to some technical reasons we are forced to consider a slight modification of our SFP. We will now work with the uniformized version of the environment process (Z_t) and allow decisions to be taken at jump times of the uniformized version (whether or not a real jump occurs). There are only very few recent papers dealing with the average cost criterion in SFP, see for example the special production model in Sethi et al. (1997) and Sethi et al. (1998). We tackle the problem again by discretizing the continuous problem and using the vanishing discount approach. Under certain assumptions, which are mainly due to Sennott (1989a) and following essentially the ideas in Schäl (1993), we prove the existence of average cost optimal policies (Theorem 3.6). This has not been done in the earlier work of Sethi et al. (1997, 1998). Since the assumptions in Theorem 3.6 are not easy to verify, we will give some sufficient conditions for them. Mainly these conditions imply positive Harris recurrence of the controlled state process. We will also show that the relative value function is a constrained viscosity solution of a HJB-equation and derive a verification Theorem (Theorem 4.4).

We will apply our results to three examples which are interesting for themselves. The first one is the previously defined *multi-product manufacturing system*. It has already been considered in Sethi/Zhang (1994), Sethi et al. (1997) and Sethi et al. (1998). However, their approach is different from ours in that they directly operate with the continuous model. In the cases of one or two products we derive the optimality of threshold and switching-curve policies respectively (cf. also Rajagopal et al. (1995)). The second example is a generic *single-server network* with routing which is the fluid analogue of the famous Klimov problem (see e.g. Klimov (1974), Walrand (1988)). The environment process influences here the inflow rates of the buffers. The purely deterministic model has been investigated in Chen/Yao (1993). It is possible to prove that the optimal policy is a so-called index-policy and the indices coincide with the indices of the Klimov-problem. This result holds for the discounted as well as for the average cost problem (here under a suitable stability condition which implies the finiteness of the average cost) and the indices are independent of the interest rate and the arrival intensity. The third application is the *routing to parallel queues*, where the arrival rate of fluid depends on the environment process. In the case of equal linear holding cost we can show that the least-loaded routing policy is optimal for both optimization criteria. In the deterministic two-buffer case we obtain the optimality of a switching curve policy, where the switching curve can be computed explicitely.

Another interesting topic that we will deal with in Section 7 is the following: If we have only one environment state, then our SFP reduces obviously to the following purely deterministic control problem

$$(F)\begin{cases} \int_0^\infty e^{-\beta t} c(y_t, a_t)\, dt \to \min \\ y_t = y_0 + \int_0^t b(a_s)\, ds \\ y_t \geq 0, \\ a_t \in U,\ t \geq 0 \end{cases}$$

which we will refer to as the *fluid problem*. In recent years it turned out that there is a close connection between the stability of stochastic queueing networks and their associated fluid problems (see e.g. Dai (1995), Bramson (1996), Maglaras (1998a)). Following this idea, together with the observation that the optimal policies in stochastic networks and the associated fluid problem often coincide, several authors have conjectured that there is a strong connection between these optimization problems (see Atkins/Chen (1995), Avram et al. (1995), Avram (1997), Meyn (1997)). Such a connection would be very helpful since optimization problems in stochastic networks are notoriously difficult to solve. Meyn (1997) proved for the average cost case that the relative value functions, when properly normalized converge against the value function in the fluid model. Since the problem (F) is relatively easy to solve (it often reduces to a so-called *separated continuous linear program* (SCLP) which can be solved quite efficiently, cf. Pullan (1993, 1995), Weiss (1996, 1997)), the crucial question is how the optimal control of (F) can be translated in a "good" policy for the stochastic network. A numerical study, where so-called "Fluid Heuristics" are used for the control of stochastic networks can be found in Atkins/Chen (1995). Alanyali/Hajek (1998) consider a special routing problem and prove that the load-balancing policy which is optimal in the associated fluid problem is asymptotically optimal in the stochastic network. In Maglaras (1998a,b, 1999) one can find a systematical way to construct asymptotic optimal policies in multi-class queueing networks for finite horizon problems (so-called *discrete review policies*). The asymptotics is w.r.t. fluid scaling, which works as follows: Suppose $y \in S$ and denote by $(\hat{Y}_t^\gamma)$ the state process starting in γy under policy $\pi^\gamma = (f_n^\gamma)$, $\gamma \in I\!N$. The scaled state and action processes are $Y_t^\gamma := \frac{1}{\gamma}\hat{Y}_{\gamma t}^\gamma$ and $\pi_t^\gamma = f_n^\gamma(\hat{Y}_{T_n}^\gamma)$ if $T_n \le \gamma t < T_{n+1}$ respectively, where (T_n) are the jump times of $(\hat{Y}_t^\gamma)$. The corresponding value function is

$$V_{\pi^\gamma}^\gamma(y) = E_y^{\pi^\gamma}\left[\int_0^\infty e^{-\beta t} c(Y_t^\gamma, \pi_t^\gamma)\, dt\right],$$

i.e. we increase the intensity of the process by factor γ and reduce the jump heights by the same factor. A sequence of policies π^γ is asymptotically optimal, if $\lim_{\gamma\to\infty} V_{\pi^\gamma}^\gamma(y) = V^F(y)$ for all $y \in S$. We will propose a different class of asymptotically optimal policies, which we will call *Tracking-policies*. The policy is instationary, however easy to implement. It relies on the fact that the optimal control of the fluid problem is often piecewise constant (see Pullan (1995)) and hence uses the corresponding control on properly defined time intervals. Since the trajectories of the so controlled stochastic network converge under fluid scaling against the trajectory of the fluid problem, we have named this policy Tracking-policy. The asymptotic optimality will be proven here only for multi-class queueing networks and admission/routing problems (Theorem 7.4 and 7.5), though this procedure works in a quite general class of optimization problems. In particular, V^F provides always an asymptotic lower bound on the value functions, i.e. $\liminf_{\gamma\to\infty} V_{\pi^\gamma}^\gamma(y) \ge V^F(y)$ for all $y \in S$ (Theorem 7.3). For practical applications, we obtain a good performance with the Tracking-policy, when we have a system with large initial state which is working under a high intensity.

The content of this paper is organized as follows: Section 2 and 3 contain the theory about the β-discounted and the average cost optimality for SFP respectively. In Section 4 we have summarized results which are useful for solving SFPs in practice. In particular, we derive a HJB equation and a Verification Theorem for both criteria together with a maximum principle for the discounted cost problem. Section 5 contains some numerical tools for solving SFPs. The purely deterministic case will be dealt with separately in Section 5.1 since it is very important for the last Section 7 about asymptotic optimality as outlined before. In particular, we have refined an algorithm which is due to Pullan (1993), in order to obtain a faster convergence for our problems. For the numerical solution of the general SFP we will explain the use of Kushner's *approximating Markov chain approach* (Kushner/Dupuis (1992)) in Section 5.2. Finally, Section 6 contains three applications for SFPs which have already been presented before.

Acknowledgment
I am grateful to my teacher Ulrich Rieder for his excellent guidance and encouragement during the last years and for some profound discussions with him which I enjoyed very much. Also I would like to thank Manfred Schäl for numerous helpful comments.

2 β-Discounted Optimality

In this section we consider *Stochastic Fluid Programs* with the β-discounted optimality criterion and infinite horizon. An informal description of the evolution of such models is the following: suppose $y \in I\!R^N$ is the starting state of the system. The local dynamics of it are influenced by an external process (Z_t) which is a continuous-time Markov chain or more general, a semi-Markov process (see Remark 2.8 c). (Z_t) will be called *environment process*. As long as $Z_t = z$, the system evolves according to $y_t = y + \int_0^t b^z(u(y,z,s))\,ds$, where u is an open-loop control which has to be chosen from a set of functions and b^z is linear. The decision time points of the model are the jump times of the environment process. At these time points a whole function has to be chosen which determines the control until the next jump. The decision is Markovian i.e. it depends only on the state of the system at that time. Finally, a cost rate function c depending on the state and action is given. The expected β-discounted cost of the system over an infinite horizon has to be minimized. A rigorous definition of the model will be given in Section 2.1. As already mentioned in the introduction, Stochastic Fluid Programs are a special class of controlled *piecewise deterministic Markov processes* (see Davis (1993), Forwick (1998)) with one exception: in our model we have constraints on the actions and the process can move along the boundary of the state space. To obtain a general solution technique we will exploit the fact that the optimization problem can be reduced to a discrete-time *Markov decision process* (see Section 2.2) as has already been done in Davis (1993), Forwick (1998), Presman et al. (1995). However, it is important to note that due to some convexity assumptions we do not need the concept of *relaxed controls*. We can deal with ordinary deterministic controls which makes the theory much easier. After investigating the relaxed optimization problem in Section 2.3 we present our main theorem (Theorem 2.5) about infinite horizon β-discounted Stochastic Fluid Programs in Section 2.4. It states that under some continuity and compactness assumptions an optimal stationary policy exists which is the solution of a deterministic control problem.

2.1 Continuous-time Definition

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 manuscript 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 $z \in Z$, *linear* 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. As we will see, the linearity of the dynamics will accomodate a rich class of interesting problems. 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 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:

Definition 2.1:
The 6-tuple (E, U, b, Q, c, β) is called *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 $0 := T_0 < T_1 < T_2 < \ldots$

$$Z_t = Z_{T_n} \text{ for } T_n \leq t < T_{n+1}$$

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

and

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

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

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

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 2.2:
Let π be a policy. For $x \in E$ define by

a)
$$V_\pi(x) := E_x^\pi \left[\int_0^\infty e^{-\beta t} c(X_t, \pi_t)\, dt \right]$$
the *expected discounted cost over an infinite horizon under policy* π*, starting the system in* x.

b)
$$V(x) := \inf_\pi V_\pi(x)$$
the *minimal expected discounted cost over an infinite horizon, starting the system in* x.

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

Remark 2.1:

a) Formally one has to consider the state process $\bar{X} := (X_t, \eta_t, \tau_t)$ on the enlarged state space $E \times E \times \mathbb{R}_+$, where X_t is as before, η_t the state of the system directly after the last jump and τ_t the time elapsed since the last jump. In particular, the evolution of the system is then given by $(\phi_t(x, u), x, t)$. To ease notation, one only considers the first component. However, there will be some cases, where we will need the extended formulation. Note that for a fixed stationary policy, (X_t) is not a Markov process, whereas $(\bar{X}_t)$ is.

b) 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.

2.2 Discrete-time Formulation

We will now show that the optimization problem in Definition 2.1 can be transfered 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. However, we will see that the action space causes some difficulties.

Suppose a SFP (E, U, b, Q, c, β) as defined in the previous section is given. U is endowed with the usual Borel σ-algebra. Denote by $A := \{a : \mathbb{R}_+ \to U \mid a \text{ measurable}\}$ 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*. 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. A σ-algebra on A will be defined in Section 2.3. $F := \{f : E \to A \mid f \text{ measurable}, \ f(x) \in D(x)\}$ is called the set of *decision rules* and $\pi = (f_n)$, where $f_n \in F$ is called a *policy* in the discrete case. After adding an absorbing state Δ which makes the transition kernel p stochastic, we obtain for a fixed policy π that there exists a family of probability measures $\{\hat{P}_x^\pi \mid x \in E + \{\Delta\}\}$ on a measurable space $(\hat{\Omega}, \hat{\mathcal{F}})$ and a discrete-time stochastic process $(\hat{X}_n) = (\hat{Y}_n, \hat{Z}_n)$ on $(\hat{\Omega}, \hat{\mathcal{F}})$ such that

(i) $\hat{P}_x^\pi(\hat{X}_0 = x) = 1$ for all $x \in E$.

(ii) $\hat{P}_x^\pi(\hat{X}_{n+1} \in B \times \{z'\} \mid \hat{X}_0, \ldots, \hat{X}_n) = p(\hat{X}_n, f_n(\hat{X}_n); B \times \{z'\})$ for all $z' \in Z$ and $B \in \mathfrak{B}(S)$.

Remark 2.2:
It is important to point out that the Markov chain (X_n) as previously defined and the process (X_t) as defined in Section 2.1 are two different objects, as well as the corresponding policies. In what follows we will skip the "ˆ" in the notation. 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.

Definition 2.3:
The 6-tuple (E, A, D, p, C, β) is called the *Discretized Stochastic Fluid Program* (DSFP).

Remark 2.3:
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. For a proof see e.g. Forwick (1998) Theorem 2.2.14.

Theorem 2.1:
Let π be a policy for the SFP and σ the corresponding policy for the DSFP. Then we obtain

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

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

Proof: 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 $\{\mathcal{F}_{T_n}\}$

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

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

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

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

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

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 $\mathcal{U}v$.

Remark 2.4:

a) Let $\pi = (f_n)$ be a policy for the DSFP. Then we have $V_\pi = \lim_{n\to\infty} U_{f_0} \dots U_{f_n} 0$. The proof is similar to the one for Theorem 2.1.

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 $v \le w$ then $\mathcal{U}_f v \le \mathcal{U}_f w$ and $\mathcal{U}v \le \mathcal{U}w$.

The next aim will be to show the existence of optimal policies in the DSFP. Therefore, we have to establish several compactness and continuity properties. This will be done under

Assumption 2.1:

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

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

However, this causes some difficulties since we have to find a topology on A which guarantees that A is compact and that also some continuity properties hold. To cope with this problem we pass over to randomized actions. The action space can then be shown to be compact w.r.t. the *Young topology*. This procedure will be explained in the next section. It will turn out that our special DSFP formulation allows for the minimum to be taken in the smaller set A of deterministic actions.

2.3 A Relaxed Problem

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

$$\mathcal{R} := \{r : I\!R_+ \to I\!P(U) \mid r \text{ measurable}\}.$$

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 2.5:

a) As usual in $\mathcal{L}^p$-spaces, r should be thought of as a representative of the λ^1-equivalence class.
b) $I\!P(U)$ is endowed with the Borel-σ-algebra which is induced by the weak topology. For a characterization of measurability of functions $r : I\!R_+ \to I\!P(U)$ see Lemma A.3.
c) $A \subset \mathcal{R}$ since the elements of A can be interpreted as the Dirac measures in the set $\mathcal{R}$. In particular we obtain $\tilde{\phi}(x, \delta_a) = \phi_t(x,a),\ \tilde{C}(x,\delta_a) = C(x,a)$ and so on.

It is possible to show that $\mathcal{R}$ is compact w.r.t. the Young-topology and $\mathcal{R}$ is metrizable. For definition of the Young-topology and a proof of these results we refer the reader to Davis (1993) Section 4.3 or Forwick (1998) chapter 2. The following Lemma will now be crucial.

Lemma 2.2:
Let a relaxed DSFP $(E, \mathcal{R}, \tilde{D}, \tilde{p}, \tilde{C}, \beta)$ be given. Under Assumption 2.1 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 or Forwick (1998) Theorem 2.2.6.

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.
Analoguously 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 $\tilde{D}$ is closed.

c) and

d) see e.g. Davis (1993) Theorem 44.11 or Forwick (1998) Theorem 2.2.11.

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 $x \in \psi^{-1}[B]$. □

Remark 2.6:
The one-step cost function C and the transition kernel p depend by definition on the interest rate β. To make the dependence explicit we will sometimes write C^β and p^β. It can easily be shown that even

(i) $(x,r,\beta) \mapsto \tilde{C}^\beta(x,r)$ is lower semicontinuous.

(ii) $(x,r,\beta) \mapsto \int_E v(x')\tilde{p}^\beta(x,r;dx')$ is continuous and bounded for every continuous, bounded function $v : E \to I\!\!R$.

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 2.3:
Let a DSFP be given and $v \in \mathfrak{C}_{lsc}$. Under Assumption 2.1 there exists an $f^* \in F$ such that

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

and $\mathcal{U}v \in \mathfrak{C}_{lsc}$.

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 A.1 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 $a_t = \int_U ur_t(du),\ t \geq 0$. Since U is convex, $a_t \in U$ for all $t \geq 0$ (see e.g. Hinderer (1984) Theorem 25.10) and is measurable due to Lemma A.3, hence $a \in A$. Moreover, since b^z is linear

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

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

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

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

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

Then $f^* : E \to A$ is measurable (see Lemma A.3) 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. □

2.4 β-Discounted Cost Optimality Equation

We will first consider the finite-horizon optimization problem for a DSFP. Define for a policy $\pi = (f_0, \ldots, f_{n-1})$, $n \in \mathbb{N}$ and $x \in E$, the *expected discounted cost over n-stages using policy π and starting in x* by

$$V_{n\pi}(x) := E_x^\pi \left[\sum_{k=0}^{n-1} C\left(X_k, f_k(X_k)\right) \right].$$

And the *minimal expected discounted cost over* n*-stages starting in* x by

$$V_n(x) := \inf_{\pi} V_{n\pi}(x)$$

where the infimum is taken over all policies π. In the sequel we will always set $V_0 := 0$. The proof of the next lemma follows e.g. along the lines of Rieder (1994), see also Hernández-Lerma/Lasserre (1996).

Lemma 2.4: *(Value iteration)*
With the preceding definitions we obtain under Assumption 2.1

a) $V_n(x) = \mathcal{U}V_{n-1}(x) = \mathcal{U}^n 0,\ n \in I\!N$.
b) If f_k is a minimizer of $V_{k-1},\ k = 1, \ldots, n$ which exists, then the policy $\pi = (f_n, \ldots, f_1)$ is optimal for the n-stage optimization problem.

We will now return to our infinite-horizon optimization problem. The following assumption is needed.

Assumption 2.2:
There exists a policy π such that $V_\pi(x) < \infty$ for all $x \in E$.

Theorem 2.5: *(β-Discounted cost optimality equation)*
Suppose that the Assumptions 2.1 and 2.2 hold. Then

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

$$V(x) = \min_{a \in D(x)} \left[C(x,a) + \int_0^\infty e^{-(\beta+q_z)t} \sum_{z' \neq z} q_{zz'} V\left(\phi_t(x,a), z'\right)\, dt \right]. \qquad (2.1)$$

b) There exists a minimizer $f^* \in F$ of V in (2.1) and the stationary policy $(f^*, f^*, \ldots)$ is β-discounted optimal.

Remark 2.7:
Let $f \in F$. For the stationary policy $\pi = (f, f, \ldots)$ we write $f^\infty := \pi$ and $V_f := V_\pi$. Moreover, for $\pi = (f, \ldots, f)$ we denote $V_{nf} := V_{n\pi}$.

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

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

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

and since the operator $\mathcal{U}$ is monotone we have $V_n \uparrow \hat{V} \leq V$. From Lemma A.2 (interchange of min and lim) together with Theorem 2.3 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 2.3 that there exists a decision rule f^* which attains the infimum in $\hat{V} = \mathcal{U}\hat{V}$. Thus we obtain

$$\hat{V} = \mathcal{U}^n_{f^*}\hat{V} \geq \mathcal{U}^n_{f^*} 0 = V_{nf^*}$$

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

Remark 2.8:

a) A natural question that arises when reading Section 2.1 is why the policies have been defined in a discrete way. A natural candidate for a policy would be a measurable mapping $\pi_t : H_t \to U,\ t \geq 0$, 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 $t \geq 0$. However, it is known that Theorem 2.5 remains valid if we would minimize over all policies $\pi = (f_n)$ such that f_{n+1} depends on the history $h_n = 0x_0f_1t_1x_1 \dots f_nt_nx_n,\ n \in I\!N$. Thus in terms of Yushkevich (1980) Theorem 2.5 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

$$p^{SM}(x, a; B \times \{z'\}) = p_{zz'} \int_0^\infty e^{-\beta t} f_{zz'}(t) 1_B(\phi_t(x, a))\, dt$$

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

All other data remains the same. In particular the optimality equation (2.1) is now of the form

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

2.5 Properties of the Value Function

Suppose a SFP as defined in Section 2.1 is given and Assumptions 2.1 and 2.2 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 2.6:
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 2.1 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 starting 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 starting 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. □

We will often need the following growth assumption on the cost rate function c

Assumption 2.3:
There exist constants $k \in \mathbb{N}$ and $C_0 \in \mathbb{R}_+$ such that for all $z \in Z, u, u' \in U$ and $y, y' \in S$

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

Lemma 2.7:
If $S = \mathbb{R}^N$ and $y \mapsto c(y,z,u)$ is continuous for all $u \in U, z \in Z$ and fulfills Assumption 2.3 then $V(y,z)$ is Lipschitz-continuous in y.

Proof: Let $y, y', h \in \mathbb{R}^N$. Moreover, denote by (π_t) the process of the optimal policy for starting in $y + h$. Due to our assumptions, (π_t) is also admissible for starting in y. Hence we obtain

$$\begin{aligned} V(y,z) - V(y+h,z) &\le V_\pi(y,z) - V_\pi(y+h,z) \\ &= E^\pi_{y+h}\left[\int_0^\infty e^{-\beta t}\Big(c(Y_t - h, Z_t, \pi_t) - c(Y_t, Z_t, \pi_t)\Big)\,dt\right] \\ &\le C_0\|h\| E^\pi_{y+h}\left[\int_0^\infty e^{-\beta t}\Big(1 + \|Y_t - h\|^k + \|Y_t\|^k\Big)\,dt\right] \end{aligned}$$

where the last term tends to zero if $\|h\| \to 0$ since $\|Y_t\| = O(t)$. With the same arguments one can show that $V(y,z) - V(y+h,z)$ has a lower bound which tends to zero as $\|h\| \to 0$ and the statement is proven. □

Lemma 2.8:
If $S = \mathbb{R}^N$ and $y \mapsto c(y,z,u)$ is continuously differentiable and convex for all $u \in U, z \in Z$ and fulfills Assumption 2.3 then $V(y,z)$ is continuously differentiable w.r.t. y.

Proof: Since V is convex due to Lemma 2.6 it suffices to show that the partial derivatives exist (cf. Rockafellar (1970)). Let $y, y' \in \mathbb{R}^N$ and $h > 0$. By e_ν we denote the ν-th unit vector. The convexity of c implies the convexity of V (Lemma 2.6), hence

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

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

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

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

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 2.3 for $|h|$ small enough

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

since the trajectories can grow at most linearly. 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

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

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

which implies the statement. $\square$

3 Average Optimality

We will now deal with the minimization of the long-term average cost of SFPs. As in the preceding section we are interested in the existence of optimal policies and an optimization principle. In continuous time, the average cost can be defined in different ways. The most natural one is to take

$$G_\pi(x) = \limsup_{t\to\infty} \frac{1}{t} E_x^\pi \left[\int_0^t c(X_s, \pi_s)\, ds \right].$$

We will define the average cost optimization problem for a uniformized version of the SFP. The reason is that we were not able to derive an optimality principle for the non-uniformized SFP. However, when the optimal policy is given by a feedback control, the minimal average cost in both models are the same. The main theorem of this section (Theorem 3.6) states the existence of an average cost optimal policy under some assumptions which are the same as the ones introduced by Sennott (1989a) for discrete-time Markov decision processes. In addition it can be shown that an accumulation point of β-discounted policies for $\beta \downarrow 0$ is average optimal. As in Section 2 we solve our problem by discretization. Unfortunately, this is more complicated here, since it is not clear whether the average cost of a policy in the continuous and the discrete setting are the same in general. However, we will give some conditions which imply the equivalence. The proof of Theorem 3.6 uses the vanishing discount approach and follows essentially as in Schäl (1993). There are only a few papers dealing with average cost for piecewise deterministic processes. Hordijk/Van der Duyn Schouten (1983) investigate the average cost problem for Markov decision drift processes. In Sethi et al. (1997, 1998) one can find special production models under the average cost criterion. However, their approach is in a continuous setting and does not deal with the question of existence of optimal policies. The section is organized as follows: in Section 3.1 we introduce the uniformized SFP and give two definitions of average cost. In the following section we investigate the relation of the stationary distributions of the continuous and the discrete model. Section 3.3 contains the main theorem and the validity of the average cost optimality inequality. Since the assumptions of Theorem 3.6 are quite technical we will give some sufficient conditions for them. Mainly these conditions imply positive Harris recurrence of the controlled state process. Finally Section 3.4 deals with the validity of the average cost optimality equation.

3.1 Definition of Average Optimality

The model we consider here is a slight modification of the model in Section 2. The difference is that we consider a *uniformized* environment process (Z_t), i.e. let

$q > \max_{z\in Z} q_z$ and $P = I + \frac{1}{q}Q$. Then (Z_t) can be constructed from a sequence (T_n) of jump times, where the random variables $(T_{n+1} - T_n),\ n \in \mathbb{N}$ are independent and exponentially distributed with parameter q and from a Markov chain (Λ_n) with transition matrix P as follows. Let $\Lambda_0 := Z_0,\ T_0 := 0$ and $t \geq 0$. Then

$$Z_t = \Lambda_n, \quad \text{if } T_n \leq t < T_{n+1}$$

is in distribution equal to a Markov chain with generator Q (cf. Bertsekas (1995)). Formally we now add further artificial jump time points to the system. Decisions have to be taken in a Markovian fashion at times T_n whether or not a real jump has occured. When a SFP is given as in Section 2.1, we will refer to its modification as the uniformized SFP. In the discrete setting, the transition kernel p and and the one-step cost function C for the uniformized β-discounted model are now given by

$$p(x, a; B \times \{z'\}) = q\, p_{zz'} \int_0^\infty e^{-(\beta+q)t} 1_B\left(\phi_t(x,a)\right)\, dt$$

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

The operator $\mathcal{U}$ for $v : E \to \mathbb{R}_+$ reads

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

Of course all the previous theorems remain valid in this case and we will use the same definitions and notations. Before we define the average cost problem for this model, we will shortly investigate its relation to the problem in Section 2 in the β-discounted case. In particular, the next theorem states that the value function of the uniformized SFP is independent of the parameter q. The statement follows from Theorem 2 in Yushkevich (1980).

Theorem 3.1:
Under Assumptions 2.1 and 2.2 the value functions of the uniformized SFP and the non-uniformized SFP coincide in the β-discounted problem.

Let us now return to the problem of defining average costs in the uniformized SFP. There is certainly more than one reasonable definition for average optimality for continuous-time processes (see e.g. Sennott (1989b), Ross (1970)). But the most appealing one is the following:

Definition 3.1:
Let a uniformized SFP be given with $\beta = 0$ and let π be a policy. For $x \in E$ define by

a)
$$G_\pi(x) = \limsup_{t\to\infty} \frac{1}{t} E_x^\pi \left[\int_0^t c(X_s, \pi_s)\, ds \right]$$
the *average cost under policy* π*, starting the system in* x*.*

b)
$$G(x) = \inf_\pi G_\pi(x)$$
the *minimal average cost, starting the system in* x*.*

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

As in the discounted case we will tackle this problem via discretization. It is important to note that the transition kernel p and the one-step cost function C in the average case are defined by

$$p(x, a; B \times \{z'\}) = q\, p_{zz'} \int_0^\infty e^{-qt} 1_B \left(\phi_t(x, a)\right)\, dt$$

$$C(x, a) := \int_0^\infty e^{-qt} c(\phi_t(x, a), z, a_t)\, dt.$$

To avoid confusions we write in this section p^β, C^β for the quantities in the β-discounted model. The average cost for a DSFP are now defined in an obvious way by:

Definition 3.2:
Let a uniformized DSFP be given with $\beta = 0$ and let $\pi = (f_n)$ be a policy. For $x \in E$ define by

a)
$$J_\pi(x) = \limsup_{m\to\infty} \frac{E_x^\pi \left[\sum_{n=0}^{m-1} C(X_n, f_n(X_n)) \right]}{E_x^\pi [T_m]}$$
the *average cost under policy* π*, starting the system in* x*.*

b)
$$J(x) = \inf_\pi J_\pi(x)$$
the *minimal average cost, starting the system in* x*.*

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

Remark 3.1:

a) If $\pi = f^\infty$ is a stationary policy we write G_f and J_f.

b) Since the sequence $(T_{n+1} - T_n)$ is iid with mean $\frac{1}{q}$ we obviously have

$$J_\pi(x) = q \limsup_{m\to\infty} \frac{1}{m} E_x^\pi \left[\sum_{n=0}^{m-1} C(X_n, f_n(X_n)) \right]$$

3.2 Stationary Distributions

In this section we choose a fixed stationary feedback policy $\pi = f^\infty$ and will hence suppress the dependence on π in our notation. It is possible to obtain a connection between the stationary distribution of the continuous-time process (X_t) and the discrete-time process (X_n) of the uniformized model. For $t \geq 0,\ B \in \mathfrak{B}(S),\ z' \in Z$ and $x \in E$ let us denote by

$$p_t(x; B \times \{z'\}) = P_x(X_t \in B \times \{z'\})$$

the transition kernel of (X_t). As before $p(x; B \times \{z'\})$ is the one-step transition probability for (X_n), i.e.

$$p(x; B \times \{z'\}) = q\, p_{zz'} \int_0^\infty e^{-qt} 1_B(\phi_t(x))\, dt.$$

Definition 3.3:

a) A distribution μ on E is a *stationary distribution* of (X_t) if for all $t \geq 0,\ A \in \mathfrak{B}(E)$

$$\mu(A) = \int_E p_t(x; A)\, \mu(dx).$$

b) A distribution ν on E is a *stationary distribution* of (X_n) if for all $A \in \mathfrak{B}(E)$

$$\nu(A) = \int_E p(x; A)\, \nu(dx).$$

An important role in determining the stationary distribution, plays the generator $\mathcal{A}$ of (X_t). $\mathcal{A}$ acts on the set of functions

$$\begin{aligned}\mathcal{D}(\mathcal{A}) \;=\; & \{v : E \to I\!R \mid v \text{ measurable, bounded }, t \mapsto v(\phi_t(x), z) \text{ is absolutely} \\ & \text{continuous for all } x \in E\}\end{aligned}$$

and a version of it is given by

$$\mathcal{A}v(x) = \lim_{t\to 0} \frac{1}{t}(E_x[v(X_t)] - v(x)) = \tilde{v}(x) + q \sum_{z'} p_{zz'}(v(y, z') - v(x))$$

where $\tilde{v} : E \to I\!R$ is such that

$$v(\phi_t(x), z) - v(x) = \int_0^t \tilde{v}(\phi_s(x), z)\, ds.$$

The following theorem states that (X_t) has a stationary distribution if and only if (X_n) has one and gives a transformation formula for them. See e.g. Costa (1990) or Davis (1993).

Theorem 3.2:

a) Let ν be a stationary distribution of (X_n), then μ defined for $B \in \mathfrak{B}(S),\ z \in Z$ by

$$\mu(B \times \{z\}) = \int_0^\infty q e^{-qt} \nu(\{y \mid \phi_t(x) \in B\} \times \{z\})\, dt$$

is a stationary distribution of (X_t).

b) Let μ be a stationary distribution of (X_t), then ν defined for $B \in \mathfrak{B}(S),\ z \in Z$ by

$$\nu(B \times \{z\}) = \sum_{z'} p_{zz'} \mu(B \times \{z'\})$$

is a stationary distribution of (X_n).

Proof:

a) Obviously μ is a distribution. It holds that μ is a stationary distribution if and only if $\int_E \mathcal{A}v(x)\mu(dx) = 0$ for all $v \in \mathcal{D}(\mathcal{A})$ (cf. Ethier/Kurtz (1986) Proposition 9.2). Let $v \in \mathcal{D}(\mathcal{A})$ be arbitrary. We obtain

$$\begin{aligned}
& \int_0^\infty \mathcal{A}v(\phi_t(x), z) e^{-qt}\, dt \\
= & \int_0^\infty \tilde{v}(\phi_t(x), z) e^{-qt}\, dt + \int_0^\infty q e^{-qt} \sum_{z'} p_{zz'} \left(v(\phi_t(x), z') - v(\phi_t(x), z)\right)\, dt \\
= & \int_0^\infty \frac{d}{dt}\left(v(\phi_t(x), z) e^{-qt}\right)\, dt + \int_0^\infty q e^{-qt} \sum_{z'} p_{zz'} v(\phi_t(x), z')\, dt \\
= & E_x[v(X_{T_1})] - v(x).
\end{aligned}$$

Since ν is a stationary distribution for (X_n) we have

$$\int_E \left[E_x[v(X_{T_1})] - v(x)\right]\ \nu(dx) = 0.$$

This yields

$$0 = \sum_z \int_S \int_0^\infty \mathcal{A}v(\phi_t(y, z), z) e^{-qt}\, dt\, \nu(dy, z) = \frac{1}{q} \int_E \mathcal{A}v(x)\, \mu(dx)$$

and the proof of part a) is complete.

b) From the definition of μ we get that ν is a distribution. For $A \in \mathfrak{B}(E)$ we have to show $\nu(A) = \int_E p(x;A)\nu(dx)$. Since μ is stationary, $\int_E \mathcal{A}v(x)\,\mu(dx) = 0$ for all $v \in \mathcal{D}(\mathcal{A})$.
We will first show that $p(x;A) \in \mathcal{D}(\mathcal{A})$ for all $A \in \mathfrak{B}(E)$. Let $A = B \times \{z'\}$, $x = (y,z)$. Since

$$p(x;A) = e^{-qt}p((\phi_t(x),z);A) + \int_0^t qe^{-qs}1_B(\phi_s(x))\,ds\;p_{zz'}$$

$p((\phi_t(x),z);A)$ is obviously absolutely continuous and we obtain at points where $1_B(y)$ is continuous

$$\lim_{t\downarrow 0}\frac{1}{t}\left(p((\phi_t(x),z);A) - p(x;A)\right) = \lim_{t\downarrow 0}\frac{1}{t}(e^{qt}-1)p(x;A) -$$

$$-\lim_{t\downarrow 0}\frac{1}{t}e^{qt}\int_0^t qe^{-qs}1_B(\phi_s(x))\,ds\;p_{zz'} = q(p(x;A) - 1_B(y)p_{zz'}).$$

Hence $p(x;A) \in \mathcal{D}(\mathcal{A})$ and

$$\mathcal{A}\,p(x;A) = q(\sum_w p_{zw}p(y,w;A) - 1_B(y)p_{zz'}).$$

Therefore

$$\int_E \mathcal{A}p(x;A)\mu(dx) = q(\sum_z \int_S \sum_w p_{zw}p(y,w;A)\mu(dy,z) - q(\sum_z \int_S 1_B(y)p_{zz'}\mu(dy;z))$$

which yields

$$\int_E p(x;A)\,\nu(dx) = \sum_z p_{zz'}\int_B \mu(dy,z) = \nu(A)$$

and the proof is complete. □

For a function $c: E \to I\!\!R_+$ we obtain now

$$\begin{aligned}\int_E c(x)\,\mu(dx) &= \sum_z \int_S \int_0^\infty c(\phi_t(x),z)qe^{-qt}\,dt\,\nu(dy,z)\\ &= q\sum_z \int_S C(x)\,\nu(dy,z) = q\int_E C(x)\,\nu(dx)\end{aligned}$$

which means that the expected cost in a stationary regime of an SFP and the corresponding DSFP differ only by a factor q.

Remark 3.2:
When we are concerned with general controlled processes we usually have piecewise-open-loop policies. Therefore we have to consider the enlarged process $\bar{X}_t =$

(X_t, η_t, τ_t) on the state space $\bar{E} = E \times S \times I\!\!R_+$ (cf. Remark 2.1). Every piecewise-open-loop policy is now a feedback control of the new state process. Theorem 3.2 can be shown in the same way for the process $\bar{X}_t$ and its discretized version (see e.g. Costa (1990), Davis (1993)). If $u : E \times I\!\!R_+ \to U$ is admissible, we obtain with $\bar{x} = (x, \eta, \tau)$

$$\int_{\bar{E}} c(x, u(\eta, \tau))\, \mu(d\bar{x}) = q \int_{\bar{E}} C(x, u(x, \cdot))\, \nu(d\bar{x})$$

where $C : E \times A \to I\!\!R_+$ is defined by

$$C(x, u(x, \cdot)) = \int_0^\infty e^{-qt} c(\phi_t(x, u), z, u(x, t))\, dt.$$

3.3 Average Cost Optimality Inequality

In this section we will prove the existence of a c- and d-average cost optimal stationary policy. Let us define for $x \in E$, $\beta > 0$ and a fixed state $\xi \in E$

$$h^\beta(x) = V^\beta(x) - V^\beta(\xi) \quad \text{and} \quad \rho(\beta) = \beta V^\beta(\xi),$$

where V^β is the value function of the β-discounted model. h^β is called *relative value function*. Assumption 3.1 has been established by Sennott (1989a) for Markov decision processes with a countable state space.

Assumption 3.1:

(i) There exists a policy π such that $G_\pi(x) < \infty$ for all $x \in E$.

(ii) There exist constants $L \in I\!\!R$, $\bar{\beta} > 0$ and a function $M : E \to I\!\!R_+$ with

$$L \le h^\beta(x) \le M(x)$$

for all $x \in E$ and $0 < \beta \le \bar{\beta}$.

The following Tauber Theorem will be a useful tool. Versions of it can be found e.g. in Hordijk/Van der Duyn Schouten (1983) and Sethi et al. (1997).

Theorem 3.3:
For all policies π and $x \in E$ we obtain

$$\limsup_{\beta \downarrow 0} \beta V_\pi^\beta(x) \le G_\pi(x)$$

Applying the Tauber Theorem we immediately obtain the following two lemmas.

Lemma 3.4:
Under Assumption 3.1 there exists a sequence of interest rates $\beta_n \downarrow 0$ such that for all $x \in E$

$$0 \leq \lim_{n\to\infty} \beta_n V^{\beta_n}(x) = \limsup_{\beta\downarrow 0} \rho(\beta) < \infty$$

Proof: From the Tauber Theorem 3.3 we obtain with Assumption 3.1 (i)

$$0 \leq \rho := \limsup_{\beta\downarrow 0} \rho(\beta) \leq G_\pi(\xi) < \infty.$$

Let $\beta_n \downarrow 0$ be a sequence such that $\lim_{n\to\infty} \rho(\beta_n) = \rho$. For $x \in E$, Assumption 3.1 (ii) yields

$$\begin{aligned} \left|\beta_n V^{\beta_n}(x) - \rho\right| &\leq \beta_n \left|h^{\beta_n}(x)\right| + |\rho(\beta_n) - \rho| \\ &\leq \beta_n \max\{L, M(x)\} + |\rho(\beta_n) - \rho|. \end{aligned}$$

The right hand side converges to 0 as $n \to \infty$ which implies the result. □

Lemma 3.5:
Under Assumption 3.1 it holds for all policies π and $x \in E$ that

$$\limsup_{\beta\downarrow 0} \rho(\beta) \leq G_\pi(x)$$

Now we are able to prove the main theorem of this section. The proof is along the lines of Schäl (1993) who established the existence of average optimal policies in discrete-time Markov decision processes with Borel state space.

Theorem 3.6: (*Average cost optimality inequality*)
Suppose that the Assumptions 2.1 2.2 and 3.1 hold. Then

a) There exists a constant $\rho \geq 0$ and a lower semicontinuous function $h : E \to \mathbb{R}$ such that the *average cost optimality inequality* holds, i.e. for all $x \in E$

$$\frac{\rho}{q} + h(x) \geq \min_{a\in D(x)} \left[C(x,a) + q \int_0^\infty e^{-qt} \sum_{z'} p_{zz'} h\left(\phi_t(x,a), z'\right)\, dt \right]. \tag{3.2}$$

Moreover, there exists a minimizer f^* of h in (3.2).

b) Suppose that $J_{f^*} \geq G_{f^*}$. Then the stationary policy $(f^*, f^*, \ldots)$ is c-average optimal and $\rho = \lim_{\beta \downarrow 0} \rho(\beta)$ are the minimal average cost, independent of x. Moreover, there exists a decision rule f^0 and sequences $\beta_m(x) \to 0$, $x_m(x) \to x$ such that
$$f^0(x) = \lim_{m\to\infty} f^{\beta_m(x)}(x_m(x)),$$
where f^β is an optimal decision rule in the β-discounted model and the stationary policy $(f^0, f^0, \ldots)$ is c-average optimal, provided $J_{f^0} \geq G_{f^0}$.

Proof: a),b) Define $\rho = \liminf_{\beta \downarrow 0} \rho(\beta) \geq 0$ which is finite because of Lemma 3.4. Take $\beta(n)$ as the subsequence such that $\rho = \lim_{n\to\infty} \rho(\beta(n))$. Define

$$h(x) := \liminf_{n\to\infty,\, x'\to x} h^{\beta(n)}(x') = \lim_{n\to\infty} \inf_{k\geq n} \inf_{d(x,x')\leq \frac{1}{n}} h^{\beta(k)}(x')$$

where d is a metric on E. Due to Assumptions 2.1 and 2.2 the β-discounted optimality equation holds and we can write it in the following form for $\beta > 0,\ x \in E$

$$\frac{\rho(\beta)}{\beta+q} + h^\beta(x) = C^\beta(x, f^\beta(x)) + q \int_0^\infty e^{-(\beta+q)t} \sum_{z'} p_{zz'} h^\beta\left(\phi_t(x, f^\beta(x)), z'\right) dt, \quad (3.3)$$

where f^β is the optimal decision rule in the β-discounted model. From Schäl (1993) Lemma 3.4 we know that there exist sequences $\{k_n\}$ of integer-valued measurable mappings and $\{x_n\}$ of E-valued measurable mappings on E such that $k_n(x) \to \infty$, $x_n(x) \to x$ for $n \to \infty$ and $h^{\beta(k_n(x))}(x_n(x)) \to h(x)$. Define $a_n(x) = f^{\beta(k_n(x))}(x_n(x))$, $x \in E$. We will now fix $x \in E$ and suppress the dependence on x in our notation. Then by (3.3)

$$\begin{aligned} &\rho(\beta(k_n))/(\beta(k_n)+q) + h^{\beta(k_n)}(x_n) = \\ &C^{\beta(k_n)}(x_n, a_n) + q \int_0^\infty e^{-(\beta(k_n)+q)t} \sum_{z'} p_{zz'} h^{\beta(k_n)}\left(\phi_t(x_n, a_n), z'\right) dt. \end{aligned} \quad (3.4)$$

Moreover, we know from Schäl (1975) that there exists a measurable function $g^0 : E \to \mathcal{R}$ such that $g^0(x)$ is an accumulation point of $\{a_n(x)\}$ and $g^0 \in F$. For a fixed $x \in E$ choose a subsequence $\{n_m\}$ of natural numbers (for simplicity denoted by m) such that $a_m(x) \to g^0(x)$. Taking $m \to \infty$ we obtain with the lower semicontinuity of C^β (see Remark 2.6)

$$\liminf_{m\to\infty} C^{\beta(k_m(x))}(x_m(x), a_m(x)) \geq C(x, g^0(x)).$$

For the next step, observe that multiplying $p^\beta(x, a; B)$ by $\frac{\beta+q}{q}$ makes the transition kernel stochastic. Since $\frac{\beta+q}{q} p^\beta(x, a; \cdot)$ is weakly continuous (see Remark 2.6) we obtain with Lemma 2.3 (ii) in Schäl (1993), cf. also Serfozo (1982)

$$\liminf_{m\to\infty} \int_E h^{\beta(k_m(x))}(x') p^{\beta(k_m(x))}(x_m(x), a_m(x); d\,x')$$

$$\geq \int_E \left[\liminf_{m\to\infty, x''\to x'} h^{\beta(k_m(x))}(x'') \right] p(x, g^0(x); d\,x') \geq \int_E h(x')p(x, g^0(x); d\,x')$$

Hence taking $\liminf_{m\to\infty}$ in (3.4) we obtain altogether

$$\frac{\rho}{q} + h(x) \geq C(x, g^0(x)) + \int_E h(x')p(x, g^0(x); d\,x')$$

$$\geq C(x, g^*(x)) + \int_E h(x')p(x, g^*(x); d\,x')$$

where g^* is the minimizer of h which exists since h is lower semicontinuous (see Schäl (1993)). As in the proof of Theorem 2.3 we can now define $f^0(x)(t) = \int_U ug^0(x)(t, du)$ (respectively $f^*(x)(t) = \int_U ug^*(x)(t, du)$) which is in F and due to the convexity of c

$$\frac{\rho}{q} + h(x) \geq C(x, g^0(x)) + \int_E h(x')p(x, g^0(x); d\,x')$$

$$\geq C(x, f^0(x)) + \int_E h(x')p(x, f^0(x); d\,x')$$

Iterating this inequality m times one gets

$$m\frac{\rho}{q} + h(x) \geq E_x^{f^0}\left[\sum_{n=0}^{m-1} C(X_n, f^0(X_n))\right] + E_x^{f^0}\left[h(X_m)\right].$$

Assumption 3.1 (ii) yields that h is finite and $h \geq L$. Hence we obtain by dividing through m and taking limit $m \to \infty$

$$\rho \geq J_{f^0}(x) \geq G_{f^0}(x) \geq \limsup_{\beta\downarrow 0} \beta V_{f^0}^{\beta}(x) \geq \limsup_{\beta\downarrow 0} \rho(\beta) \geq \rho$$

where the second inequality follows from the assumption and the third inequality from Theorem 3.3. The last but one inequality is a consequence of Lemma 3.4. Hence we have equality. Using Lemma 3.5 this implies that the stationary policy $(f^0, f^0, \ldots)$ is c-average optimal and the minimal average cost are $\rho = \lim_{\beta\downarrow 0} \rho(\beta)$. The same holds for g^* which completes the proof. □

To show that $(f^0, f^0, \ldots)$ and $(f^*, f^*, \ldots)$ from Theorem 3.6 are also d-average optimal we need some further assumptions.

Theorem 3.7:
Suppose that the Assumptions 2.1 2.2 and 3.1 are valid. Moreover, we assume

(i) If $y \in S$ then $\lambda y \in S$ for $\lambda \in [0, 1]$.
(ii) M of Assumption 3.1 is locally uniformly bounded.

(iii) The cost rate function c satisfies

$$c(\lambda y, z, u) \leq K(\lambda) c(x, u)$$

for $\lambda \in (\underline{\lambda}, 1)$ with $0 < \underline{\lambda} < 1$ and a function $K : I\!R_+ \to I\!R$ with $K(\lambda) \to 1$ if $\lambda \uparrow 1$.

Then the stationary policies $(f^*, f^*, \ldots)$ and $(f^0, f^0, \ldots)$ of Theorem 3.6 are also d-average optimal and ρ are the minimal average cost independent of x.

Proof: The proof proceeds by using a coupling argument. Denote by $(\hat{Z}_t)$ a continuous-time Markov chain with generator $\frac{\beta+q}{q}Q$ and uniformization parameter $\beta + q$. Therefore we have $(\hat{Z}_t) \stackrel{d}{=} (Z_{t\frac{\beta+q}{q}})$ and in particular the jump times satisfy $\hat{T}_n \stackrel{d}{=} \frac{q}{\beta+q} T_n,\ n \in I\!N$. For an arbitrary decision rule $f \in F$, let $\hat{f}(\frac{q}{\beta+q}y, z)(t) := f(x)(t\frac{\beta+q}{q})$ for $t \geq 0,\ x \in E$. Thus for $x \in E$

$$\begin{aligned} \phi_t(\frac{q}{\beta+q}y, z, \hat{f}) &= \frac{q}{\beta+q}y + \int_0^t b^z(\hat{f}(\frac{q}{\beta+q}y, z)(s))\, ds \\ &= \frac{q}{\beta+q}y + \frac{q}{\beta+q}\int_0^{t\frac{\beta+q}{q}} b^z(f(x)(s))\, ds = \frac{q}{\beta+q}\phi_{t\frac{\beta+q}{q}}(x, f) \end{aligned}$$

Hence $\hat{Y}_t = \frac{q}{\beta+q} Y_{t\frac{\beta+q}{q}}$ for all $t \geq 0$ and $\hat{Y}_t \in S$ due to our Assumption (i). Moreover, using Assumption (iii) we get

$$\begin{aligned} \int_{\hat{T}_n}^{\hat{T}_{n+1}} c(\hat{Y}_t, \hat{Z}_t, \hat{\pi}_t)\, dt &= \frac{q}{\beta+q}\int_{T_n}^{T_{n+1}} c(\frac{q}{\beta+q}Y_t, Z_t, \pi_t)\, dt \\ &\leq \frac{q}{\beta+q}K(\frac{q}{\beta+q})\int_{T_n}^{T_{n+1}} c(Y_t, Z_t, \pi_t)\, dt \end{aligned}$$

If we define for $v : E \to I\!R_+,\ f \in F$, the operators

$$\tilde{\mathcal{U}}_f(x) = C^0(x, f) + \frac{q}{\beta+q}\int_0^\infty e^{-qt} q \sum_{z'} p_{zz'} v(\phi_t(x, f), z')\, dt$$

$$\mathcal{U}_f(x) = C^\beta(x, f) + \frac{q}{\beta+q}\int_0^\infty e^{-(\beta+q)t}(\beta+q)\sum_{z'} p_{zz'} v(\phi_t(x, f), z')\, dt,$$

we obtain from Theorem 3.6

$$\begin{aligned} \rho &\geq J_{f^0}(x) = q \limsup_{N\to\infty} E_x^{f^0}\left[\sum_{n=0}^{N-1} C(X_n, f^0(X_n))\right] \\ &\geq q \limsup_{\alpha\to 1}(1-\alpha) E_x^{f^0}\left[\sum_{n=0}^{\infty} \alpha^n C(X_n, f^0(X_n))\right] \\ &= q \limsup_{\beta\to 0}\frac{\beta}{\beta+q} E_x^{f^0}\left[\sum_{n=0}^{\infty}(\frac{q}{\beta+q})^n C(X_n, f^0(X_n))\right] = q \limsup_{\beta\to 0}\frac{\beta}{\beta+q}\tilde{U}_{f^0}^\infty 0(x) \\ &\geq \limsup_{\beta\to 0}\frac{\beta}{K(\frac{q}{\beta+q})} U_{\hat{f}}^\infty 0(\frac{q}{\beta+q}y, z) \geq \limsup_{\beta\downarrow 0}\beta V^\beta(\frac{q}{\beta+q}y, z) \geq \rho \end{aligned}$$

The second inequality follows from the Tauber Theorem in Sznajder/Filar (1992). The last inequality follows as in Lemma 3.4 using the assumption on M. Hence we have equality in the preceding expression. In particular, when we omit the first inequality and apply the same arguments to an arbitrary policy π we obtain $J_\pi(x) \geq \rho$. Hence $J \geq \rho$ and the d-average optimality of f^0 follows. The same proof can be used for $(f^*, f^*, \ldots)$. $\square$

Assumption 3.1 is often difficult to verify directly. However, we can give some sufficient conditions which will prove extremely useful in our applications. Besides, it is possible to provide conditions which yield that $J_f = G_f$. For the next lemma suppose that $c \geq 1$, otherwise replace c by $c + 1$.

Lemma 3.8:
Suppose that Assumptions 2.1 and 2.2 are valid and that there exists a decision rule $f \in F$ such that for $\pi = f^\infty$ there exists a state $\xi \in E$ with

$$E_x^f \left[\int_0^{\tau_\xi} c(X_t, \pi_t) \, dt \right] < \infty \tag{3.5}$$

for all $x \in E$, where $\tau_\xi = \inf\{t \geq 0 \mid X_t = \xi\}$. Then there exists a constant $\bar{\beta} > 0$ and a function $M : E \to I\!R_+$ such that for all $x \in E$ and $0 < \beta < \bar{\beta}$

$$h^\beta(x) = V^\beta(x) - V^\beta(\xi) \leq M(x).$$

Moreover, if (3.5) holds for the stopping time $\sigma_\xi = \inf\{T_n \,|\, X_{T_n} = \xi\}$, then $G_f(x) = J_f(x) < \infty$ for all $x \in E$.

Proof: Let $\pi^\beta = (f^\beta)^\infty$ be the optimal stationary policy for the β-discounted model and denote by (π_t^β) the process of the optimal control, starting in ξ. (π_t) is the process of the control starting in x under policy π. Now define for $t \geq 0$

$$\tilde{\pi}_t^\beta = \begin{cases} \pi_t & , \text{if } t < \tau_\xi \\ \pi_{t-\tau_\xi}^\beta & , \text{if } t \geq \tau_\xi \end{cases}$$

For arbitrary $\beta > 0$ we obtain for $x \in E$ (cf. Remark 2.8a))

$$V^\beta(x) \leq V_{\tilde{\pi}^\beta}^\beta(x) \leq E_x^f \left[\int_0^{\tau_\xi} c(X_t, \pi_t) \, dt \right] + V^\beta(\xi).$$

Hence we can define $M(x) := E_x^f \left[\int_0^{\tau_\xi} c(X_t, \pi_t) \, dt \right]$ which is finite due to our assumption.

If (3.5) holds for σ_ξ, then denoting by N_ξ the number of jumps up to time σ_ξ we obtain

$$\begin{aligned} & E_x^f\left[\int_0^{\sigma_\xi} c(X_t,\pi_t)\,dt\right] = E_x^f\left[\sum_{n=0}^{N_\xi}\int_{T_n}^{T_{n+1}} c(X_t,\pi_t)\,dt\right] \\ = \;& E_x^f\left[\sum_{n=0}^{\infty}\int_{T_n}^{T_{n+1}} c(X_t,\pi_t)\,dt\; 1_{[N_\xi \geq n]}\right] \\ = \;& \sum_{n=0}^{\infty} E_x^f\left[E_x^f\left[\int_{T_n}^{T_{n+1}} c(X_t,\pi_t)\,dt \mid \mathcal{F}_{T_n}\right] 1_{[N_\xi\geq n]}\right] \\ = \;& E_x^f\left[\sum_{n=0}^{\infty} C(X_n, f(X_n)) 1_{[N_\xi\geq n]}\right] = E_x^f\left[\sum_{n=0}^{N_\xi} C(X_n, f(X_n))\right] < \infty \end{aligned}$$

Hence it follows from Proposiotion 9.1.7 and Theorem 10.2.2 in Meyn/Tweedie (1993a) that the controlled Markov chain (X_n) is positive Harris recurrent. In particular, we know that the expected cost until the state $\bar\xi = (\xi,\xi,0)$ is reached by the enlarged process are finite. From Theorem 4.1 in Meyn/Tweedie (1993c) we know that $(\bar X_t)$ is positive Harris recurrent with a stationary distribution μ and that $\int c\,d\mu < \infty$. Applying Proposition 4.2 in Glynn/Sigman (1992) we obtain for all $x \in E$.

$$G_f(x) = \int_{\bar E} c(\bar x, f(\eta)(\tau))\mu(d\bar x) < \infty.$$

Theorem 3.2 then gives us $G_f = J_f$. □

The assumption that $L \leq h^\beta(x)$ for $0 < \beta < \bar\beta, x \in E$ is clearly fulfilled, if we have monotonicity, i.e. $V^\beta(x) \geq V^\beta(\xi)$ for all $x \in E$ and $0 < \beta < \bar\beta$. Another important case where this condition is fulfilled emerges when the cost rate function is coercive (see e.g. Kitaev/Rykov (1995)).

Definition 3.4:
The cost rate function $c : E \times U \to I\!R_+$ will be called *coercive* when the set $B_K := \{x \in E \mid \inf_{u\in U} c(x,u) \leq K\}$ is compact for all $K \in I\!R_+$.

Remark 3.3:
Since c is lower semicontinuous and U compact (Assumption 2.1), we obtain with Theorem A.1 that $x \mapsto \min_{u\in U} c(x,u)$ is lower semicontinuous and hence B_K is closed. Therefore, under Assumption 2.1, a growth condition on c like the one in Assumption 3.2 is sufficient for the coercivity of c.

Assumption 3.2:
There exist constants $k \in \mathbb{N}$ and $C_1, C_2 \in \mathbb{R}_+$ such that for all $z \in Z, u \in U$ and $y \in S$

$$c(y,z,u) \geq C_1 \|y\|^k - C_2.$$

Lemma 3.9:
Suppose that Assumption 2.1, 2.2 and 3.1 (i) hold and let $\bar{\beta} > 0$. Assume that there exists an upper semicontinuous function $M : E \to \mathbb{R}_+$ such that $-M(x) \leq h^\beta(x) \leq M(x)$ for all $x \in E$ and $0 < \beta < \bar{\beta}$. If the cost rate function satisfies Assumption 3.2 , then there exists a constant $L \in \mathbb{R}$ such that

$$L \leq h^\beta(x)$$

for all $x \in E,\ 0 < \beta < \bar{\beta}$.

The proof uses ideas of Sennott (1989b) Proposition 3.

Proof: Define $\rho = \limsup_{\beta \downarrow 0} \rho(\beta)$. Note that ρ is finite due to the proof of Lemma 3.4. Choose $K > \max\{\rho + \epsilon, \min_u c(\xi, u)\}$ for $\epsilon > 0$. Hence $\xi \in B_K$. Since B_K is compact, M upper semicontinuous and V^β lower semicontinuous due to our assumptions we can define

$$-L = \max_{x \in B_K} M(x), \quad V^\beta(x^\beta) = \min_{x \in B_K} V^\beta(x).$$

From our assumptions we have

$$-M(x) \leq V^\beta(x) - V^\beta(\xi) \leq M(x)$$

for all $0 < \beta < \bar{\beta}$ and $x \in E$. Hence for all $x \in B_K$

$$\beta V^\beta(x) \geq \beta \left(-M(x) + V^\beta(\xi)\right) \geq \beta \left(L + V^\beta(\xi)\right) \tag{3.6}$$

$$\beta V^\beta(x) \leq \beta \left(M(x) + V^\beta(\xi)\right) \leq \beta \left(-L + V^\beta(\xi)\right)$$

and $\limsup_{\beta \downarrow 0} \beta(-L + V^\beta(\xi)) = \rho$. Therefore, we can conclude that there exists a $\bar{\beta} > 0$ such that $\beta V^\beta(x) \leq \rho + \epsilon$ for all $x \in B_K,\ 0 < \beta < \bar{\beta}$. In particular $\beta V^\beta(x^\beta) \leq \rho + \epsilon$ if $0 < \beta < \bar{\beta}$. Now suppose $x \notin B_K$ and $0 < \beta < \bar{\beta}$ and define $\tau := \inf\{t \geq 0 \mid X_t \in B_K\}$ where (X_t) is the state process induced by the β-discounted optimal policy π_t^β. Thus

$$\begin{aligned} V^\beta(x) &\geq E_x^{\pi^\beta} \left[\int_0^\tau e^{-\beta t} c(X_t, \pi_t^\beta)\, dt + e^{-\beta\tau} V^\beta(x^\beta) \right] \\ &\geq E_x^{\pi^\beta} \left[(\rho + \epsilon) \frac{1 - e^{-\beta\tau}}{\beta} + e^{-\beta\tau} V^\beta(x^\beta) \right] \geq V^\beta(x^\beta). \end{aligned}$$

Notice, that the statement is true even if $\tau = \infty$. Altogether we have for $x \in B_K$ from (3.6)

$$V^\beta(x) - V^\beta(\xi) \geq L$$

and for $x \notin B_K$

$$V^\beta(x) - V^\beta(\xi) \geq V^\beta(x^\beta) - V^\beta(\xi) \geq L$$

which implies the statement. □

Lemma 3.10:
Suppose that the Assumptions 2.1, 2.2, 3.1 and 3.2 are valid. Then $f^0 \in F$ and $f^* \in F$ defined in Theorem 3.6 satisfy

$$G_{f^0} = J_{f^0} \quad \text{and} \quad G_{f^*} = J_{f^*},$$

provided the transition kernels $p(x, f^0; \cdot)$ and $p(x, f^*; \cdot)$ are weakly continuous.

Proof: Since $\tilde{C} : \tilde{D} \to \bar{I\!\!R}_+$ is lower semicontinuous, we obtain as in Remark 3.3 with Assumption 3.2 that $\tilde{C}(x, r)$ is coercive. Denote by $(\bar{X}_t)$ and (X_n) the continuous and discrete Markov chains respectively, which are induced by the stationary policy $(f^0, f^0, \ldots)$ ($(\bar{X}_t)$ is the enlarged process (X_t, η_t, τ_t) cf. Remark 3.2). Let $K > \rho := \limsup_{\beta \downarrow 0} \rho(\beta)$ and define $B_K = \{x \in E \mid \min_{r \in \tilde{D}(x)} \tilde{C}(x, r) \leq K\}$ and $\tau_{B_K} = \inf\{n \in I\!\!N \mid X_n \in B_K\}$. From the proof of Theorem 3.6 we know that $J_{f^0}(x) \leq \rho < \infty$ for all $x \in E$. This immediately implies that $P_x(\tau_{B_K} < \infty) > 0$ for all $x \in E$. Thus, if we denote by ψ the maximal irreducibility measure of (X_n), it follows that $\psi(B_K) > 0$. In particular, *supp* ψ has non-empty interior and since (X_n) is a weak Feller chain by our assumption we obtain with Theorem 6.2.9 in Meyn/Tweedie (1993a) that (X_n) is a T-chain. Moreover, we also know that $P_x(X_n \to \infty) = 0$ for all $x \in E$ and Theorem 9.0.2 in in Meyn/Tweedie (1993a) implies that (X_n) is a Harris recurrent chain which possesses a unique invariant measure ν. Using the coercivity of $C(x, a)$ and the fact that $J_{f^0}(x) < \infty$, Proposition 12.1.3 in Meyn/Tweedie (1993a) tells us that ν is indeed a stationary distribution and thus (X_n) is positive Harris recurrent. Therefore, we obtain with Theorem 14.3.3 in Meyn/Tweedie (1993a) that

$$\lim_{m\to\infty} \frac{1}{m} \sum_{n=0}^{m-1} E_x^{f^0}\left[C(X_n, f^0(X_n))\right] = \frac{1}{q} J_{f^0}(x) = \int C \, d\,\nu$$

independent of the initial state $x \in E$. Obviously $P_x^{f^0}(X_t \to \infty) = 0$ is satisfied for each $x \in E$, hence with Theorem 3.2 (i) in Meyn/Tweedie (1993b) it follows that $(\bar{X}_t)$ is positive Harris recurrent. In analogy to Theorem 4.2 in Meyn/Tweedie (1993b) it can be shown that $(\bar{X}_t)$ is a T-process. Finally, Proposition 4.2 in Glynn/Sigman (1992) implies now

$$\lim_{t\to\infty} \frac{1}{t} E_x^{f^0}\left[\int_0^t c(X_s, \pi_s)\, ds\right] = \int c \, d\,\mu$$

From Theorem 3.2 we know that

$$\int c \, d\,\mu = q \int C \, d\,\nu.$$

Thus $J_{f^0} = G_{f^0}$ independent of $x \in E$. □

3.4 Average Cost Optimality Equation

This section provides a further condition under which the limit h is continuous. This enables us to obtain a HJB-equation for the average cost case. Fortunately, this assumption is not too hard to satisfy for the examples we have in mind (cf. Hernández-Lerma/Lasserre (1996)). Suppose that the sequence $\beta_n \downarrow 0$ is such that $\lim_{n\to\infty} \rho(\beta_n) = \rho := \limsup_{\beta\downarrow 0} \rho(\beta)$.

Assumption 3.3:

(i) The sequence $\{h^{\beta_n}\}$, $\beta_n \to 0$ of the relative value functions is equicontinuous.

(ii) The function M which appears in Assumption 3.1 fulfills for all $x \in E, a \in D(x)$

$$\int_E M(x')p(x, a; d\,x') < \infty.$$

Theorem 3.11: (*Average cost optimality equation*)
Suppose that the Assumptions of Theorem 3.6 and Assumption 3.3 hold. Then

a) There exists a constant $\rho > 0$ and a continuous function $h : E \to I\!R$ such that the *average cost optimality equation* holds, i.e. for all $x \in E$

$$\frac{\rho}{q} + h(x) = \min_{a\in D(x)} \left[C(x, a) + q \int_0^\infty e^{-qt} \sum_{z'} p_{zz'} h\left(\phi_t(x, a), z'\right) \, dt \right]. \tag{3.7}$$

Moreover, there exists a minimizer f^* of h in (3.7).

b) Suppose that $J_{f^*} \geq G_{f^*}$. Then the stationary policy $(f^*, f^*, \ldots)$ is c-average optimal and $\rho = \lim_{\beta\downarrow 0} \rho(\beta)$ are the minimal average cost, independent of x. Moreover, there exists a decision rule f^0 and sequences $\beta_m(x) \to 0, x_m(x) \to x$ such that

$$f^0(x) = \lim_{m\to\infty} f^{\beta_m(x)}(x_m(x)),$$

where f^β is the optimal decision rule in the β-discounted model and the stationary policy $(f^0, f^0, \ldots)$ is c-average optimal, provided $J_{f^0} \geq G_{f^0}$.

Proof: a), b) From Assumption 3.3(i) and the Arzela-Ascoli Theorem we know that there exists a subsequence of $\{h^{\beta_n}\}$ (for convenience still denoted by $\{h^{\beta_n}\}$) and a continuous function h such that for all $x \in E$

$$\lim_{n\to\infty} h^{\beta_n}(x) = h(x)$$

and the convergence is uniform on compact sets. In particular, if $x_n \to x$ for $n \to \infty$ we have $h^{\beta_n}(x_n) \to h(x)$. Thus h coincides with the limit defined in Theorem 3.6 and we have

$$\begin{aligned} \frac{\rho}{q} + h(x) &\geq C(x, g^0(x)) + \int_E h(x')p(x, g^0(x); d\,x') \\ &\geq \min_{a\in D(x)} \left[C(x,a) + \int_E h(x')p(x,a; d\,x') \right]. \end{aligned}$$

As in the proof of Theorem 3.5 we obtain that $f^0(x)(t) = \int_u ug^0(x)(t, du)$ defines a stationary average optimal policy, as well as the minimizer f^* of h. To complete the proof, we have to show the reverse inequality. From the discounted cost optimality equation we have for all $x \in E$, $a \in D(x)$

$$\begin{aligned} \frac{\rho(\beta_n)}{q} + h^{\beta_n}(x) &= \min_{a\in D(x)} \left[C^{\beta_n}(x,a) + \int_E h^{\beta_n}(x')p^{\beta_n}(x,a; d\,x')\, dt \right] \\ &\leq C(x,a) + \int_E h^{\beta_n}(x')p(x,a; d\,x') \end{aligned}$$

Taking $n \to \infty$ we obtain with Assumption 3.3 and Dominated Convergence

$$\frac{\rho}{q} + h(x) \leq C(x,a) + \int_E h(x')p(x,a; d\,x')$$

for all $x \in E$, $a \in D(x)$ which implies the result. □

Remark 3.4:
Assumption 3.3 (i) is for example fulfilled when $S = I\!R^N$ and $(y,u) \mapsto c(y,z,u)$ is convex for all $z \in Z$. This follows from Hernández-Lerma/Lasserre (1996) Remark 5.5.3.

4 Solution Methods

In this section we provide different results which help computing the optimal policy in the β-discounted and in the average cost case. Since we know from Theorem 3.6 that the average cost problem can be solved via the β-discounted problem, we will mainly restrict to solution methods for the β-discounted problem. Reference to the average cost problem will be mentioned explicitly. We will first establish the well-known method of policy iteration. Afterwards solutions to the one-step optimization problems are investigated. Since these problems are deterministic control problems, we will derive a Hamilton-Jacobi-Bellman equation. Further on we explain how to solve the one-step optimization problems via Pontryagin's maximum principle.

4.1 Policy Iteration for β-Discounted Problems

Suppose that the problem of Section 2 is given and Assumptions 2.1 and 2.2 hold. Schäl (1975) showed that the optimal stationary policy can be taken as an accumulation point of the sequence of minimizers f_n of the V_{n-1}, $n \in I\!N$. In particular if all minimizers f_n are equal f then f^∞ is an optimal stationary policy. To formulate this procedure we need the following notions. For $x \in E$, $a \in D(x)$ and $v \in M$ we introduce the operator $L : M \to M$ as

$$Lv(x,a) := C(x,a) + \int_0^\infty e^{-(\beta+q_z)t} \sum_{z' \neq z} q_{zz'} v(\phi_t(x,a), z')\, dt.$$

Let $D_n^*(x) := \{a \in D(x) \mid a \text{ minimizes } a \to LV_{n-1}(x,a)\}$ for $n \in I\!N$, $x \in E$ and $D^*(x) := \{a \in D(x) \mid a \text{ minimizes } a \to LV(x,a)\}$ for all $x \in E$. Moreover, for a compact metric space A and a sequence (A_n) of nonempty subsets of A define

$$LsA_n := \{x \in A \mid x \text{ is an accumulation point of } (x_n) \text{ with } x_n \in A_n\ \forall n \in I\!N\}.$$

Theorem 4.1: (*Policy Iteration*)
Suppose that Assumption 2.1 and 2.2 hold. For all $x \in E$ we obtain

$$\emptyset \neq LsD_n^*(x) \subset D^*(x).$$

Moreover, if $f_n \in D_n^*$, then there exists an accumulation point $f \in F$ such that $f \in LsD_n^*$ and every stationary policy f^∞ with $f \in D^*$ is optimal.

Proof: Fix $x \in E$. Since $a \mapsto LV_{n-1}(x,a)$ is lower semicontinuous for $n \in I\!N$ (see proof of Theorem 2.5), $D_n^*(x) \neq \emptyset$ for all $n \in I\!N$. Thus $LsD_n^*(x) \neq \emptyset$ because $D(x)$

is compact. Let $a_0 \in LsD_n^*(x)$, i.e. there exists a sequence (a_{n_k}) with $a_{n_k} \in D_{n_k}^*(x)$ and $\lim_{k\to\infty} a_{n_k} = a_0$. For $m \in I\!N$ we obtain

$$\begin{aligned} \lim_{n\to\infty} V_n(x) &= \lim_{k\to\infty} V_{n_k}(x) = \lim_{k\to\infty} LV_{n_k-1}(x, a_{n_k}) \geq \liminf_{k\to\infty} LV_m(x, a_{n_k}) \\ &\geq LV_m(x, a_0). \end{aligned}$$

Hence we have

$$V(x) = \lim_{n\to\infty} V_n(x) \geq \lim_{m\to\infty} LV_m(x, a_0) = LV(x, a_0) \geq V(x)$$

and therefore $LsD_n^*(x) \subset D^*(x)$. From Schäl (1975) the existence of f follows and the proof is complete. □

4.2 A Hamilton-Jacobi-Bellman Equation

From Theorem 2.5 we know that the value function V is the solution of a deterministic control problem. Therefore, it is possible to derive a Hamilton-Jacobi-Bellman (HJB) equation for the problem and thus to obtain a Verification Theorem. The approach is standard and can be found e.g. in the text books of Bardi/Capuzzo-Dolcetta (1997), Fleming/Soner (1992), Fleming/Rishel (1975). Throughout the section we will assume that the environment state $z \in Z$ is fixed and therefore sometimes suppress it in the notation. For our convenience we will now introduce the following abbreviations. The function $\mathfrak{c}_V : S \times U \to I\!R_+$ is defined by

$$\mathfrak{c}_V(y, u) := c(y, z, u) + \sum_{z' \neq z} q_{zz'} V(y, z').$$

For $x \in E$ and $a \in D(x)$ the operator L in this section is defined by

$$LV(x, a) := \int_0^\infty e^{-(\beta + q_z)t} \mathfrak{c}_V(\phi_t(x, a), a_t)\, dt.$$

The one-step optimization problem of Theorem 2.5 can now be written in a control theoretic framework as follows

$$(CP) \begin{cases} \int_0^\infty e^{-(\beta+q_z)t} \mathfrak{c}_V(\phi_t(x, a), a_t)\, dt \;\to\; \min \\ \phi_t(x, a) = y + \int_0^t b^z(a_s)\, ds \\ \phi_t(x, a) \in S \\ a_t \in U, \text{ for all } t \geq 0 \end{cases}$$

In order to obtain a Verification Theorem we first have to show

Lemma 4.2: *(Bellman principle)*
Under Assumption 2.1 and 2.2, the following relation holds for all $x \in E,\ T > 0$

$$V(x) = \inf_{a \in D(x)} \left[\int_0^T e^{-(\beta+q_z)t} \mathfrak{c}_V(\phi_t(x,a), a_t)\, dt + e^{-(\beta+q_z)T} V(\phi_T(x,a), z) \right] \tag{4.8}$$

Proof: Let us name $W(x)$ the right hand side of (4.8). Let $T > 0, x \in E$ and $a \in D(x)$ be an arbitrary control. Then

$$\begin{aligned} LV(x,a) &= \int_0^T e^{-(\beta+q_z)t} \mathfrak{c}_V(\phi_t(x,a), a_t)\, dt + \int_T^\infty e^{-(\beta+q_z)t} \mathfrak{c}_V(\phi_t(x,a), a_t)\, dt \\ &= I + e^{-(\beta+q_z)T} LV(\phi_T(x,a), z, \tilde{a}) \geq I + e^{-(\beta+q_z)T} V(\phi_T(x,a), z) \end{aligned}$$

where I is the first integral on the right hand side and $\tilde{a}_t = a_{T+t}$. Taking the infimum over $a \in D(x)$ we obtain $V(x) \geq W(x)$. For the reverse inequality let $a \in D(x)$ be arbitrary. Fix $\epsilon > 0$ and take $\tilde{a} \in D(\phi_T(x,a), z)$ such that $LV(\phi_T(x,a), z, \tilde{a}) - \epsilon < V(\phi_T(x,a), z)$. Define

$$a_t^* = \begin{cases} a_t & , t \leq T \\ \tilde{a}_{t-T} & , t > T \end{cases}$$

Then we obtain

$$\begin{aligned} V(x) &\leq LV(x, a^*) \\ &= \int_0^T e^{-(\beta+q_z)t} \mathfrak{c}_V(\phi_t(x,a), a_t)\, dt + \int_T^\infty e^{-(\beta+q_z)t} \mathfrak{c}_V(\phi_t(x,a^*), a_t^*)\, dt \\ &= I + e^{-(\beta+q_z)T} LV(\phi_T(x,a), z, \tilde{a}) \leq I + e^{-(\beta+q_z)T} V(\phi_T(x,a), z) + \tilde{\epsilon} \end{aligned}$$

where I is the first integral on the right hand side. Since ϵ and a are arbitrary we get $V(x) \leq W(x)$ for all $x \in E$. □

Part b) of the following theorem is a so-called *Verification Theorem.*

Theorem 4.3:
We suppose that Assumptions 2.1, 2.2 hold and that $(y,u) \mapsto c(y,z,u)$ and $y \mapsto V(y,z)$ are continuous and

$$V(x) \leq C_0(1 + \|y\|^k)$$

for some constants $C_0 \in \mathbb{R}_+,\ k \in \mathbb{N}$. Then we obtain

a) The value function V is a *constrained viscosity solution* of the HJB equation

$$(\beta + q_z)V(x) = \min_{u\in U}\left[\mathfrak{c}_V(y,u) + b^z(u)V_y(x)\right]. \tag{4.9}$$

b) If the continuously differentiable function $W(\cdot, z) : S \to I\!R$ satisfies (4.9) and $\lim_{t\to\infty} e^{-(\beta+q_z)t}W(\phi_t(x,a)) = 0$, for all $a \in D(x)$ and $E_x^\pi[W(X_n)] \to 0$ when $n \to \infty$ for all policies π, $x \in E$, then $W(x) \le V(x)$ for all $x \in E$. Moreover, if there exists a decision rule $f^* \in F$ such that for all $x \in E$

$$(\beta+q_z)W(\phi_t(x,f^*),z) = \mathfrak{c}_W(\phi_t(x,f^*), f^*(x)(t)) + b^z(f^*(x)(t))W_y(\phi_t(x,f^*),z)$$

for almost every $t \ge 0$, then $V = W$ and $(f^*, f^*, \dots)$ is a stationary, β-discounted optimal policy.

Proof:

a) Fix $x_0 = (y_0, z) \in E, y_0 \in \overset{\circ}{S}$ and let ψ^1 be continuously differentiable on $I\!R^N$, such that $V(y,z) - \psi^1(y)$ attains its maximum at $y = y_0$ in a neighborhood $N(y_0)$ of y_0 in S. Consider a control $a \in D(x_0)$ with $a_t \equiv u$ for $0 < t < T$, where T is small enough such that a is admissible and $\phi_t(x_0,a) \in N(y_0)$ for $0 < t < T$. From Lemma 4.2 we have

$$V(x_0) - e^{-(\beta+q_z)T}V(\phi_T(x_0,a),z) \le \int_0^T e^{-(\beta+q_z)t}\mathfrak{c}_V(\phi_t(x_0,a),a_t)\,dt.$$

Since $\phi_T(x_0,a) \in N(y_0)$ one gets

$$V(\phi_T(x_0,a),z) \le \psi^1(\phi_T(x_0,a)) + V(x_0) - \psi^1(y_0).$$

Hence

$$V(x_0) - e^{-(\beta+q_z)T}\Big(\psi^1(\phi_T(x_0,a)) + V(x_0) - \psi^1(y_0)\Big) \le \int_0^T e^{-(\beta+q_z)t}\mathfrak{c}_V(\phi_t(x_0,a),a_t)\,dt.$$

Dividing the preceding inequality by T and letting T tend to zero we obtain

$$(\beta + q_z)V(x_0) - \psi^1_y(y_0)b^z(u) \le \mathfrak{c}_V(y_0,u)$$

Since the inequality holds for all $u \in U$ we obtain

$$(\beta + q_z)V(x_0) \le \min_{u\in U}\Big[\mathfrak{c}_V(y_0,u) + \psi^1_y(y_0)b^z(u)\Big].$$

Thus, $V(\cdot, z)$ is a viscosity subsolution on $\overset{\circ}{S}$.
To show that $V(\cdot, z)$ is viscosity supersolution on S, we suppose the contrary. Hence, there exists an $x_0 = (y_0, z) \in E$, a continuously differentiable function

ψ^2 on $I\!R^N$ such that $V(y,z) - \psi^2(y)$ attains its minimum at y_0 in a neighborhood $N(y_0)$ of y_0 in S and a constant $\delta >$ such that for all $u \in U$

$$(\beta + q_z)V(x_0) - \psi^2_y(y_0)b^z(u) + \delta \leq \mathfrak{c}_V(y_0, u).$$

Since all functions are continuous this implies

$$(\beta + q_z)V(y, z_0) - \psi^2_y(y)b^z(u) + \delta \leq \mathfrak{c}_V(y, u)$$

for all $y \in N(y_0)$ and $u \in U$. Let $a \in D(x_0)$ and $T > 0$ small enough, then

$$\begin{aligned} LV(x_0, a) &\geq \int_0^T e^{-(\beta+q_z)t} \mathfrak{c}_V(\phi_t(x_0, a), a_t)\, dt + e^{-(\beta+q_z)T} V(\phi_T(x_0, a), z) \\ &\geq \int_0^T e^{-(\beta+q_z)t} \Big(\delta + (\beta + q_z)V(\phi_t(x_0, a), z) - \\ &\quad \psi^2_y(\phi_t(x_0, a))b^z(a_t)\Big)\, dt + e^{-(\beta+q_z)T} V(\phi_T(x_0, a), z). \end{aligned}$$

On the other hand we have

$$\begin{aligned} &V(x_0) - e^{-(\beta+q_z)T} V(\phi_T(x_0, a), z) \\ &\leq V(x_0) - e^{-(\beta+q_z)T} \Big(\psi^2(\phi_T(x_0, a)) + V(x_0) - \psi^2(y_0)\Big) \\ &\leq \int_0^T e^{-(\beta+q_z)t} \Big((\beta + q_z)V(\phi_t(x_0, a), z) - \psi^2_y(\phi_t(x_0, a))b^z(a_t)\Big)\, dt. \end{aligned}$$

Taking the infimum over all $a \in D(x_0)$ we obtain from the last two inequalities

$$V(x_0) \geq V(x_0) + \delta \frac{1 - e^{-(\beta+q_z)T}}{\beta + q_z}$$

which is a contradiction. Therefore $V(\cdot, z)$ is a viscosity supersolution on S and hence a constrained viscosity solution.

b) Let W be as assumed and $a \in D(x)$ an arbitrary control. Hence

$$\begin{aligned} e^{-(\beta+q_z)T} W(\phi_T(x, a), z) &= W(x) - \int_0^T e^{-(\beta+q_z)t} \Big((\beta + q_z)W(\phi_t(x, a), z) - \\ &\quad W_y(\phi_t(x, a), z)b^z(a_t)\Big)\, dt \\ &\geq W(x) - \int_0^T e^{-(\beta+q_z)t} \mathfrak{c}_W(\phi_t(x, a), a_t)\, dt. \end{aligned}$$

Letting $T \to \infty$ we obtain with our assumption $W \leq \mathcal{U}W$. Thus, for an arbitrary policy π we get

$$E^\pi_x [W(X_{n+1}) \mid X_n] = \int_E W(x)p(X_n, f(X_n); dx) =$$

$$C(X_n, f_n(X_n)) + \int_0^\infty e^{-(\beta+q_{Z_n})t} \sum_{z' \neq Z_n} q_{Z_n z'} W(\phi_t(X_n, f_n(X_n)))dt - C(X_n, f_n(X_n))$$
$$\geq W(X_n) - C(X_n, f_n(X_n))$$

Hence

$$C(X_n, f_n(X_n)) \geq E_x^\pi \left[W(X_n) - W(X_{n+1}) \mid X_n\right].$$

Summing over n from 0 to m and taking expectation w.r.t. P_x^π, we obtain

$$E_x^\pi \left[\sum_{n=0}^m C(X_n, f_n(X_n))\right] \geq W(x) - E_x^\pi \left[W(X_{m+1})\right].$$

Taking limit $m \to \infty$ and applying our assumption we obtain $V \geq W$.
If there exists an $f^* \in F$ which provides the infimum in (4.9) then we derive in addition $W = \mathcal{U}W$, thus $W \geq V$ (cf. Theorem 2.5) and the proof is complete. □

Remark 4.1:
If the value function is continuously differentiable w.r.t. y, then equation (4.9) reduces to the HJB equation in the classical sense.

As far as the average cost problem is concerned we obtain a similar Verification Theorem. In this case we define $\mathfrak{c}_h(y, u) := c(x, u) + q \sum_{z'} p_{zz'} h(y, z')$.

Theorem 4.4:
Suppose that Assumption 2.1, 2.2, 3.1 and 3.3 are valid and that M in 3.1 satisfies

$$M(x) \leq C_0(1 + \|y\|^k)$$

for some constants $C_0 \in \mathbb{R}_+$, $k \in \mathbb{N}$. Define $\rho := \limsup_{\beta \downarrow 0} \rho(\beta)$. If $(y, u) \mapsto c(y, z, u)$ is continuous, then

a) The limit h of the relative value functions is a constrained viscosity solution of the HJB equation

$$\rho + qh(x) = \min_{u \in U} \left[\mathfrak{c}_h(y, u) + b^z(u) h_y(x)\right]. \tag{4.10}$$

b) If the continuously differentiable function $w(\cdot, z) : S \to \mathbb{R}$ satisfies (4.10) and if there exists a decision rule $f^* \in F$ such that for all $x \in E$, $J_{f^*} \geq G_{f^*}$ and

$$\rho + qw(\phi_t(x, f^*), z) = \mathfrak{c}_w(\phi_t(x, f^*), f^*(x)(t)) + b^z(f^*(x)(t)) w_y(\phi_t(x, f^*), z) \tag{4.11}$$

for almost every $t \geq 0$ and

$$\lim_{m\to\infty} \frac{1}{m} E_x^{f^*}[w(X_{m+1})] = 0, \tag{4.12}$$

then $\rho = G_{f^*}(x)$ for $x \in E$ and $(f^*, f^*, \ldots)$ is a c-average optimal policy.

Proof: Under our assumption, the average cost optimality equation (Theorem 3.11) holds and can be written in the following form

$$h(x) = \min_{a\in D(x)} \left[\int_0^\infty e^{-qt}(\mathfrak{c}_h(\phi_t(x,a), a_t) - \rho)\, dt \right].$$

As in Theorem 4.3 we obtain part a). For part b) we proceed as in Theorem 4.3 b). First we obtain for all $x \in E$

$$w(x) = \int_0^\infty e^{-qt}(\mathfrak{c}_w(\phi_t(x,f^*), f^*(x,t)) - \rho)\, dt.$$

Then we get

$$E_x^{f^*}[w(X_{n+1}) \mid X_n] = w(X_n) + \frac{\rho}{q} - C(X_n, f^*(X_n))$$

and therefore

$$\frac{1}{m} E_x^{f^*} \left[\sum_{n=0}^{m-1} C(X_n, f^*(X_n)) \right] = \frac{1}{m} w(x) - \frac{1}{m} E_x^{f^*}[w(X_{m+1})] + \frac{\rho}{q}.$$

Letting $\limsup_{m\to\infty}$ and using our assumption we obtain $G_{f^*} \leq J_{f^*} = \rho$ and the proof is complete. □

4.3 Necessary and Sufficient Conditions for Optimality

For the control problem (CP) of the previous section we also have a maximum principle which provides necessary conditions for the optimal control. Under further assumptions, we also obtain sufficient conditions. Here we will restrict to the cases $S = \mathbb{R}^N$ and $S = \mathbb{R}_+^N$. The following theorems can be found in Seierstad/Sydsæter (1987). Throughout this section we assume

Assumption 4.1:
The mapping $y \mapsto \mathfrak{c}_V(y,z,u)$ is differentiable and the derivative is continuous w.r.t. all variables.

The Hamiltonian of problem (CP) is defined by

$$H(y,u,p) = p_0 \mathfrak{c}_V(y,z,u) + p^T b(u).$$

Theorem 4.5: *(Maximum principle)*
Suppose that $S = I\!R^N$ and Assumptions 2.1, 2.2 and 4.1 are valid. Let a_t^* be a piecewise continuous control and y_t^* the associated trajectory. If a_t^* is optimal, then there exists a constant p_0 and a continuous and piecewise continuously differentiable vector function $p_t = (p_1(t), \dots, p_n(t))$ such that for all $t \geq 0$

(i) $(p_0, p_t) \neq 0$, $p_0 = 1$ or $p_0 = 0$.
(ii) a_t^* minimizes $a_t \mapsto H(y_t^*, a_t, p_t), \ a_t \in U$.
(iii) $\dot{p}_t - (\beta + q_z) p_t = -p_0 \frac{\partial}{\partial y} \mathfrak{c}_V(y_t^*, a_t^*)$ except at points of discontinuity of a_t^*.

Theorem 4.6: *(Sufficient conditions for optimality)*
In addition to the assumptions of Theorem 4.5 we suppose that $u \mapsto \mathfrak{c}_V(y,u)$ is continuously differentiable and $y \mapsto c(y,u)$ is convex. The admissible control a_t^* with the associated trajectory y_t^* is optimal for (CP) if there exists a continuous and piecewise continuously differentiable vector function $p_t = (p_1(t), \dots, p_N(t))$ such that for all $t \geq 0$

(i) a_t^* minimizes $a_t \mapsto H(y_t^*, a_t, p_t), \ a_t \in U$.
(ii) $\dot{p}_t - (\beta + q_z) p_t = -\frac{\partial}{\partial y} \mathfrak{c}_V(y_t^*, a_t^*)$, except at points of discontinuity of a_t^*.
(iii) $\liminf_{t \to \infty} e^{-(\beta + q_z)t} p_t (y_t^* - y_t) \geq 0$ for all admissible trajectories y_t.

Since S and the mapping $y \mapsto c(y,u)$ are convex, we obtain from Lemma 2.6 that $(y,u) \mapsto H(y,u,p)$ is convex. Hence Theorem 4.6 follows from Theorem 3.13 in Seierstad/Sydsæter (1987). In various applications we have $S = I\!R_+^N$. Here the following theorem is of use

Lemma 4.7: *(Sufficient conditions for optimality)*
Suppose that $S = I\!R_+^N$ and Assumptions 2.1, 2.2 and 4.1 are valid. Further on assume that $u \mapsto \mathfrak{c}_V(y,u)$ is continuously differentiable and $c(y,u) = c_1(y) + c_2(u)$

where $y \mapsto c_1(y)$ and $u \mapsto c_2(u)$ are convex. The control a_t^* with the associated trajectory y_t^* is optimal for (CP) if there exists a continuous and piecewise continuously differentiable vector function $p_t = (p_1(t), \ldots, p_N(t))$ as well as a piecewise continuous vector function $\eta_t = (\eta_1(t), \ldots, \eta_N(t))$ such that for all $t \geq 0$

(i) a_t^* minimizes $a_t \mapsto H(y_t^*, a_t, p_t),\ a_t \in U$.

(ii) $\dot{p}_t - (\beta + q_z)p_t = -\frac{\partial}{\partial y}\mathfrak{c}_V(y_t^*, a_t^*) + \eta_t$, when $\dot{p}_t$ exists at t.

(iii) $\eta_t \geq 0$.

(iv) $\eta_t y_t^* = 0$.

(v) $\liminf_{t\to\infty} e^{-(\beta+q_z)t} p_t(y_t^* - y_t) \geq 0$ for all admissible trajectories y_t.

5 Numerical Methods

In this section we investigate numerical methods for solving our SFP. In a first part we consider the special case of a deterministic fluid program. The reason is twofold. On the one hand this is a very important problem (cf. also Section 7), on the other hand there exist very efficient algorithms to solve it. In a second part we deal with the general SFP. In both cases we apply the numerical methods to examples. The second application is an admission control problem which has been investigated in Bäuerle (1998b). For this model we also present a sensitivity analysis w.r.t. some stochastic parameters.

5.1 Numerical Methods for Deterministic Fluid Programs

We present an algorithm which solves the purely deterministic fluid optimization problem, i.e. if the dynamics $b^z(u) = b(u)$ are not influenced by the environment process. This will be important in Section 7. Several papers have dealt with this problem. In Chen/Yao (1993) a myopic solution procedure is explained. Weiss (1996, 1997) develops an algorithm for re-entrant lines by formulating the problem as a separated continuous linear program and using results of Pullan (1993, 1995). In Avram/Bertsimas/Ricard (1995) a heuristical approach is given, using Pontryagins maximum principle. For our algorithm we use essentially the formulation as a separated continuous linear program combined with the sufficient conditions for deterministic control problems presented in the last section.
In this subsection we assume now that $S = \mathbb{R}_+^N$, that the cost rate function c is linear in y and u, i.e. $c(y,u) = c_1 y + c_2 u$ and that the set of controls U is a bounded polyhedron, i.e. U can be written as $U = \{u \in \mathbb{R}^K \mid Au \le b,\ u \ge 0\}$ and U is bounded. We get the following optimization problem:

$$\begin{cases} \int_0^\infty e^{-\beta t}\,[c_1 y_t + c_2 a_t]\; dt \;\rightarrow\; \min \\ y_t = y_0 + \int_0^t b(a_s)\, ds \\ A a_t \le b \\ y_t, a_t \ge 0 \text{ for all } t \ge 0 \end{cases}$$

It is now possible to write the objective function in a slightly different form:

$$\begin{aligned} \int_0^\infty e^{-\beta t} c_1 y_t\, dt &= \frac{c_1 y_0}{\beta} + \int_0^\infty \int_0^t e^{-\beta t} c_1 b(a_s)\, ds\, dt = \\ &= \frac{c_1 y_0}{\beta} + \int_0^\infty \int_s^\infty e^{-\beta t} c_1 b(a_s)\, dt\, ds = \frac{c_1 y_0}{\beta} + \frac{1}{\beta}\int_0^\infty e^{-\beta s} c_1 b(a_s)\, ds \end{aligned}$$

To get a numerically tractable problem we replace the infinite time horizon by a finite but large horizon $T > 0$ and obtain the following problem with a new objective function:

$$
(SCLP) \begin{cases} \int\limits_0^T e^{-\beta t} c a_t \, dt \;\to\; \min \\ y_t = y_0 + \int\limits_0^t b(a_s)\, ds \\ A a_t \le b \\ y_t, a_t \ge 0 \text{ for all } t \ge 0 \end{cases}
$$

where $c \in I\!R^K$. This is now a so-called separated continuous linear program $(SCLP)$. The optimal solution consists of two functions (a_t, y_t). However, since y_t can be determined from a_t, when speaking of an optimal solution for $(SCLP)$ we only mean the component a_t. This point of view will also be taken for the other programs which appear during this section. Problems of this type have already been extensively investigated in the literature, see e.g. Pullan (1993, 1995) and the references given there. Weiss (1996, 1997) used this formulation to give an algorithm for solving re-entrant problems. The following result which states that the optimal control can be chosen w.l.o.g. as a piecewise constant function is due to Pullan (1995) Theorem 3.3

Theorem 5.1:
There exists an optimal solution of $(SCLP)$ with a_t piecewise constant on $[0, T]$ and with a finite number of breakpoints.

Remark 5.1:
If we have a finite horizon problem with $\beta = 0$, the objective function of $(SCLP)$ is $\int_0^T (T-t) c a_t \, dt$ and Theorem 5.1 is still valid.

Besides the algorithm of Weiss (1996, 1997) for re-entrant lines, there exists a general algorithm for solving $(SCLP)$, see e.g. Pullan (1993). The key issue is the following discrete formulation. Let $P = \{t_0, \ldots, t_m\}$ with $t_0 = 0 < t_1 < \ldots < t_m = T$ be an arbitrary, fixed partition of the interval $[0, T]$. We define the ordinary linear program $LP(P)$ by

$$
LP(P) \begin{cases} \sum\limits_{i=1}^m \frac{1}{\beta}\left(e^{-\beta t_{i-1}} - e^{-\beta t_i}\right) c a(t_{i-1}+) \;\to\; \min \\ (t_1 - t_0) b(a(t_0+)) - y(t_1) = -y_0 \\ (t_i - t_{i-1}) b(a(t_{i-1}+)) - y(t_i) + y(t_{i-1}) = 0, \; i = 2, \ldots, m \\ A a(t_{i-1}+) \le b, \; i = 1, \ldots, m \\ y(t_i), a(t_{i-1}+) \ge 0, \; i = 1, \ldots, m \end{cases}
$$

To relate optimal solutions of $LP(P)$ to optimal solutions of $(SCLP)$ we need the following definitions:

Definition 5.1:
Let $P = \{t_0, \ldots, t_m\}$ be a partition of $[0, T]$. Suppose we have $m + 1$ vectors $\hat{a}(t_0), \ldots, \hat{a}(t_m)$, then the function $a(t)$ defined by

$$a(t) = \begin{cases} \hat{a}(t_{m-1}+) & , t = T \\ \hat{a}(t_{i-1}+) & , t_{i-1} \leq t < t_i,\ i = 1, \ldots, m \end{cases}$$

is called *piecewise constant extension* of $\hat{a}$. The function $a(t)$ defined by

$$a(t) = \left(\frac{t_i - t}{t_i - t_{i-1}} \right) \hat{a}(t_{i-1}) + \left(\frac{t - t_{i-1}}{t_i - t_{i-1}} \right) \hat{a}(t_i), \text{ for } t_{i-1} \leq t \leq t_i,\ i = 1, \ldots, m$$

is called *piecewise linear extension* of $\hat{a}$.

For arbitrary P, the piecewise constant extension a of the optimal solution $\hat{a}$ of $LP(P)$ is admissible for $(SCLP)$ (see Pullan (1993), Lemma 3.1) and

$$\int_0^T e^{-\beta t} c a_t \, dt = \sum_{i=1}^{m} \frac{1}{\beta} \left(e^{-\beta t_{i-1}} - e^{-\beta t_i} \right) c \hat{a}(t_{i-1}+),$$

i.e. the value of $LP(P)$ gives us an upper bound for the value of $(SCLP)$. If P contains the breakpoints of the optimal solution of $(SCLP)$, the piecewise constant extension of the optimal solution of $LP(P)$ obviously gives the optimal solution a_t of $(SCLP)$. On the other hand, a steady refinement of the set of breakpoints P leads to an approximation of the optimal solution of $(SCLP)$. The idea of the algorithm which we present in this section is now to compute time points which come close to potential breakpoints of the optimal solution. This can be done since breakpoints are either depletion times of buffers or time points of switching priorities (see Avram/Bertsimas/Ricard (1995)). Switching priorities will be detected by using the sufficient conditions for optimality presented in Section 4. We will now give a discrete version thereof for our deterministic model.

Lemma 5.2:
Suppose that a partition P of $[0, T]$ is given. If the variables $a(t_{i-1}+), y(t_i), \eta(t_{i-1}+)$, $p(t_m), p(t_{i-1}),\ i = 1, \ldots, m$ fulfill

(i) $(t_1 - t_0) b(a(t_0+)) - y(t_1) = -y_0$

$(t_i - t_{i-1}) b(a(t_{i-1}+)) - y(t_i) + y(t_{i-1}) = 0,\ i = 2, \ldots, m.$

(ii) $Aa(t_{i-1}+) \leq b,\ i = 1, \ldots, m.$

(iii) $p(t_i) - e^{\beta(t_i - t_{i-1})}p(t_{i-1}) = \frac{1}{\beta}(c - \eta(t_{i-1}+))(e^{\beta(t_i-t_{i-1})} - 1), \ \ i = 1, \ldots, m$ $p(t_m) = 0$.

(iv) $\eta(t_{i-1}+)y(t_{i-1}) = \eta(t_{i-1}+)y(t_i) = 0, \ i = 1, \ldots, m$.

(v) $a(t_{i-1}+)$ minimizes $a \mapsto p(t_{i-1})b(a)$ and $a \mapsto p(t_i)b(a)$ with $a \in U$.

(vi) $y(t_i), a(t_{i-1}+), \eta(t_{i-1}+) \geq 0, \ i = 1, \ldots, m$.

then the piecewise constant extension of $a(t_{i-1}+), \ i = 1, \ldots, m$ is an optimal solution of $(SCLP)$.

Proof: Let a_t, η_t be the piecewise constant extension of $a(t_{i-1}+), \ i = 1, \ldots, m$ and $\eta(t_{i-1}+), \ i = 1, \ldots, m$ respectively and y_t the piecewise linear extension of $y_0, y(t_i), \ i = 1, \ldots, m$. We define for $t_{i-1} \leq t < t_i$ the function

$$\alpha_t = (e^{\beta(t-t_{i-1})} - 1)/(e^{\beta(t_i - t_{i-1})} - 1)$$

(obviously $0 \leq \alpha_t \leq 1$) and $p_t = \alpha_t p(t_i) + (1 - \alpha_t)p(t_{i-1})$. The statement follows, when we can show that the process (a_t, y_t, η_t, p_t) satisfies for all $0 \leq t \leq T$

(i) $y_t = y_0 + \int_0^t b(a_s)\, ds \geq 0$.

(ii) $a_t \in U$.

(iii) $\dot{p}_t - \beta p_t = -c + \eta_t, \ p_T = 0$

(iv) $\eta_t y_t = 0$.

(v) a_t minimizes $a \mapsto p_t b(a), \ a \in U$.

(i) and (ii) follow from Lemma 3.1 in Pullan (1993). (iv) holds since y_t is linear on $[t_{i-1}, t_i], \ i = 1, \ldots, m$ which implies that if $y_j(t_{i-1}) = y_j(t_i) = 0$, then $y_j(t) \equiv 0$ on $[t_{i-1}, t_i]$. Since for $t_{i-1} \leq t < t_i$, $p_t = \alpha_t p(t_i) + (1 - \alpha_t)p(t_{i-1})$ and $a_t \equiv a(t_{i-1}+)$ on $[t_{i-1}, t_i]$ minimizes both $a \mapsto p(t_{i-1})b(a)$ and $a \mapsto p(t_i)b(a)$, we obtain (v). Thus it is left to show (iii). However, a short calculation gives us that for all $0 \leq t \leq T$

$$p_t = \int_t^T e^{\beta(t-s)}(-c + \eta_s)\, ds$$

which implies (iii). □

Our algorithm now proceeds as follows. For $y \in I\!R_+^N$ define $\Upsilon(y) = \{1 \leq j \leq N \mid y_j = 0\}$. Suppose we have a partition set P of $[0, T]$. We can solve the $LP(P)$ with starting state y and partition set P. The corresponding value function will be denoted by $V(P, y)$. If the solution is optimal, we would have $p(t_0) = \frac{\partial}{\partial y}V(P, y(t_0))$. Hence we approximate the real value $p(t_0)$ by $\frac{\partial}{\partial y}V(P, y(t_0))$ and compute a new control $a(t_0+)$ for the first time interval by solving

$$(LP)\begin{cases} p(t_0)b(a) \;\to\; \min \\ Aa \le b \\ b_j(a) \ge 0, \text{ for all } j \in \Upsilon(y) \\ a \ge 0 \end{cases}$$

The first depletion time under this control is then given by

$$t' := \min\{\frac{y_j}{-b_j(a(t_0+))} \mid j \notin \Upsilon(y),\ b_j(a(t_0+)) < 0,\ j = 1, \ldots, N\}.$$

Determine $\eta(t_0+) = c + \dot{p}(t_0) - \beta p(t_0)$ and define $p_t = e^{\beta(t-t_0)}p(t_0) + \frac{1}{\beta}(c - \eta(t_0+))(e^{\beta(t-t_0)} - 1)$. Using sensitivity analysis, we can determine the time interval $[t_0, t'']$ on which $a(t_0)$ is a solution of (LP) with $p(t_0)$ replaced by p_t. Thus t'' gives the time where priorities switch. $t := \min\{t', t''\}$ is therefore a potential breakpoint for the optimal control and will be added to P. Starting from t we can compute in the same way further potential breakpoints. In order to obtain a good initial solution, we determine in a first step the myopic solution as in Chen/Yao (1993). The procedure is as follows

PROCEDURE Myopic Solution

$t := t_0,\ y := y_0,\ i := 0$

WHILE $(t < T)$ **DO**

Solve

$$(LP)\begin{cases} ca \;\to\; \min \\ Aa \le b \\ b_j(a) \ge 0, \text{ for all } j \in \Upsilon(y) \\ a \ge 0 \end{cases}$$

$a_i := a$

$\Delta t := \min\{\frac{y_j}{-b_j(a)} \mid j \notin \Upsilon(y),\ b_j(a) < 0,\ j = 1, \ldots, N\}$

$t_{i+1} := t_i + \Delta t$

$t := t_{i+1}$

$y := y + \Delta t\, b(a)$

$i = i + 1$

END

OUTPUT (a_i, t_i).

In the myopic solution procedure above, we minimize the cost at a breakpoint time, subject to getting a feasible pair of control and trajectory and determine the next breakpoint as the first depletion time under this control. Sometimes the myopic solution is already optimal, as for example in the case of the index policy (see Chen/Yao (1993)). Here we are only interested in the breakpoints generated by the algorithm which we will use to obtain an improved solution. The whole algorithm for solving $(SCLP)$ works as follows:

ALGORITHM SCLP

STEP 1: Compute Myopic Solution

P_0 = breakpoints of Myopic Solution.

STEP 2: Computation of new breakpoints:
$t := t_0,\ y := y_0,\ P := P_0$

WHILE $(t < T)$ **DO**
Solve $LP(P)$. Objective value $V(P, y)$. $p_t := \frac{\partial}{\partial y} V(P, y)$
Solve

$$(LP)\begin{cases} p_t b(a) \ \rightarrow \ \min \\ Aa \leq b \\ b_j(a) \geq 0, \ \text{for all } j \in \Upsilon(y) \\ a \geq 0 \end{cases}$$

$t' := \min\{\frac{y_j}{-b_j(a)} \mid j \notin \Upsilon(y),\ b_j(a) < 0,\ j = 1, \ldots, N\}$

$\eta_j := 0$, if $j \notin \Upsilon(y)$ or $j \in \Upsilon(y)$ and $b_j(a) > 0$ else
$\eta_j := c_j + \dot{p}_j(t) - \beta p_j(t),\ j = 1, \ldots, m$

$p_s := e^{\beta(s-t)} p_t + \frac{1}{\beta}(c - \eta)(e^{\beta(s-t)} - 1)$

$t'' := \max\{s \geq 0 \mid a$ still solves (LP) with p_t replaced by $p_s\}$

$\Delta t := \min\{t', t''\}$

$y := y + \Delta t\, b(a)$

$t := t + \Delta t$

$P_0 := P_0 \cup \{t\}$

$P := \{\hat{t} - \Delta t \geq 0 \mid \hat{t} \in P\} \cup \{0\}$
END

IF the solution (a, y, η, p) satisfies (i)-(vi) of Lemma 5.2 **THEN** Stop.
The piecewise constant extension of a is optimal for $(SCLP)$.

ELSE GOTO STEP 2.

Of course, if the number of breakpoints which have been generated in one step is very small, one can add further breakpoints, like $\frac{t_i+t_{i-1}}{2}$, $i = 1, \ldots, m$. In any case one should always remove breakpoints which have not led to an improvement in the next step.

Example:
We have used the algorithm to solve the following small network

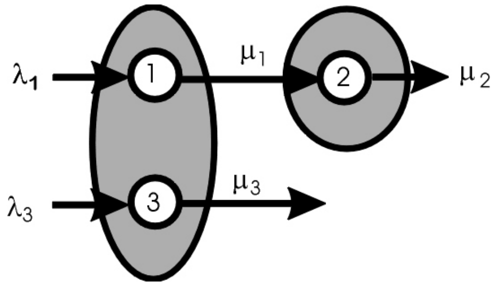

fig.5.1 : Network with blocking

with $\lambda_1 = \lambda_3 = 0, \mu_1 = 8, \mu_2 = 2, \mu_3 = 10, c_1 = \mathbb{1}, c_2 = 0, y_0 = (8, 0, 10), \beta = 0$. It can be shown for the ordinary stochastic network problem that the optimal policy is a switching policy. However, the proof is surprisingly difficult (see Bäuerle/Brüstl/Rieder (1998)). Avram/Bertsimas/Ricard (1995) have shown that the same result holds for the deterministic fluid problem, but the proof is only correct for the interior of the state space. The data of this problem is chosen in such a way that the solution obtained from Avram/Bertsimas/Ricard (1995) is not correct (indeed they get the myopic solution). The result we obtain with our algorithm is summarized in the following table

	Myopic Solution	$LP(P)$ Solut.	New Breakpoints	$LP(P)$ Solut.
P	$\{0, 1, 2, 5\}$	-	$\{0, 1, 1.33, 2, 4, 5\}$	-
$a(t_0+)$	$(0, 0, 1)$	$(\frac{1}{4}, 1, \frac{3}{4})$	-	$(\frac{1}{4}, 1, \frac{3}{4})$
$a(t_1+)$	$(1, 1, 0)$	$(\frac{3}{4}, 1, \frac{1}{4})$	-	$(\frac{1}{4}, 1, \frac{3}{4})$
$a(t_2+)$	$(0, 1, 0)$	$(0, \frac{2}{3}, 0)$	-	$(\frac{1}{4}, 1, 0)$
$a(t_3+)$	-	-	-	$(\frac{1}{4}, 1, 0)$
$a(t_4+)$	-	-	-	$(0, 0, 0)$
$y(t_1)$	$(8, 0, 0)$	$(6, 0, 2.5)$	-	$(6, 0, 2.5)$
$y(t_2)$	$(0, 6, 0)$	$(0, 4, 0)$	-	$(5.33, 0, 0)$
$y(t_3)$	$(0, 0, 0)$	$(0, 0, 0)$	-	$(4, 0, 0)$
$y(t_4)$	-	-	-	$(0, 0, 0)$
Obj. value	29	25.5	-	22.67

First we have computed the myopic solution with breakpoints $\{0, 1, 2, 5\}$ and objective value 29. Using these breakpoints to solve $LP(P)$ we get an improved solution with objective value 25.5. In a next step we have computed the new breakpoints 1.33, 2 and 4 which we have added to our set P. Solving $LP(P)$ gives the objevtive value 22.67 and the algorithm detects optimality.

5.2 Numerical Methods for Stochastic Fluid Programs

Solving SFPs is much more complicated numerically than solving deterministic ones. It is even not clear whether the optimal solution is again piecewise constant for every environment state. Of course one would expect this. In principle, there are several different ways in which one could tackle this problem. For example one could try and solve the fixed point equation 2.1 by iteration. However, we use a direct approach to this problem, namely the *Approximating Markov chain approach* (see Kushner/Dupuis (1992)). This is a general method for stochastic control problems. In what follows, we give a short outline of how to apply it to our SFP as defined in Section 2.1. 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

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

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:

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

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

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

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

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

Under Assumptions 2.1 and 2.2 there exists a minimizer f of V^h and the stationary policy $\pi = f^\infty$ is optimal. From Kushner/Dupuis (1992) we know that for $h \to 0$, $V^h(x)$ converges to $V(x)$ for every environment state z. We use this method to compute the optimal policy for the following admission control problem.

Example:
A controller has to decide upon acceptance of fluid for N buffers in parallel. If the environment process is in state z at time t there is a demand for inflow into buffer j at rate $\lambda_j(z)$. The outflow rate μ_j for buffer j is fixed $j = 1, \ldots, N$. The controller obtains a reward r_j for each unit of accepted fluid for buffer j, but has to pay holding costs $\hat{c}(y)$ which depend on the common buffer content y and are increasing. The control is hence a vector (u, v), where u_j is the fraction of fluid that is admitted to buffer j and v_j is the activation rate of server j (of course $v_j = 1$ if there is fluid in the buffer). In terms of our SFP, the data is given by

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

It has been shown in Bäuerle (1998b) that the optimal policy is of switching type under suitable assumptions on the cost rate function. In particular in the one buffer case ($N = 1$) we obtain a threshold policy (see also Rajagopal et al. (1995)).
The following numerical computation of the optimal policy has been done for the one-and two-buffer case with two environment states. Figure 5.2 and 5.3 refer to the **one-buffer case** with $c(y) = (y + 0.5)^2$, $\beta = 0.9$, $r = \frac{20}{9}$, $\mu = 2$. In figure 5.2 we have fixed $q_0 = q_1 = 2, \lambda(0) = 4$ and have varied the input 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 $\frac{\partial}{\partial y} c(y^*) = \beta r$ 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 input rate $\lambda(1)$.
In figure 5.3 we have fixed the two input 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 input rate $\bar{\lambda} = \frac{1}{2}(\lambda(0) + \lambda(1)) = 2.5$. Hence we would expect

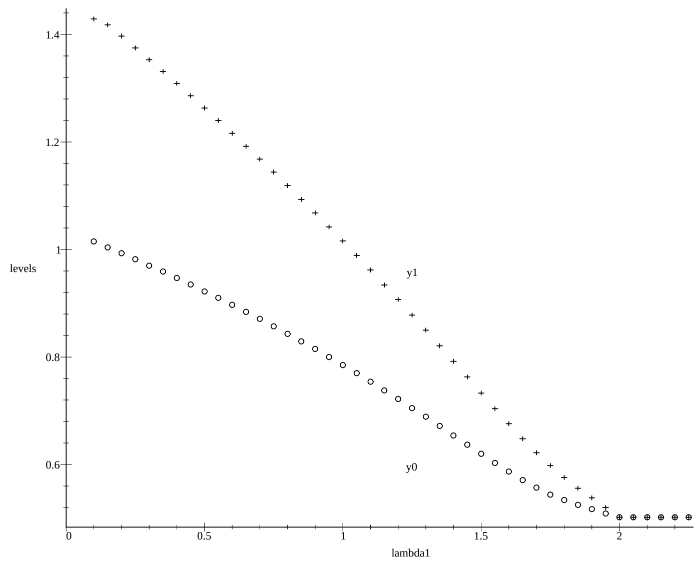

fig.5.2 : Optimal thresholds - variation of $\lambda(1)$

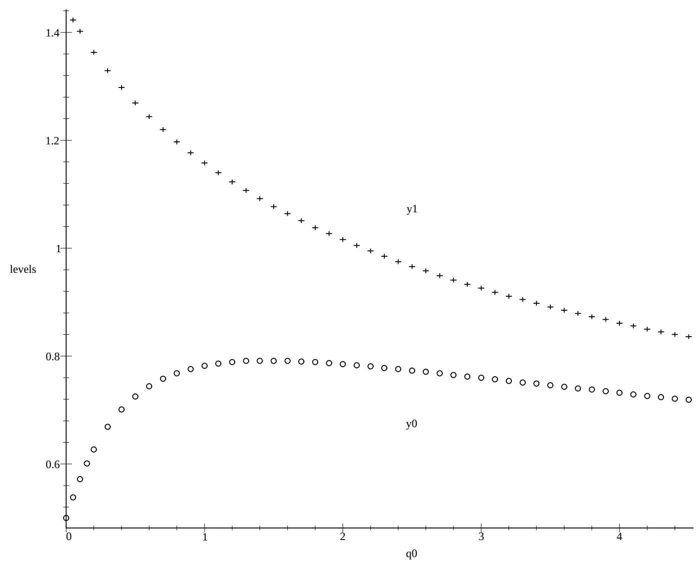

fig.5.3 : Optimal thresholds - variation of q_0

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 input rate $\bar{\lambda}$. This can be observed from the numerical data. 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.

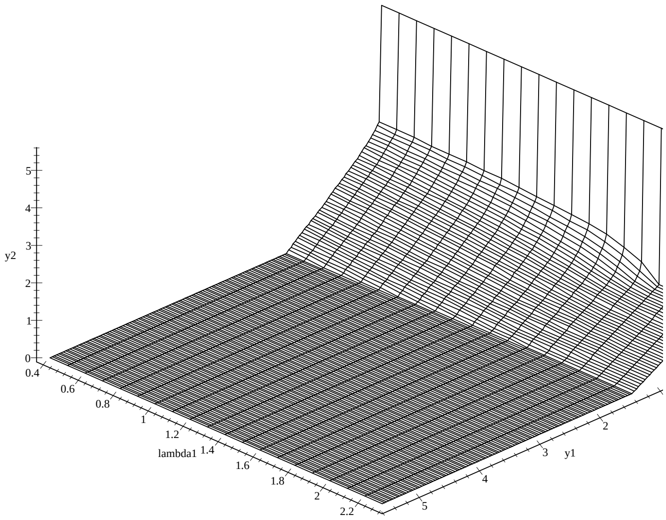

fig.5.4 : Optimal policy in state $z = 0$ - variation of λ_1

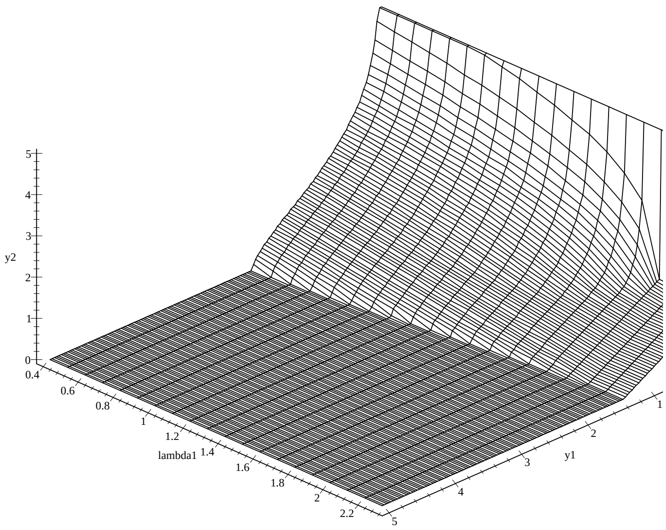

fig.5.5 : Optimal policy in state $z = 1$ - variation of λ_1

The figures 5.4-5.7 for the **two-buffer case** are quite similar. 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 Bäuerle (1998b) we know that the optimal policy is characterized by 4 switching-curves $S_1^0(y_1), S_2^0(y_1), S_1^1(y_1)$ and $S_2^1(y_1)$ where $y_2 \leq S_j^z(y_1)$ if and only if the whole fluid is accepted into buffer j in environment state z, when the buffer levels are y_1 and y_2 respectively. Since the data is symmetric in buffer 1 and 2, the optimal policy is also symmetric. Hence we can restrict w.l.o.g. to the policy for the first buffer. In figure 5.4 and 5.5 we see the optimal policy for buffer 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 acceptance region.

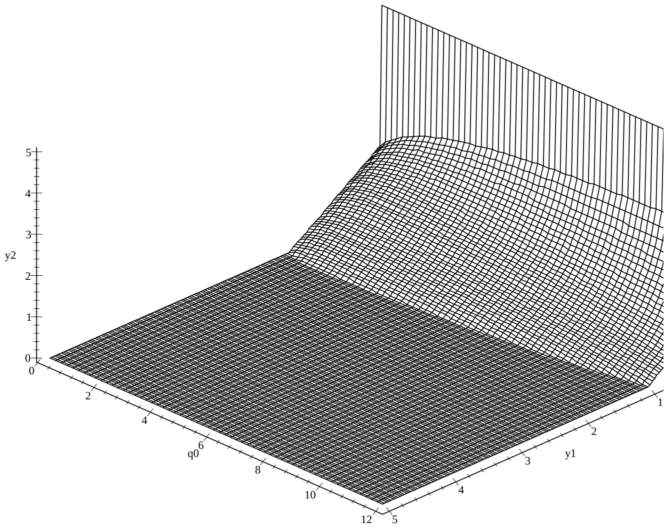

fig.5.6 : Optimal policy in state $z = 0$ - variation of q_0

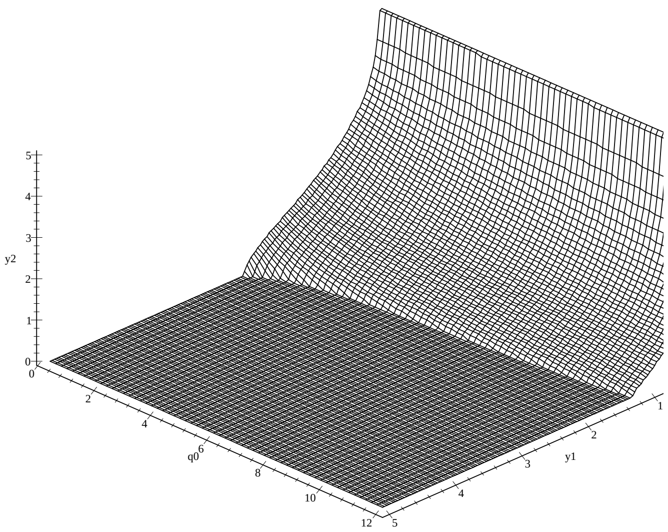

fig.5.7 : Optimal policy in state $z = 1$ - variation of q_0

It seems that the optimal policy in the two-buffer case has the same properties as in the one-buffer case, that is: as soon as $\lambda_1(1) > \mu_1$, the policy does not change;

both acceptance regions increase when $\lambda_1(1)$ decreases and the acceptance region in environment state 1 is always greater than the one in environment state 0.
In figure 5.6 and 5.7 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 5.6 refers to the optimal policy in environment state 0, figure 5.7 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 acceptance 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.

6 Applications

We present now some important examples for SFPs and apply our results of the previous sections. SFPs typically arise in models for manufacturing and telecommunication systems. In both cases we encounter events that occur on different time scales, where the faster one is modeled as a deterministic flow. For example the cell stream sources in ATM multiplexers are often modeled as on-off sources, where we have a certain inflow rate into a buffer when the source is on (talkspurt state) and no inflow, when the source is off (silent state). The duration of the state length is random.
In the sequel we investigate *three examples*. The first one is formulated in the framework of manufacturing models and is a so-called parallel machine problem with backlog. In the one and two product model we show that the optimal production policies are threshold and switching-curve policies, respectively. The second example is a single-server network. Here we can prove that the optimal scheduling policy is an index policy. The last example is the problem of routing to parallel queues. In the case of equal holding cost we show the optimality of the least-loaded routing policy, in the case of two buffers and arbitrary cost we derive the optimality of a switching curve policy.

6.1 Multi-Product Manufacturing Systems

A typical example for SFPs are stochastic manufacturing systems with machine failures and/or demand changes. Since there are different timescales for the occurence of events (for example the production process itself evolves faster than random breakdowns of the machines), quantities that vary faster are replaced with their averages. The cost of such systems are inventory/backlog cost as well as production cost which are often assumed to be convex. The task is to find the optimal production rate of the machines, which involves a scheduling problem in dynamic jobshop systems. For a discussion of recent models and approaches see Sethi/Zhang (1994).
In this section we will investigate a parallel machine system with backlog. The example is taken from Sethi et al. (1998). We consider a multi-product manufacturing system with stochastic production capacity and constant demand for each product over time. The vector $y = (y_1, \dots, y_N)$ gives the inventory/backlog of each product and we assume $S = I\!R^N$. There are a number of parallel machines for manufacturing which are subject to random breakdown and repair. Hence $\lambda(z) \in I\!R_+, z \in Z$ gives the production capacity of the system that is available. The vector $u \in U = \{u \in [0,1]^N \mid \sum_{j=1}^N u_j \leq 1\}$ contains the percentages of the production capacity that are assigned to each of the products. If we denote by $\mu \in I\!R_+^N$ the constant demand rate, the dynamics of the system are for $a \in A, x \in E, t \geq 0$

$$y_t = \phi_t(x,a) = y + \int_0^t \lambda(z)a_s - \mu\, ds.$$

The function $c : I\!R^{2N} \to I\!R_+$ denotes the surplus (inventory/backlog) and production cost. In our notation we can summarize the SFP as follows

$$\begin{aligned} E &= I\!R^N \times Z \\ U &= \{u \in [0,1]^N \mid \sum_{j=1}^N u_j \leq 1\} \\ b^z(u) &= \lambda(z)u - \mu \end{aligned}$$

We will first look at the discounted model with interest rate $\beta > 0$. In order to apply our results we have to impose the following assumptions on the cost rate function:

Assumption 6.1:

(i) $(y, u) \mapsto c(y, z, u)$ is convex for all $z \in Z$.

(ii) c satisfies the growth conditions of Assumption 2.3 and 3.2.

(iii) $y \mapsto c(y, z, u)$ is continuously differentiable for all $z \in Z, u \in U$.

Under this assumption it has been shown in Sethi et al. (1998) that the value function V satisfies the following growth condition

$$|V(y,z) - V(y',z)| \leq C_0\Big(1 + \|y\|^{k+1} + \|y'\|^{k+1}\Big)\|y - y'\|. \tag{6.13}$$

Obviously our Assumptions 2.1 and 2.2 are fulfilled. From Theorem 2.5, Lemma 2.8 and Theorem 4.3 we immediately obtain Theorem 6.1.

Theorem 6.1: (β*-Discounted case*)
Under the preceding Assumption 6.1 it holds for the parallel machine problem that

a) There exists a β-discounted optimal policy.

b) The value function V is continuously differentiable w.r.t. y and satisfies the HJB-equation

$$(\beta + q_z)V(x) = \min_{u \in U}[c(x,u) + V_y(x)(\lambda(z)u - \mu)] + \sum_{z' \neq z} q_{zz'}V(y, z') \tag{6.14}$$

Moreover, V fulfills the growth condition (6.13).

c) If there exists an $f^* \in F$ such that for almost every $t \geq 0$

$$\begin{aligned} (\beta + q_z)V(\phi_t(x, f^*), z) \;&=\; c(\phi_t(x,f^*), z, f^*(x)(t)) + V_y(\phi_t(x,f^*))(\lambda(z)f^*(x)(t) \\ &\quad -\mu) + \sum_{z' \neq z} q_{zz'}V(\phi_t(x,f^*), z') \end{aligned}$$

then $(f^*, f^*, \ldots)$ is an optimal policy.

Now we will look at the same model with the c-average cost optimality criterion. In this case we need a further assumption. Note that due to our previous assumptions, the cost rate function is coercive (see Remark 3.3).

Assumption 6.2:
Suppose that ν is the stationary distribution of the environment process (Z_t). Then we assume $\sum_z \lambda(z)\nu_z > \sum_{j=1}^N \mu_j$.

It has been shown in Sethi et al. (1998) Theorem 3 (cf. also Sethi et al. (1997) Theorem 3.3) that with $\xi = (0,0) \in E$ (w.l.o.g. suppose $0 \in Z$) we have

$$|h^\beta(x)| \le C_0(1 + \|y\|^{k+2}) =: M(y)$$

for all $x \in E, \beta > 0$, where $C_0 \in I\!R_+$ is independent of β. Moreover, from Theorem 2 in Sethi et al. (1998) we can conclude that there exists a policy π such that $G_\pi(x) < \infty$ for all $x \in E$ (cf. Lemma 3.8). Hence using Lemma 3.9, Assumption 3.1 is fulfilled. Since the convexity assumptions of Remark 3.4 are fulfilled we obtain with Theorem 3.6, Theorem 4.4 (notice that Assumption 3.3 (ii) is also valid) and Remark 3.4

Theorem 6.2: (*Average case*)
Under the preceding Assumptions 6.1 and 6.2 it holds for the parallel machine problem that

a) There exists a decision rule f^0 and sequences $\beta_m(x) \to 0$, $x_m(x) \to x$ such that

$$f^0(x) = \lim_{m\to\infty} f^{\beta_m(x)}(x_m(x)),$$

where f^β is an optimal decision rule in the β-discounted model and the stationary policy $(f^0, f^0, \ldots)$ is c-average optimal, provided $J_{f^0} \ge G_{f^0}$.

b) There exists a continuous and convex function $h : E \to I\!R$ and a constant $\rho \ge 0$ such that h is a viscosity solution of the HJB-equation

$$\rho + qh(x) = \min_{u\in U} [c(x,u) + h_y(x)(\lambda(z)u - \mu)] + q \sum_{z'} p_{zz'} h(y,z') \tag{6.15}$$

c) If some continuously differentiable function $w(\cdot, z) : S \to I\!R$ satisfies (6.15) and if there exists a decision rule $f^* \in F$ such that (4.11) and (4.12) hold and $J_{f^*} \ge G_{f^*}$, then $\rho = G_{f^*}(x)$ for $x \in E$ and $(f^*, f^*, \ldots)$ is a c-average optimal policy.

Remark 6.1:
The same analysis can be carried out if we allow for a stochastically varying demand, i.e. there is a second continuous-time Markov chain (Z_t^2) with finite state space which is independent of everything else and determines the demand rate $\mu(z)$. Obviously it is possible to construct a common continuous-time Markov chain (Z_t) such that the dynamics of the system is given by $b^z(u) = \lambda(z)u - \mu(z)$.

In the cases of one ($N = 1$) or two products ($N = 2$) it is possible to further derive some structured properties for the optimal policies under suitable assumptions on the cost rate function.

A) One-Product System
Suppose now that $N = 1$, i.e. we have a one-product system and that

Assumption 6.3:
The cost rate function is of the form $c(y, z, u) = c(y) + \tilde{c}u$, with $\tilde{c} \in I\!R_+$.

Let us first look at the β-discounted problem. Assumption 6.1 also implies that V is convex in y (cf. Lemma 2.6), hence V_y is increasing. In this case we obtain from the HJB-equation (6.14) that the optimal policy is given by a threshold feedback control. A feedback control is a function $g : E \to U$ such that $y_t = y + \int_0^t b^z(g(y_s))ds$ has a unique solution and the open loop control $u(x, t) = g(y_t)$ is admissible.

Corollary 6.3: *(β-Discounted case)*
In addition to the assumptions of Theorem 6.1 we assume that $N = 1$ and that the cost rate function satisfies Assumption 6.3. Then the optimal stationary policy $(f^*, f^*, \ldots)$ in the β-discounted model is given by a threshold feedback control g, i.e. there exists a function $S : Z \to I\!R$ such that

$$g(x) = \begin{cases} 1 & , y < S(z) \\ \min\{1, \frac{\mu}{\lambda(z)}\} & , y = S(z) \\ 0 & , y > S(z) \end{cases}$$

and $f^*(x)(t) = g(\phi_t(x, f^*))$.

Remark 6.2:
The definition of $g(x)$ at $y = S(z)$ is arbitrary. We have chosen it in such a way that the inventory - if possible - stays the same. If $|Z| = 2$, the function $S(z)$ can be computed explicitely. This has been done in Akella/Kumar (1986). For arbitrary Z it is possible to derive monotonicity properties of S. In Sethi/Zhang (1994) one

finds statements if (Z_t) is a birth-and-death process. For a more general concept using stochastic orderings, see Rajagopal et al. (1995).

In the average cost case we obtain

Corollary 6.4: *(Average case)*
In addition to the assumptions of Theorem 6.2 we assume that $N = 1$ and that the cost rate function satisfies Assumption 6.3. Then the optimal stationary policy $(f^*, f^*, \ldots)$ in the c-average cost model is given by a threshold feedback control g, i.e. there exists a function $S : Z \to I\!R$ such that

$$g(x) = \begin{cases} 1 & , y < S(z) \\ \min\{1, \frac{\mu}{\lambda(z)}\} & , y = S(z) \\ 0 & , y > S(z) \end{cases}$$

and $f^*(x)(t) = g(\phi_t(x, f^*))$. Moreover, there exists a sequence $\beta_m \to 0$ such that $S^{\beta_m}(z) \to S(z)$ for $z \in Z$, where S^{β_m} is the optimal threshold function in the β_m-discounted model.

Proof: From Corollary 6.3 we know that for $\beta > 0$, the optimal policy is given by a threshold feedback control with threshold function $S^\beta : Z \to I\!R$, i.e.

$$f^\beta(x)(t) = \begin{cases} 1 & , \phi_t(x, f^\beta) < S^\beta(z) \\ \min\{1, \frac{\mu}{\lambda(z)}\} & , \phi_t(x, f^\beta) = S^\beta(z) \\ 0 & , \phi_t(x, f^\beta) > S^\beta(z) \end{cases}$$

Hence we can choose in the proof of Theorem 3.11 a subsequence $\{\beta_m\}$ of $\{\beta_n\}$ such that $S^{\beta_m}(z) \to S(z)$ for $z \in Z$ and $m \to \infty$. First we have to verify that for g, the integral equation

$$\phi_t(x) = y + \int_0^t \lambda(z) g(\phi_s(x), z) - \mu \, ds$$

has a unique solution. As far as existence is concerned, it is possible to compute a solution explicitly. However this involves to distinguish several cases. For example if $\lambda(z) > \mu$ and $y > S(z)$ we have

$$\phi_t(x) = \begin{cases} y - t\mu & , 0 \le t \le t(x) \\ S(z) & , t > t(x) \end{cases}$$

with $t(x) = (y - S(z))/\mu$. Uniqueness follows e.g. from Hartman (1982) Theorem 6.2. Using Lemma A.4 it can be readily verified that $f^{\beta_m(x)}(x_m(x)) \to f(x)$ for $m \to \infty$, where f is constructed from a feedback control g with threshold function

S. This again makes it necessary to distinguish several cases. We will only look at the case $\lambda(z) > \mu$ and $y > S(z) \in I\!R$. We have to show

$$\int_0^\infty \int_U \psi(t,u) f^{\beta_m}(x_m)(t,du)dt \to \int_0^\infty \int_U \psi(t,u) f^0(x)(t,du)dt$$

for all measurable functions ψ, with $u \mapsto \psi(t,u)$ is continuous for all $t \geq 0$ and $\int_0^\infty \sup_{u\in U} |\psi(t,u)|\ dt < \infty$. W.l.o.g. suppose $y - S(z) > 3\epsilon$ for $\epsilon > 0$. Choose $N_0(\epsilon)$ big enough such that for all $m \geq N_0(\epsilon)$:

$$z_m = z,\ |y - y_m| \leq \epsilon,\ |S^{\beta_m}(z) - S(z)| \leq \epsilon$$

and thus $y_m > S^{\beta_n}(z)$ for all $m, n \geq N_0(\epsilon)$. Hence we obtain with $t_m(x) := (y_m - S^{\beta_m}(z))/\mu$

$$\left| \int_0^{t(x)} \psi(t,0)dt + \int_{t(x)}^\infty \psi(t, \frac{\mu}{\lambda(z)})dt - \int_0^{t_m(x)} \psi(t,0)dt - \int_{t_m(x)}^\infty \psi(t, \frac{\mu}{\lambda(z)})dt \right| \leq$$

$$\leq \int_{t(x)}^{t_m(x)} |\psi(t,0)|dt + \int_{t(x)}^{t_m(x)} |\psi(t, \frac{\mu}{\lambda(z)})|dt \to 0 \text{ for } m \to \infty,$$

since $t_m(x) \to t(x)$ for $m \to \infty$ which implies the statement. Last but not least, we have to show that the transition kernel induced by the feedback control g is weakly continuous which yields the result by Lemma 3.10. Hence we have to show

$$x \mapsto \sum_{z'} p_{zz'} \int_0^\infty e^{-qt} v(\phi_t(x), z')dt$$

is continuous and bounded, for all continuous and bounded $v : E \to I\!R$, where ϕ_t is generated by g. Boundedness is clear. As far as continuity is concerned we will again look at the case $\lambda(z) > \mu$ and $y > S(z) \in I\!R$. Thus for N_0 big enough, $z_m = z$ for all $m \geq N_0$ and it remains to show that

$$y \mapsto \int_0^\infty e^{-qt} v(\phi_t(x), z')dt$$

is continuous for all $z' \in Z$. We have

$$\lim_{m\to\infty} \int_0^\infty e^{-qt} v(\phi_t(y_m, z), z')dt = \lim_{m\to\infty} \left(\int_0^{t_m(x)} e^{-qt} v(y_m - t\mu, z')dt + \frac{e^{-qt_m(x)}}{q} v(S_m(z)) \right)$$

$$= \lim_{m\to\infty} \left(\int_0^{t(x)} e^{-qt} v(y_m - t\mu, z')dt + \frac{e^{-qt(x)}}{q} v(S_m(z)) \right) = \int_0^\infty e^{-qt} v(\phi_t(x), z')dt$$

where we have used dominated convergence in the last step. □

B) Two-Product System

An investigation of this model can be found in Rajagopal (1995) section 4.9.3. We will restrict to the β-discounted model here. The average cost model can be solved in the same spirit as in the one-product system. Suppose now that $N = 2$, i.e. we have a two-product system and that the cost rate function has the following properties

Assumption 6.4:

(i) $c(y, z, u) = c(y) + \tilde{c}u,\ \tilde{c} \in I\!R^2_+$.
(ii) c satisfies the growth condition of Assumption 2.3 and 3.2.
(iii) $y \to c(y)$ is strictly convex and continuously differentiable.
(iv) $\left(\frac{\partial c}{\partial y_1} - \frac{\partial c}{\partial y_2}\right)(y)$ is strictly increasing in y_1 and strictly decreasing in y_2.
(v) $\frac{\partial c}{\partial y_1}$ is strictly increasing in y_2 and $\frac{\partial c}{\partial y_2}$ is strictly increasing in y_1.
(vi) $\frac{\partial c}{\partial y_1}(y + h(e_1 - e_2))$ is strictly increasing in $h > 0$, $\frac{\partial c}{\partial y_2}(y + h(e_1 - e_2))$ is strictly decreasing in $h > 0$.

The property in Assumption 6.4 (v) is also called supermodularity or one says that c has monotone differences. The assumptions are e.g. fulfilled if c is separable, i.e. $c(y_1, y_2) = c_1(y_1) + c_2(y_2)$ and each $c_i, i = 1, 2$ is strictly convex, continuously differentiable and satisfies the growth conditions. Here it is possible to show

Theorem 6.5:
Under the preceding Assumption 6.4 the optimal policy is given by a feedback control of switching-type. The feedback control can be characterized by a partition $P_1 + P_2 + SP$ of $I\!R^2$ such that

a) if $y \in SP$, then it is optimal to stop production and $y \in SP,\ y' \geq y$ implies $y' \in SP$.
b) if $y \in P_1$, then it is optimal to produce item 1 only and if $y' \notin SP,\ y'_1 \leq y_1, y'_2 > y_2$ then $y' \in P_1$.
c) if $y \in P_2$, then it is optimal to produce item 2 only and if $y' \notin SP,\ y'_2 \leq y_2, y'_1 > y_1$ then $y' \in P_2$.

The proof that this policy is ε-optimal can be found in Rajagopal (1995). However, he did not prove that V is continuously differentiable. Since this is true due to Lemma 2.8 we can complete his proof.

6.2 Single-Server Networks

A well-known problem in queueing theory is the so-called Klimov Problem (see e.g. Klimov (1974), Walrand (1988)). Here we consider its fluid analogue which is a generic single-server fluid network (fig.6.1). There are N buffers with infinite capacity which receive fluid from outside at rate $\lambda(z) = (\lambda_1(z), \ldots, \lambda_N(z))$, if at time t the environment process $Z_t = z$. The vector $y = (y_1, \ldots, y_N)$ gives the buffer content and we assume $S = I\!R_+^N$. A single server has to be splitted among the buffers.

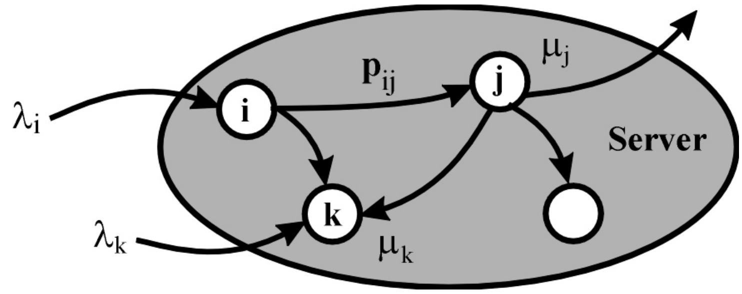

fig.6.1 : Single-server network

The potential service rate of buffer j is assumed to be $\mu_j > 0,\ j = 1, \ldots, N$ which means that if a fraction $u_j \in (0,1)$ of the server is allocated to buffer j, there is an output of rate $u_j\mu_j$. For abbreviation denote the matrix $D = diag(\mu_j)$ as the diagonal matrix with elements μ_j on the diagonal. The fluid that is leaving buffer j is divided and a fraction of $p_{ji} \in [0,1),\ i = 1, \ldots, N$ is instantaneously flowing into buffer i. Denote the matrix $P = (p_{ji})$ and define $U = \{u \in [0,1]^N \mid \sum_{j=1}^N u_j \leq 1\}$. Hence, given a fixed server allocation $u \in U$ and a fixed environment state z, the input rate into buffer j is $\lambda_j(z) + \sum_{i=1}^N p_{ij}\mu_i u_i$ and the output rate is equal to $\mu_j u_j$. In matrix notation this is $\lambda(z) + P^T Du$ and Du respectively.
We suppose that a linear cost of rate $c_j \in I\!R$ is incurred, when holding fluid in buffer j. Denote $c = (c_1, \ldots, c_N)$. In our notation we can summarize the SFP as follows, if we introduce the matrix $A := D(I - P)$ (I denotes the identity matrix)

$$
\begin{aligned}
E &= I\!R_+^N \times Z \\
U &= \{u \in [0,1]^N \mid \sum_{j=1}^N u_j \leq 1\} \\
b^z(u) &= \lambda(z) - A^T u \\
c(x,u) &= cy
\end{aligned}
$$

Let us first investigate the β-discounted problem. A different optimization problem is obtained, when we suppose that we obtain a reward of $r_j \in I\!R$ for each unit u_j, we allocate the server to buffer j. Denote $r = (r_1, \ldots, r_N)$. The aim here is to maximize the expected discounted reward of the system over an infinite horizon. In Bäuerle/Rieder (1999) it has been shown that if we define $r = Ac$, the optimal policies for both optimization problems are the same. We will impose the following assumptions, where we fix $r = Ac$ throughout the section.

Assumption 6.5:

(i) Suppose that P is transient, i.e. $\sum_{n=0}^{\infty} P^n < \infty$.
(ii) The rewards r are non-negative.

Assumption 6.5(i) is for example fulfilled if $\sum_{i=1}^{N} p_{ji} < 1$, for all $j = 1, \dots, N$ i.e. a positive fraction of $1 - \sum_{i=1}^{N} p_{ji}$ is leaving the system. Our aim is to prove that the optimal policy is a priority index policy. The deterministic version of this problem has also been investigated in Chen/Yao (1993). Using a linear programming approach the authors there showed that the index policy is a myopic solution of the optimization problem and gave conditions under which the myopic solution is also globally optimal.
We will next give a definition of the indices and some important properties. In the following subsections we will define the index policy and present a proof of its optimality.

Definition and Properties of the Indices
Due to Assumption 6.5 we have that $(I-P)^{-1} = \sum_{n=0}^{\infty} P^n \geq 0$ and hence

$$A^{-1} = \sum_{n=0}^{\infty} P^n D^{-1} \geq 0.$$

Obviously this relation holds for arbitrary quadratic submatrices of A. For means of short notation, let us introduce the following abbreviation: For a subset $S \subset \{1, \dots, N\}$ we denote

$$a_i^S = (-a_{ij})_{j \in S} = (\mu_i p_{ij})_{j \in S}, \ i \notin S \quad \text{and} \quad A_S = (a_{ij})_{i,j \in S}.$$

An analogous definition is used for vectors.

Now we will give a recursive definition of the indices $I^1, \dots, I^N$, the so-called **largest remaining index algorithm** (the name will be justified by Lemma 6.6 a)). It is closely connected to the reward model. By $\mathbb{1}$ we denote the vector consisting of 1's only - the dimension should be clear from the context.

Largest remaining index algorithm:

(i) $I^1 = \max_{1 \leq j \leq N} \ r_j, \ \ i_1 = \text{argmax}_{1 \leq j \leq N} \ r_j, \ \ S_1 = \{i_1\}$.
(ii) For $k = 1, \dots, N-1$ let

$$I_j^{k+1} = \frac{r_j + a_j^{S_k} A_{S_k}^{-1} r_{S_k}}{1 + a_j^{S_k} A_{S_k}^{-1} \mathbb{1}}, \quad j \notin S_k$$

$$I^{k+1} = \max_{j \notin S_k} I_j^{k+1}, \quad i_{k+1} = \text{argmax}_{j \notin S_k} \ I_j^{k+1}$$

Set $S_{k+1} = S_k + \{i_{k+1}\}$.

Buffer i_k is now assigned the index $I^k,\ k = 1, \ldots, N$.

The indices have the following nice interpretation: I^1 is simply the maximal reward rate in the model. Suppose that the indices $I^1, \ldots, I^k$ have already been determined. Given that we have to keep the buffer contents of the buffers in S_k at zero we can now look at the reduced network which consists of the buffers in $\{1, \ldots, N\} - S_k$. If we allocate a unit of the server to buffer $j \notin S_k$, in order to keep the buffers in S_k empty we have to assign to them a server capacity u_{S_k} which can be computed from

$$0 = A_{S_k}^T u_{S_k} - a_j^{S_k}.$$

Therefore $u_{S_k} = a_j^{S_k} A_{S_k}^{-1}$. Hence I_j^{k+1} is the reward rate of buffer j in the reduced network.
The following properties of the indices will be crucial in the sequent proofs.

Lemma 6.6:
Under Assumption 6.5, the indices computed by the largest remaining index algorithm fulfill

a) $I^1 \geq I^2 \geq \ldots \geq I^N \geq 0$.
b) For $1 \leq k \leq N-1$

$$A_{S_k}^{-1}(I^{k+1}\mathbb{1} - r_{S_k}) \leq 0.$$

c) For $1 \leq j < k \leq N-1$

$$e_j A_{S_k}^{-1}(I^{k+1}\mathbb{1} - r_{S_k}) \leq e_j A_{S_{k-1}}^{-1}(I^k\mathbb{1} - r_{S_{k-1}}).$$

Proof: a) and b) see Bäuerle/Rieder (1999).
c) W.l.o.g. we assume that $i_k = k$ for $k = 1, \ldots, N$. Hence we have

$$A_{S_k} = \begin{pmatrix} A_{S_{k-1}} & a_S \\ -a_k^{S_{k-1}} & a_{kk} \end{pmatrix}.$$

When we define $w_k = a_k^{S_{k-1}} A_{S_{k-1}}^{-1}$ we can write

$$I^k = \frac{r_k + w_k r_{S_{k-1}}}{1 + w_k \mathbb{1}}.$$

Moreover, the definition $z^j = e_j A_{S_k}^{-1}$, $1 \le j < k$ gives us

$$z^j = \begin{pmatrix} w_k \alpha + e_j A_{S_{k-1}}^{-1} \\ \alpha \end{pmatrix} \quad \text{where} \quad \alpha = \frac{-a_S e_j A_{S_{k-1}}^{-1}}{a_{kk} + a_S w_k}.$$

Using the introduced notation we obtain

$$\begin{aligned}
& e_j A_{S_{k-1}}^{-1} \left(I^k \mathbb{1} - r_{S_{k-1}} \right) - e_j A_{S_k}^{-1} \left(I^{k+1} \mathbb{1} - r_{S_k} \right) \\
= \;& e_j A_{S_{k-1}}^{-1} \left(I^k \mathbb{1} - r_{S_{k-1}} \right) - z^j \left(I^{k+1} \mathbb{1} - r_{S_k} \right) \\
= \;& e_j A_{S_{k-1}}^{-1} \left(I^k \mathbb{1} - r_{S_{k-1}} \right) - (w_k \alpha + e_j A_{S_{k-1}}^{-1}) \left(I^{k+1} \mathbb{1} - r_{S_{k-1}} \right) - \alpha (I^{k+1} - r_k) \\
= \;& (I^k - I^{k+1}) e_j A_{S_{k-1}}^{-1} \mathbb{1} + \alpha (w_k \mathbb{1} + 1)(I^k - I^{k+1}) \\
= \;& (I^k - I^{k+1}) z^j \mathbb{1} = (I^k - I^{k+1}) e_j A_{S_k}^{-1} \mathbb{1} \ge 0
\end{aligned}$$

where the non-negativity is valid due to a) and the fact that $A_{S_k}^{-1} \ge 0$. □

Definition of the Priority Index Policy

The priority index policy is now defined as follows: assign the complete server to the non-empty buffer with highest index as long as there is fluid in this buffer. When the buffer is empty, assign to it only the capacity that is needed to hold the buffer at zero and assign the rest of the server to the buffer with second highest index and so on. Since there can be re-entrants from the newly processed buffer to buffers with higher priority, this procedures makes it necessary to re-assign the server capacity to all the buffers at each time point when a buffer empties.
Assume that the buffers have been rearranged such that the natural order coincides with the priority order i.e. $i_k = k$ and $S_k = \{1, \ldots, k\}$, $k = 1, \ldots, n$. Under different environment states the number of buffers which can be emptied is different.

Definition 6.1:

The environment state $z \in Z$ is called *d-stable*, $d \in \{0, 1, \ldots, N\}$ if there exists a $u \in U$ such that $\lambda_j(z) < (A^T u)_j$ for $j = 1, \ldots, d$ and $\lambda_j(z) \ge (A^T u)_j$ for $j = d+1, \ldots, N$ and all $u \in U$.

In a d-stable environment, the server can empty the first d buffers irrespective of the initial buffer contents. Let $z \in Z$ be a given d-stable environment state. We will define the following server allocations

$$u^*(z, 1) = (1, 0, \ldots, 0)$$

$$u^*(z, k+1) = (\lambda_{S_k}(z) A_{S_k}^{-1} + \varepsilon a_{k+1}^{S_k} A_{S_k}^{-1}, \varepsilon, 0, \ldots, 0), \quad k = 1, \ldots, d < N$$

where $\varepsilon > 0$ is chosen such that $\mathbb{1} u^*(z, k+1) = 1$

and

$$u^*(z, N+1) = \lambda(z) A^{-1} \text{ if } z \text{ is } N-\text{stable}.$$

Note that all server allocations are admissible due to our definition of d-stability. $u^*(z, k+1)$ is exactly the server allocation which can hold buffers $1, \ldots, k$ at zero and work on the fluid in buffer $k+1$. $u^*(z, N+1)$ can hold the complete system empty.
Formally we can define the *priority index policy* as the stationary feedback policy $\pi = f^\infty$ with

$$f(x)(t) = u^*(z, j) \quad \text{if} \quad j = \min\{i \mid y_i(t) > 0\} \wedge (d+1),\ z \text{ is } d-\text{stable},$$

where $y_t = \phi_t(x, f(x))$. For a fixed initial state we will call the action $f(x)$ index rule.

Theorem 6.7: *(Optimality of the priority index policy - β-discounted case)*
Under Assumption 6.5, the priority index policy is optimal for the discounted cost model.

Remark 6.3:
An important special case is $P = 0$, i.e. there is no routing and processed fluid leaves the system immediately. Assumption 6.5 (i) is then fulfilled. Moreover, we obtain $r_j = \mu_j c_j$ and $a_j^S = 0$ for arbitrary $j \notin S$. Therefore, the LRI-algorithm gives $I^{k+1} = \max_{j \notin S_k} r_j$ and the priority index policy is the well-known μc-rule. In the deterministic setting this result can also be found in Avram et al. (1995).

Example:
We will illustrate the operation of the priority index policy by means of the following deterministic 5 buffer example:

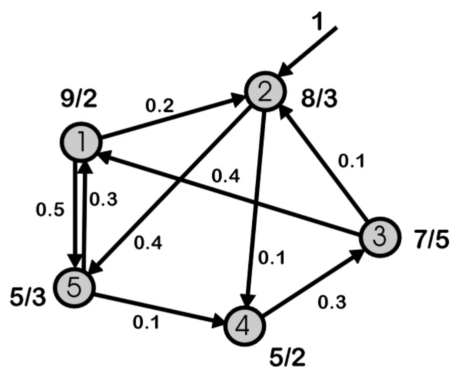

fig.6.2 : Single-server network

The numbers next to the buffer give the reward rate r_j and the server rate μ_j respectively. We have $\lambda_1 = 1$, all other external input rates are zero. The routing probabilities are indicated at the edges. The indices of the buffers computed with the LRI-algorithm are: $I(1) = 9, I(2) = 8, I(3) = 8, I(4) = 5.7, I(5) = 6.3$. Figure 6.3 gives the path of the optimal trajectory. A vertical cut at time t gives the amount of fluid in the buffers at time t, hence $y = (1, 1, 1, 1, 1)$ is the buffer content at $t = 0$.

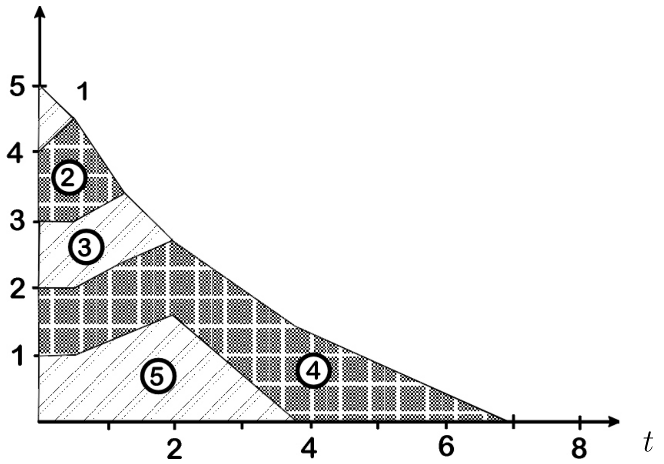

fig.6.3 : Optimal trajectory

Proof of the optimality of the Priority Index Policy
For the proof we use the method of policy iteration which is explained in Section 4. Assumption 2.1 and 2.2 are obviously fulfilled. Let us first look at the following generic control problem with function $l : I\!R_+^N \to I\!R_+$

$$(CP)\begin{cases} \int_0^\infty e^{-(\beta+q_z)t} l(\phi_t(x,a))\, dt \;\to\; \min \\ \phi_t(x,a) = y + \int_0^t \lambda(z) - A^T a_s\, ds \\ \phi_t(x,a) \in I\!R_+^N \\ a_t \in U \text{ for all } t \geq 0 \end{cases}$$

In the sequel, $z \in Z$ is fixed and we will suppress it in our notation. $V(y)$ denotes now the value function of (CP).

Assumption 6.6:

(i) l is increasing and convex.

(ii) There exists a constant $C_0 \in I\!R_+$ such that for all $y, y' \in I\!R_+^N$

$$|l(y) - l(y')| \leq C_0\,(1 + \|y\| + \|y'\|)\,\|y - y'\|.$$

(iii) l is continuously differentiable.

Lemma 6.8:
Suppose that the index rule is optimal for problem (CP) and l fulfills Assumption 6.6. Then

a) $y \mapsto V(y)$ is increasing and convex.
b) There exists a constant $C_0' \in I\!R_+$ such that for all $y, y' \in I\!R_+^N$

$$|V(y) - V(y')| \le C_0' \,(1 + \|y\| + \|y'\|)\, \|y - y'\|.$$

c) $y \mapsto V(y)$ is continuously differentiable.

Proof: Again let $i_k = k$ for $k = 1, \dots, N$.
a) The convexity follows immediately from Lemma 2.6. For the monotonicity let $y \in I\!R_+^N, h > 0$. It suffices to show that $V(y+he_j) \ge V(y)$ for $j = 1, \dots, N$. Denote by $a_h(t)$ the optimal index rule, starting in $y + he_j$. Suppose that the additional amount of fluid h is colored red, all other fluid has color blue. Moreover, we assume that red fluid is always the last to process in the buffers. Define the control

$$a(t) = \begin{cases} a_h(t) & \text{, if } a_h(t) \text{ processes blue fluid only.} \\ (\lambda_{S_k} A_{S_k}^{-1}, 0) & \text{, if } a_h(t) \text{ processes red fluid and buffer } 1, \dots, k \text{ are empty.} \end{cases}$$

Obviously $a \in D(y)$ and $\phi_t(y, a) \le \phi_t(y + he_j, a_h)$ for all $t \ge 0$. Hence we have due to the monotonicity of l that $V(y) \le LV(y, a) \le LV(y + h, a_h) = V(y + h)$.

b) We first show the statement for $y' = y + he_j,\ h > 0,\ j = 1, \dots, N$. Denote by a_t the optimal index rule, starting in y. Obviously $a \in D(y + he_j)$ and $\|\phi_t(y, a) - \phi_t(y + he_j, a)\| = h$ for all $t \ge 0$, where $\| \cdot \|$ is the L_1-norm here. Due to the assumptions on l we obtain

$$\begin{aligned} |V(y + he_j) - V(y)| &\le LV(y + he_j, a) - LV(y, a) \\ &= \int_0^\infty e^{-(\beta+q_z)t} |l(\phi_t(y + he_j, a)) - l(\phi_t(y, a))| \, dt \\ &\le \int_0^\infty e^{-(\beta+q_z)t} C_0 \Big(1 + \|\phi_t(y + he_j, a)\| + \|\phi_t(y, a)\|\Big) h \, dt \end{aligned}$$

Since the dynamics are linear we have that $\|\phi_t(y, a)\| \le \|y\| + bt$, with $b > 0$. Thus

$$\begin{aligned} |V(y + he_j) - V(y)| &\le \int_0^\infty e^{-(\beta+q_z)t} C_0 \Big(1 + \|y + he_j\| + \|y\| + 2bt\Big) h \, dt \\ &= \frac{C_0}{\beta + q_z} \Big(1 + \|y + he_j\| + \|y\|\Big) h + \frac{2bC_0}{(\beta + q_z)^2} h \\ &\le C_j \Big(1 + \|y + he_j\| + \|y\|\Big) h \end{aligned}$$

Now let $y, y' \in I\!R_+^N$ be arbitrary. Define $\Delta_j := y_j - y'_j,\ y^0 = y$ and for $k = 1, \dots, N$ define $y^k := y - \Delta_1 e_1 - \dots - \Delta_k e_k,$. Then

$$\begin{aligned} |V(y) - V(y')| &= |V(y) - \sum_{k=1}^{N-1} \Big(V(y^k) - V(y^k) \Big) - V(y')| \\ &\le \sum_{k=1}^{N} |V(y^{k-1}) - V(y^k)| \le \sum_{k=1}^{N} C_k \Big(1 + \|y^{k-1}\| + \|y^k\| \Big) |\Delta_k| \end{aligned}$$

Since $\|y^k\| \le \|y\| + |\Delta_1| + \dots + |\Delta_k| \le \|y\| + \|y - y'\| \le 2\|y\| + \|y'\|$ we obtain

$$|V(y) - V(y')| \le \tilde{C}_0 \Big(1 + 4\|y\| + 2\|y'\| \Big) \sum_{k=1}^{N} |\Delta_k| \le C'_0 \Big(1 + \|y\| + \|y'\| \Big) \|y - y'\|$$

which completes the proof.

c) Since V is convex, it suffices to prove that the partial derivatives exist (cf. Rockafellar (1970)). Define $\phi_t(y) = \phi_t(y, a)$, where a is the optimal index rule. We will first show that $\phi_t(y)$ has continuous partial derivatives with respect to y for almost all $t \ge 0$. Therefore, suppose z is d-stable and denote by $T_k(y)$ and $\phi^k(y)$ the depletion time of buffer $k \le d$ and the state of the buffers at time $T_k(y)$ under the optimal index rule. An easy induction shows that $T_k(y)$ and $\phi^k(y)$ are linear in y, i.e. there exist $B^k \in I\!R^{N \times N},\ b^k \in I\!R^N$ such that $T_k(y) = b^k y,\ \phi^k(y) = B^k y$. Obviously, T_k is increasing in y. Therefore, we obtain with $|h|$ small enough and $t \in (T_k(y), T_{k+1}(y)),\ k < d,\ j = 1, \dots, N$

$$\begin{aligned} \frac{\phi_t(y + he_j) - \phi_t(y)}{h} &= \frac{1}{h} \Big(\phi^k(y + he_j) - \phi^k(y) + (t - T_k(y + he_j)(\lambda - A^T u^*(k)) \\ &\quad -(t - T_k(y))(\lambda - A^T u^*(k)) \Big) = B^k e_j - b^k e_j (\lambda - A^T u^*(k)) \end{aligned}$$

Thus if $t \ne T_k(y),\ k = 1, \dots, N,\ \phi_t(y)$ has continuous partial derivatives. Finally, we get now with Assumption 6.6 (ii) and Bounded Convergence

$$\begin{aligned} \lim_{h \to 0} \frac{V(y + he_j) - V(y)}{h} &= \lim_{h \to 0} \int_0^\infty e^{-(\beta + q_z)t} \frac{1}{h} \left(l(\phi_t(y + he_j)) - l(\phi_t(y)) \right) dt \\ &= \int_0^\infty e^{-(\beta + q_z)t} \lim_{h \to 0} \frac{1}{h} \left(l(\phi_t(y + he_j)) - l(\phi_t(y)) \right) dt \\ &= \int_0^\infty e^{-(\beta + q_z)t} \frac{\partial}{\partial y} l(\phi_t(y)) \frac{\partial}{\partial y_j} \phi_t(y)\, dt \end{aligned}$$

which implies the statement. □

Proof of Theorem 6.7: According to the policy iteration it is sufficient to show that the priority index rule is a minimizer of the n-stage value function V_n for all $n \in I\!N$. Hence we have to show that the priority index rule is the optimal control for (CP) where we define $l(y) := cy + \sum_{z'\neq z} q_{zz'} V_n(y,z'),\ n \in I\!N$. We do this by induction on n using Lemma 4.7, where (i) reads

$$a_t^* \text{ maximizes } a_t \mapsto a_t A p_t \text{ for } a_t \in U.$$

First let us introduce the following definitions and notations, where we assume w.l.o.g. that $i_k = k$ for $k = 1, \dots, N$.
$T(x) \in I\!R_+$ is the time it takes to empty buffer $1, \dots, N$ if z is N-stable and y the initial state, under the index rule. If z is d-stable, $d < N$, then $T(x) := \infty$.
$\phi_t(x) := \phi_t(x, a)$, where a is the priority index rule.
Moreover, we define the following matrix

$$B = (b_1, \dots, b_N) \quad = \quad \begin{pmatrix} c_1 & b_{12} & b_{13} & \cdots & b_{1N} \\ c_2 & c_2 & b_{23} & \cdots & b_{2N} \\ \vdots & \vdots & \vdots & & \vdots \\ c_N & c_N & c_N & \cdots & c_N \end{pmatrix}$$

with

$$(b_{1,k+1}, \dots, b_{k,k+1}) = A_{S_k}^{-1}\left(I^{k+1}1\!\!1 - r_{S_k}\right) + c_{S_k},$$

$k = 1, \dots, N-1$. Due to Lemma 6.6 b), c) we know that $b_1 \geq b_2 \geq \dots \geq b_N$.
Now consider the optimization problem for $n = 0$. Since $V_0 = 0$ we have to solve the control problem for the purely deterministic model (formally we have then also to define $Q := 0$). As adjoint functions we take for $t \geq 0$ and fixed $x \in E$

$$p_t^0(x) := \int_t^{T(x)} e^{(\beta+q_z)(t-s)} c(\phi_s(x), z)\, ds,$$

where the function $c : E \to \{b_1, \dots, b_N\}$ is defined by $c(x) := b_{k\wedge(d+1)}$, if $y_1 = \dots = y_{k-1} = 0,\ y_k > 0$ and z is d-stable, $k = 1, \dots, N+1$. Throughout the proof we will understand that $\int_{t_1}^{t_2} \dots ds =: 0$ if $t_1 > t_2$. $p_t^0(x)$ is obviously piecewise continuously differentiable since $\phi_t(x)$ is continuous in t and

$$\dot{p}_t^0(x) = (\beta + q_z) p_t^0(x) - c(\phi_t(x), z).$$

Set $\eta_t^0 := c - c(\phi_t(x), z)$. Since $b_k \leq c$, we have $\eta_t^0 \geq 0$. If $c(\phi_t(x), z) = b_k$, then $[\phi_t(x)]_j = 0, j = 1, \dots k-1$ and $\eta_t^0 \phi_t(x) = 0$. The continuity of $\phi_t(x)$ in t implies that η_t is piecewise constant. Hence conditions (ii), (iii) and (iv) of Lemma 4.7 are fulfilled. To verify (i) we have to look at

$$Ap_t^0(x) = \int_t^{T(x)} e^{(\beta+q_z)(t-s)} Ac(\phi_s(x), z)\, ds.$$

Now $u^* \in U$ with $u_1^*, \dots, u_k^* > 0,\ u_{k+1}^* = \dots = u_N^* = 0,\ \sum_j u_j^* = 1$ maximizes $u \to uw,\ u \in U$ if and only if $0 < w_1 = \dots = w_k \geq w_{k-1}, \dots, w_N$. Therefore, we have to compute $Ab_k,\ k = 1, \dots, N$. $Ab_1 = r$ by definition and

$$Ab_{k+1} = \begin{pmatrix} I^{k+1}1\!\!1_k \\ a_{k+1}^{S_k} A_{S_k}^{-1} r_{S_k} - I^{k+1} a_{k+1}^{S_k} A_{S_k}^{-1} 1\!\!1 + r_{k+1} \\ \vdots \\ a_N^{S_k} A_{S_k}^{-1} r_{S_k} - I^{k+1} a_N^{S_k} A_{S_k}^{-1} 1\!\!1 + r_N \end{pmatrix}$$

Because of the definition of I^{k+1}, it holds that

$$e_{k+1} Ab_{k+1} - I^{k+1} = \left(r_{k+1} + a_{k+1}^{S_k} A_{S_k}^{-1} r_{S_k}\right) - I^{k+1}\left(1 + a_{k+1}^{S_k} A_{S_k}^{-1} 1\!\!1\right) = 0$$

and since $a_j^{S_k} \geq 0,\ A_{S_k}^{-1} \geq 0$, using the maximality of I^{k+1} we have for $j = k+2, \dots, N$:

$$e_j Ab_{k+1} - I^{k+1} = \left(r_j + a_j^{S_k} A_{S_k}^{-1} r_{S_k}\right) - I^{k+1}\left(1 + a_j^{S_k} A_{S_k}^{-1} 1\!\!1\right) \leq 0$$

and from Lemma 6.6 a) we get $I^{k+1} \geq 0$. Hence we have shown that

$$0 \leq e_1 Ab_1 \geq e_2 Ab_1, \dots, e_N Ab_1$$

$$0 \leq e_1 Ab_k = \dots = e_k Ab_k \geq e_{k+1} Ab_k, \dots, e_N Ab_k,$$

$k \in \{2, \dots, N\}$. Since $c(\phi_s(x), z)$ is decreasing in s, it can be readily verified that (i) holds.
It is easy to see that p_t^0 is bounded and all admissible trajectories y_t can at most grow linearly. Therefore, the limit in (v) is always 0 which implies that (v) is also valid. Thus, we have shown that (i)-(v) of Lemma 4.7 are satisfied and the index rule is a minimizer of V_0. Suppose now that this is valid for $k = 0, \dots, n-1$ and denote the corresponding adjoint functions by $p_t^k(x)$. We will prove that it is also true for n. In this case we have to define further functions $q^k : E \times \{b_1, \dots, b_N\} \to I\!R^N,\ k \in I\!N_0$ which help represent the adjoint functions.

$$q^0(x, b_k) := \int_0^{T(x)} e^{-(\beta+q_z)s} \left[b_k \wedge c(\phi_s(x), z)\right]\ ds$$

and by recursion

$$p_t^n(x) := p_t^0(x) + \int_t^{T(x)} e^{(\beta+q_z)(t-s)} \sum_{z' \neq z} q_{zz'} q^{n-1}\left(\phi_s(x), z', c(\phi_s(x), z)\right)\ ds$$

$$\begin{aligned} q^n(x, b_k) &:= q^0(x, b_k) \\ &\quad + \int_0^{T(x)} e^{-(\beta+q_z)s} \sum_{z' \neq z} q_{zz'} q^{n-1}\left(\phi_s(x), z', c(\phi_s(x), z) \wedge b_k\right)\ ds. \end{aligned}$$

We will now show that the adjoint function p_t^n together with a suitably defined Lagrange function η_t^n satisfy the sufficient optimality conditions (i)-(v) in Lemma

4.7 with $l(y) = cy + \sum_{z' \neq z} q_{zz'} V_n(y, z')$.

(i) Define $w^0 := A \cdot q^0(x, b_k)$. Since $[b_k \wedge c(\phi_s(x), z)] \le b_k$ we obtain with the properties of AB: $0 \le w_1^0 = \ldots = w_k^0 \ge w_{k-1}^0 \ge \ldots \ge w_N^0$. Suppose this is true for $A \cdot q^j(x, b_k),\ j = 0, \ldots, n-1$. Define

$$\begin{aligned} w^n &:= A \cdot q^n(x, b_k) = A \cdot q^0(x, b_k) \\ &\quad + \int_0^{T(x)} e^{-(\beta + q_z)s} \sum_{z' \neq z} q_{zz'} A \cdot q^{n-1}\left(\phi_s(x), z', c(\phi_s(x), z) \wedge b_k\right)\, ds. \end{aligned}$$

Since $q_z, q_{zz'} \ge 0$ for $z' \neq z$ we again obtain $0 \le w_1^n = \ldots = w_k^n \ge w_{k-1}^n \ge \ldots \ge w_N^n$ by the induction hypothesis. Since $c(\phi_s(x), z)$ is decreasing in s we can conclude that Ap_t^n satisfies (i).

(ii) $p_t^n(x)$ is obviously continuous and piecewise continuously differentiable with

$$\dot{p}_t^n(x) = (\beta + q_z) p_t^n(x) - c(\phi_t(x), z) - \sum_{z' \neq z} q_{zz'} q^{n-1}\left(\phi_t(x), z', c(\phi_t(x), z)\right).$$

Since by Lemma 6.8 V_n is continuously differentiable, we have that $\frac{\partial}{\partial y} V_n(\phi_t(x), z') = p_0^{n-1}(\phi_t(x), z')$. Therefore,

$$\eta_t^n(x) := c - c(\phi_t(x), z) + \sum_{z' \neq z} q_{zz'} \left\{ p_0^{n-1}(\phi_t(x), z') - q^{n-1}\left(\phi_t(x), z', c(\phi_t(x), z)\right) \right\}$$

and η_t is piecewise continuous.

(iii) Due to the definition of the q^j we have for all $j \in I\!N_0$, $k = 1, \ldots, N$

$$p_0^j(x) \ge q^j(x, b_k).$$

Since $c \ge c(\phi_t(x), z)$ for all $t \ge 0$, we obtain $\eta_t^n(x) \ge 0$ for all $t \ge 0$.

(iv) Define $w := p_0^0(x) - q^0(x, b_k)$. The definition of b_k implies that $w_k = \ldots = w_N = 0$. Moreover, if we define $w^0 := q^0(x, b_j) - q^0(x, b_k)$ with $b_k \le b_j$ then again $w_k^0 = \ldots = w_N^0 = 0$. Suppose this is valid for $w^0, \ldots, w^{n-1}$. Define

$$\begin{aligned} w^n &:= q^n(x, b_j) - q^n(x, b_k) = q^0(x, b_j) - q^0(x, b_k) \\ &\quad + \int_0^{T(x)} e^{-(\beta + q_z)s} \sum_{z' \neq z} q_{zz'} \left\{ q^{n-1}\left(\phi_s(x), z', c(\phi_s(x), z) \wedge b_j\right) \right. \\ &\quad \left. - q^{n-1}\left(\phi_s(x), z', c(\phi_s(x), z) \wedge b_k\right) \right\}\, ds. \end{aligned}$$

Using the induction hypothesis we can conclude that $w_k^n = \ldots = w_N^n = 0$. Thus the vector $p_0^n(x) - q^n(x, b_k)$ has the same property. Now if $c(\phi_t(x), z) = b_k$ we know that $[\phi_t(x)]_j = 0,\ j = 1, \ldots, k-1$. Therefore $\eta_t^n(x)\phi_t(x) = 0$ for all $t \ge 0$.

(v) It is easy to see that all $p_t^n(x)$ are bounded. In particular

$$|p_t^n(x)| \leq \frac{1}{\beta + q_z} \sum_{k=0}^{N} \left(\frac{q_z}{q_z + \beta} \right)^k \|c\|$$

for all $x \in E$ and $t \geq 0$. Hence the same argument as for $n = 0$ applies and (v) is valid. Altogether we have now proven the theorem. $\square$

Remark 6.4:
In fact it can be shown that the index policy is optimal in a very strong sense: it minimizes the cost on each sample path, i.e. if π^* is the index policy and π an arbitrary policy, it holds for all $\omega \in \Omega$ that

$$\int_0^\infty e^{-\beta t} c Y_t^{\pi^*}(\omega)\, dt \leq \int_0^\infty e^{-\beta t} c Y_t^{\pi}(\omega)\, dt.$$

Hence we have in particular that the cost under the index policy are stochastically smaller than the cost under any arbitrary policy. For the proof it is possible to take almost the same construction for the adjoint functions p_t^n. Moreover, it is important to note that the indices do neither depend on the inflow rate λ nor on the interest rate β.

Now we will investigate the single-server problem under the c-average cost criterion. We have to find a further stability assumption which guarantees the validity of Assumption 3.1. For this purpose, let us define by $\alpha_j(z),\ j = 1, \ldots, N$ the nominal total arrival rate to buffer j in environment state z, which can be computed from the traffic equation $\alpha(z) = \lambda(z) + \alpha(z)P^T$. Hence, $\alpha(z) = (I - P^T)^{-1}\lambda(z)$. We will denote by $(\nu_z)_{z \in Z}$ the stationary distribution of (Z_t) which exists since (Z_t) is irreducible. It is not surprising that the priority index policy is again optimal. The stability condition is simply that the mean traffic intensity of the buffer is less than 1.
The following general Lemma will be useful (see Lemma 7.1 in Sethi (1997)):

Lemma 6.9:
Assume that $\sum_{z \in Z} z\, \nu_z < \mu$ with $\mu \in I\!R_+$. For $l_0 > 0$ let $\tau = \inf\{t \geq 0 \mid \int_0^t (\mu - Z_s)\, ds \geq l_0\}$. Then for any $k \in I\!N$, there exists a constant C_0 independent of l_0 such that $E[\tau^k] \leq C_0(1 + l_0^k)$.

Theorem 6.10: *(Optimality of the priority index policy - average case)*
Suppose that Assumption 6.5 is valid and that

$$\sum_{z\in Z}\sum_{j=1}^{N}\frac{\alpha_j(z)}{\mu_j(z)}\nu_z < 1.$$

Then the priority index policy is c-average optimal and $G(x) = g < \infty$ for $x \in E$.

Proof: We prove the result with the help of Theorem 3.6. Denote by $\mathcal{Z} \subset Z$, the set of states which are N-stable ($\mathcal{Z} \neq \emptyset$ due to our stability assumption). W.l.o.g. assume $0 \in \mathcal{Z}$. We will verify Assumption 3.1 with $\xi = (0, 0)$. Since $y \mapsto V^\beta(y, z)$ is increasing by Lemma 6.8, the relative value function h^β is bounded below by $\min_z V^\beta(0, z) - V^\beta(\xi)$. Thus it suffices to verify the assumptions of Lemma 3.8. We will do this with $\xi = (0, 0)$ and f taken as the priority index rule. Now denote by Y_t the state process of the buffer content under the index policy $\pi = f^\infty$ and define $\Lambda_t := \int_0^t \lambda(Z_s)\, ds$. Moreover, the process $T_t = (T_1(t), \dots, T_N(t))$ is referred to as the allocation process under policy π, i.e. $T_j(t) = \int_0^t \pi_j(s)\, ds$. Thus we have for almost all ω (cf. Chen (1995)):

$$\begin{aligned}
&Y_t = Y_0 + \Lambda_t - [I - P^T]DT_t = Y_0 + \Lambda_t - A^T T_t \geq 0\\
&t \mapsto T_t \quad \text{is increasing, } T_0 := 0.\\
&t \mapsto [\mathbb{1}t - T_t] \quad \text{is increasing.}\\
&[\mathbb{1}Y_t]\, d\,[t - \mathbb{1}T_t] = 0
\end{aligned}$$

In particular we have with $Q_t := Y_t A^{-1}\mathbb{1}$ and $\tilde{\Lambda}_t := \Lambda_t A^{-1}\mathbb{1}$ for almost all ω

$$\begin{aligned}
&Q_t = Q_0 + \tilde{\Lambda}_t - \mathbb{1}T_t = Q_0 + \tilde{\Lambda}_t - t + (t - \mathbb{1}T_t) \geq 0\\
&t \mapsto \mathbb{1}T_t \quad \text{is increasing, } T_0 := 0.\\
&t \mapsto [t - \mathbb{1}T_t] \quad \text{is increasing.}\\
&Q_t\, d\,[t - \mathbb{1}T_t] = 0
\end{aligned}$$

Hence Q_t can be interpreted as the buffer content of a single server system under a work-conserving policy (see last condition) and we know that if $\tilde{\Lambda}_{t_0} + Q_0 \leq t_0$ for a $t_0 > 0$, then there exists a $t \in [0, t_0]$ such that $Q_t = 0$ and thus $Y_t = 0$. However, from our assumption (note that the stability condition can be written as $\sum_z \lambda(z)A^{-1}\mathbb{1}\nu_z < 1$) and Lemma 6.9 we have that $\tilde{\tau} := \inf\{t \geq 0 \mid t \geq Q_0 + \tilde{\Lambda}_t\}$ has the property that $E_x^f[\tilde{\tau}^k] \leq C_0(1 + \|y\|^k)$ for all $k \in \mathbb{N}$ and the constant C_0 is independent of y. Therefore, with $\tau_1 = \inf\{t \geq 0 \mid Y_t = 0\}$ we obtain $\tau_1 \leq \tilde{\tau}$ and hence

$$E_x^f[\tau_1^k] \leq C_0(1 + \|y\|^k),\ k = 1, 2.$$

However, we do not know Z_{τ_1} and are interested in $\tau_{(0,0)} := \inf\{t \geq 0 \mid X_t = (0, 0)\}$. Obviously $Z_{\tau_1} \in \mathcal{Z}$. If $\mathcal{Z} = Z$, the statement follows easily. Now suppose $\mathcal{Z} \neq Z$.

We define now the following sequence of stopping times:

$$\begin{aligned}\sigma_1 &:= \inf\{t \geq \tau_1 \mid Z_t \notin \mathcal{Z}\}\\ \tau_n &:= \inf\{t > \sigma_n \mid Y_t = 0\}\\ \sigma_n &:= \inf\{t \geq \tau_n \mid Z_t \notin \mathcal{Z}\}.\end{aligned}$$

Then $0 \leq \tau_1 \leq \sigma_1 \leq \ldots < \tau_n \leq \sigma_n$ and obviously $Y_t = 0$ for $t \in [\tau_n, \sigma_n),\ n \in I\!N$. Moreover, from the theory of Markov chains we get, since (Z_t) is ergodic

(i) $P_x^f(Z_s \neq 0,\ \tau_n \leq s < \sigma_n) \leq \alpha < 1$, for all $n \in I\!N$.
(ii) $E_x^f[(\sigma_n - \tau_n)^k] \leq C_1$, for all $k = 1, 2,\ n \in I\!N,\ x \in E$.

From (i) we conclude that

$$\begin{aligned}P_x^f(\tau_{(0,0)} > \sigma_n) &= P_x^f(\tau_{(0,0)} > \sigma_n, \ldots, \tau_{(0,0)} > \sigma_1) = P_x^f(\tau_{(0,0)} > \sigma_n \mid \tau_{(0,0)} > \sigma_{n-1})\\ &\quad \cdot \ldots \cdot P_x^f(\tau_{(0,0)} > \sigma_2 \mid \tau_{(0,0)} > \sigma_1) P_x^f(\tau_{(0,0)} > \sigma_1) \leq \alpha^n.\end{aligned}$$

And we know that $\max_z E_{(0,z)}^f[\tau_1^k] \leq C_0,\ k = 1, 2$. Hence together with (ii) we obtain

$$E_x^f[(\tau_{n+1} - \tau_n)^k] \leq C_2, \text{ for all } k = 1, 2,\ n \in I\!N$$

where C_2 is independent of x. Thus we have for $k = 1, 2$

$$\begin{aligned}E_x^f[\tau_{(0,0)}^k] &= k\int_0^\infty t^{k-1} P_x^f(\tau_{(0,0)} > t)\, dt = k\sum_{n=1}^\infty E_x^f\left[\int_{\tau_{n-1}}^{\tau_n} t^{k-1} P_x^f(\tau_{(0,0)} > t)\, dt\right] \leq\\ &\leq E_x^f[\tau_1^k] + k\sum_{n=2}^\infty E_x^f\left[\int_{\tau_{n-1}}^{\tau_n} t^{k-1} P_x^f(\tau_{(0,0)} > \sigma_{n-2})\, dt\right] \leq\\ &\leq C_0(1 + \|y\|^k) + \sum_{n=2}^\infty \alpha^{n-2} E_x^f\left[(\tau_n - \tau_{n-1})^k\right] \leq\\ &\leq C_0(1 + \|y\|^k) + C_2\frac{1}{1-\alpha} \leq C_3(1 + \|y\|^k)\end{aligned}$$

Since Y_t can grow at most linear, there exists a constant $\tilde{c} \in I\!R_+$ such that

$$\begin{aligned}E_x^f\left[\int_0^{\tau_{(0,0)}} cY_t\, dt\right] &\leq E_x^f\left[\int_0^{\tau_{(0,0)}} cy + \tilde{c}t\, dt\right]\\ &= cyE_x^f[\tau_{(0,0)}] + \frac{1}{2}\tilde{c}E_x^f[\tau_{(0,0)}^2] \leq C_4(1 + \|y\|^2) < \infty\end{aligned}$$

with a constant C_4 independent of y. In the same way we obtain with $\sigma_{(0,0)} = \inf\{T_n \mid X_{T_n} = (0,0)\ \}$ that

$$E_x^f\left[\int_0^{\sigma_{(0,0)}} cY_t\, dt\right] < \infty.$$

Thus, all assumptions of Theorem 3.6 are fulfilled. In particular $G_f(x) = J_f(x) < \infty$ for all $x \in E$. Now take a sequence $x_m(x) \in E, x_m(x) \to x \in E$ and $\beta_m(x) \downarrow 0$.

From the β-discounted case we know that the optimal policy $f^{\beta_m(x)}(x_m(x))$ is the priority index policy which is independent of $\beta_m(x)$. Using Lemma A.4 and the special structure of the index policy it is easily seen that $\lim_{m\to\infty} f^{\beta_m(x)}(x_m(x)) = f(x)$. Hence the result follows. $\square$

Remark 6.5:

(i) According to Theorem 3.7 the priority index policy is also d-average optimal.

(ii) From the proof of Theorem 6.10 we see that the state process (X_t) is positive Harris recurrent under the stability assumption, for all policies which are work-conserving, i.e. fulfill for almost all ω : $[\mathbb{1}Y_t]\, d\,[t - \mathbb{1}T_t] = 0$ for $t \geq 0$.

6.3 Routing to Parallel Queues

Another classical problem in queueing theory is the problem of routing to parallel queues. For a literature survey see e.g. Stidham/Weber (1993). In the stochastic setting there are only a few analytical results on the problem. It is known that in the symmetric case, i.e. if all queues are equal, the Join-the-shortest queue policy is optimal. This has been shown in Winston (1977). In the case of two buffers, the optimal policy is given by a switching curve (cf. Stidham/Weber (1993)) which however, cannot be given explicitely. In terms of SFPs we have one randomly varying input process $\lambda(Z_t)$ which has to be split up and routed to N different buffers. Each buffer has a single server with a fixed potential output rate μ_j, $j = 1, \ldots, N$. As before let D be the diagonal matrix with elements μ_j on the diagonal. We assume linear holding cost of rate $c_j \in \mathbb{R}_+$ when holding fluid in buffer j. Denote $c = (c_1, \ldots, c_N)$. In order to define the problem correctly one has to decide upon the splitting $u = (u_1, \ldots, u_N)$ of the arriving fluid and on the activation $v = (v_1, \ldots, v_N)$ of the servers, hence $U = \{(u, v) \in [0, 1]^{2N} \mid \sum_{j=1}^N u_j = 1\}$. Summing up our SFP has the following form

$$
\begin{aligned}
E &= \mathbb{R}_+^N \times Z \\
U &= \{(u, v) \in [0, 1]^{2N} \mid \sum_{j=1}^{N} u_j = 1\} \\
b^z(u, v) &= \lambda(z)u - Dv \\
c(x, u) &= cy
\end{aligned}
$$

We will look at two special cases of this model.

A) Equal holding cost

Let us assume that $c_1 = \ldots = c_N = c \in \mathbb{R}_+$. Note that in this subsection c is in $\mathbb{R}_+$ and not a vector. We will show that the Least Loaded Routing (LLR) policy is optimal which we define in the following way. For $y \in \mathbb{R}_+^N$ denote by

$$I(y) = \{1 \leq j \leq N \mid \frac{y_j}{\mu_j} = \min_{1 \leq i \leq N} \frac{y_i}{\mu_i}\}$$

the buffer numbers with least load, if y gives the current buffer contents. The allocation vectors $u^*(y)$ and $v^*(y, z)$ are defined by

$$u_j^*(y) = \begin{cases} \frac{\mu_j}{\sum_{i \in I(y)} \mu_i} & , \text{ if } j \in I(y) \\ 0 & , \text{ if } j \notin I(y). \end{cases}$$

$$v_j^*(y, z) = \begin{cases} 1 & , \text{ if } y_j > 0 \\ \min\{1, \frac{\lambda(z) u_j^*(y)}{\mu_j}\} & , \text{ if } y_j = 0 \end{cases}$$

for $j = 1, \ldots, N$. The LLR policy $\pi = f^\infty$ is a stationary feedback policy with

$$f(x)(t) = (u^*(y_t), v^*(y_t, z))$$

where $y_t := \phi_t(x, f(x))$.

Theorem 6.11: *(Optimality of the LLR policy - β-discounted case)*
The LLR policy is optimal for the discounted cost model.

Example:

Suppose we have the following deterministic network:

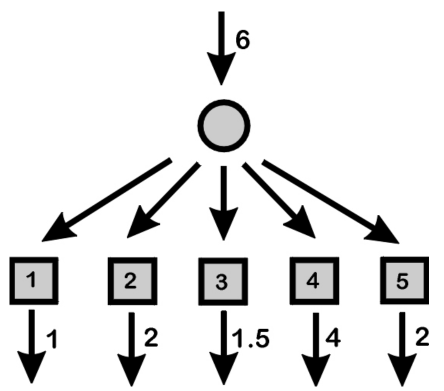

fig.6.4 : Routing network

The number next to the buffer gives the server rate μ_j. We have choosen $\lambda = 6$. Figure 6.5 gives the path of the optimal trajectory with $y = (1, 1, 1, 1, 1)$ at $t = 0$.

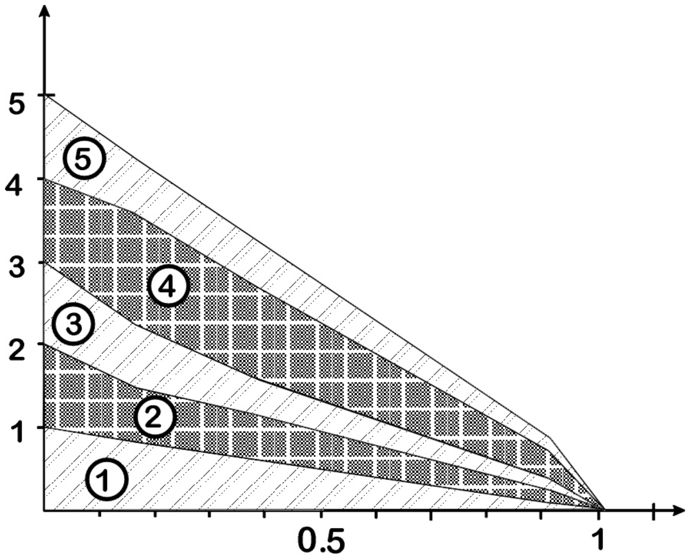

fig.6.5 : Optimal trajectory

The method of the proof is the same as in Section 6.2. Let $l : I\!R_+^N \to I\!R_+$ be arbitrary. We will look at the following problem

$$(CP) \begin{cases} \int\limits_0^\infty e^{-(\beta+q_z)t} l(\phi_t(x,a))\, dt \;\to\; \min \\ \phi_t(x,a) = y + \int\limits_0^t (\lambda(z)a_1(s) - D\, a_2(s))\; ds \\ \phi_t(x,a) \in I\!R_+^N \\ a_t \in U \text{ for all } t \geq 0 \end{cases}$$

In the following, $z \in Z$ is fixed and we will suppress it in our notation. $V(y)$ is the value function of (CP).

Lemma 6.12:
Suppose that the LLR control is optimal for problem (CP) and l fulfills Assumption 6.6. Then

a) $y \mapsto V(y)$ is increasing and convex.
b) There exists a constant $C'_0 \in I\!R_+$ such that for all $y, y' \in I\!R_+^N$

$$|V(y) - V(y')| \leq C'_0\,(1 + \|y\| + \|y'\|)\, \|y - y'\|.$$

c) $y \mapsto V(y)$ is continuously differentiable.

Proof: a) The convexity follows immediately from Lemma 2.6. For the monotonicity let $y \in I\!R_+^N, h > 0$. It suffices to show that $V(y+he_j) \geq V(y)$ for $j = 1, \ldots, N$. Denote by $a_h(t) = (a_1^h(t), a_2^h(t))$ the optimal LLR control, starting in $y + he_j$. Define the control $a_1(t) = a_1^h(t)$ and $a_2(t) = v^*(\phi_t(y,a), z)$. Obviously $a \in D(y)$ and $\phi_t(y,a) \leq \phi_t(y+he_j, a_h)$ for all $t \geq 0$. Hence we have due to the monotonicity of l that $V(y) \leq LV(y,a) \leq LV(y+h, a_h) = V(y+h)$.

b) Follows as in Lemma 6.8 b).

c) Recalling the proof of Lemma 6.8 c) it remains to show that $\phi_t(y)$ has almost everywhere continuous partial derivatives w.r.t. y. Fix two buffers k and j, $1 \leq k, j \leq N$. Define $t_0 := \inf\{t > 0 \mid [\phi_t(y)]_k = 0\}$ and $t_1 = \inf\{t \geq 0 \mid j \in I(\phi_t(y))\}$. Denote by a and a^h the LLR control, starting in y and $y + he_j$ respectively. Then we have for $t > t_0$:

$$\lim_{h \to 0} \frac{[\phi_t(y+he_j) - \phi_t(y)]_k}{h} = 0$$

and for $t < t_0$:

$$\lim_{h \to 0} \frac{[\phi_t(y+he_j) - \phi_t(y)]_k}{h} = \begin{cases} \delta_{jk} & , \text{ if } j \notin I(\phi_{t-}(y)) \\ \tilde{c} & , \text{ if } j \in I(\phi_{t-}(y)) \end{cases}$$

with a constant $\tilde{c}$. Thus, the assertion follows as in Lemma 6.8c). □

Proof of Theorem 6.11 We will use policy iteration and show that the LLR control is optimal for (CP) where we define $l(y) := cy + \sum_{z' \neq z} q_{zz'} V_n(y, z')$, $n \in I\!N$. We do this by induction on n using Lemma 4.7, where (i) reads

$$(u_t^*, v_t^*) \text{ minimizes } (u_t, v_t) \mapsto \lambda(z) p_t u_t - p_t D\, v_t \text{ for } (u_t, v_t) \in U.$$

Let $x \in E$ and denote by $T_j(x) \in \bar{I\!R}_+$ the time it takes to empty buffer $j = 1, \ldots, N$ under the LLR control. Note that either $T_j(x) < \infty$ for all j or $T_j(x) = \infty$ for all j. Moreover, let $\phi_t(x) := \phi_t(x,a)$, where a is the LLR control. Now consider the optimization problem for $n = 0$. Since $V_0 = 0$ we have to solve the control problem for the purely deterministic model (formally we have then also to define $Q := 0$). As adjoint functions we take for $t \geq 0$ and $j = 1, \ldots, N$

$$p_j^0(x,t) := c \int_t^{T_j(x)} e^{(\beta + q_z)(t-s)}\, ds,$$

and define the Lagrange multipliers for $t \geq 0$ and $j = 1, \ldots, N$ as

$$\eta_j^0(x,t) := c 1_{[t > T_j(x)]}.$$

As before we assume that $\int_{t_1}^{t_2} \ldots ds =: 0$ if $t_1 > t_2$. Obviously (ii)-(v) of Lemma 4.7 are satisfied (cf. proof of Theorem 6.7). For (i) note that $T_j(x) \geq T_i(x)$ implies $p_j^0 \geq p_i^0$ and $p_j^0(x,t) = 0$ if $[\phi_t(x)]_j = 0$, hence the LLR control solves the minimization problem in (i). Hence we have shown that (i)-(v) of Lemma 4.7 are satisfied and the LLR control is a minimizer of V_0. Suppose now that this is valid for $k = 0, \dots, n-1$ and denote the corresponding adjoint functions by $p_t^k(x)$. We will prove that it is also true for n. Define for $t \geq 0$ and $j = 1, \dots, N$

$$p_j^n(x,t) := \int_t^{T_j(x)} e^{(\beta+q_z)(t-s)} \left(c + \sum_{z' \neq z} q_{zz'} p_j^{n-1}(\phi_s(x), z', 0) \right) ds$$

$$\eta_j^n(x,t) := 1_{[t > T_j(x)]} \left(c + \sum_{z' \neq z} q_{zz'} p_j^{n-1}(\phi_s(x), z', 0) \right)$$

Properties (ii)-(v) of Lemma 4.7 follow directly (cf. proof of Theorem 6.7). Since under the LLR control $t_1 \leq t_2$ implies $I(\phi_{t_1}(x)) \subset I(\phi_{t_2}(x))$ we have by the induction hypothesis for $j \in I(\phi_s(x))$ and all $z' \in Z$ (note that the order of depletion times is independent of z'):

$$\begin{aligned} 0 \leq p_j^{n-1}(\phi_s(x), z', 0) &= p_i^{n-1}(\phi_s(x), z', 0), \text{ if } i \in I(\phi_s(x)) \\ 0 \leq p_j^{n-1}(\phi_s(x), z', 0) &\leq p_i^{n-1}(\phi_s(x), z', 0), \text{ if } i \notin I(\phi_s(x)) \end{aligned}$$

This implies (i), because $q_{zz'} \geq 0$ for $z \neq z'$. Altogether we have proved Theorem 6.11. □

Let us now look at the average cost problem. The analysis is in the same spirit as in Theorem 6.10. As before, denote by $(\nu_z)_{z \in Z}$ the stationary distribution of (Z_t).

Theorem 6.13: *(Optimality of the LLR policy - average cost case)*
Suppose the following stability assumption is valid

$$\sum_{z \in Z} \lambda(z) \nu_z < \sum_{j=1}^N \mu_j.$$

Then the LLR policy is c-average optimal and $G(x) = g < \infty$ for $x \in E$.

Proof: We prove the result with the help of Theorem 3.6. Denote by $\mathcal{Z} \subset Z$, the set of states with $\lambda(z) < \sum_j \mu_j$ ($\mathcal{Z} \neq \emptyset$ due to our stability assumption). W.l.o.g. assume $0 \in \mathcal{Z}$. We will verify Assumption 3.1 with $\xi = (0,0)$. Since $y \mapsto V^\beta(y,z)$ is increasing by Lemma 6.12, the relative value function h^β is bounded below by $\min_z V^\beta(0,z) - V^\beta(\xi)$. Thus it suffices to verify the assumptions of Lemma 3.8. We will do this with $\xi = (0,0)$ and $\hat{f}$, which will be defined in the following way: let

$$u := \left(\frac{\mu_1}{\sum_{j=1}^N \mu_j}, \dots, \frac{\mu_N}{\sum_{j=1}^N \mu_j}\right)$$

and $\hat{f}(x)(t) := (u, v^*(\hat{\phi}_t(x), z))$, where $\hat{\phi}_t(x) = \phi_t(x, \hat{f}(x))$. Denote by Y_t the state process of the buffer content under policy $\hat{\pi} = \hat{f}^\infty$, with $Y_0 = y$. W.l.o.g. suppose $\frac{y_1}{\mu_1} = \max_{1 \le j \le N}\{\frac{y_j}{\mu_j}\} > 0$ and

$$\tilde{\tau} := \inf\{t \ge 0 \mid \int_0^t (1 - \frac{\lambda(Z_s)}{\sum_{j=1}^N \mu_j})\, ds \ge \frac{y_1}{\mu_1}\}.$$

Since

$$Y_1(t) = y_1 + \int_0^t \left(\lambda(Z_s)\frac{\mu_1}{\sum_{j=1}^N \mu_j}\right) ds - \mu_1 t$$

for $0 \le t \le \tilde{\tau}$ a.s., it holds that that there exists a $t_0 \le \tilde{\tau}$ with $Y_{t_0} = 0$ a.s. and we can proceed as in the proof of Theorem 6.10 to obtain

$$E_x^f \left[\int_0^{\tau(0,0)} cY_t\, dt\right] \le C_1(1 + \|y\|^2) < \infty$$

with a constant C_1 independent of y. Using the same arguments as in Theorem 6.10 we obtain $G_{\hat{f}} < \infty$. Now take a sequence $x_m(x) \in E, x_m(x) \to x \in E$ and $\beta_m(x) \downarrow 0$. From the β-discounted case we know that the optimal policy $f^{\beta_m(x)}(x_m(x))$ is the LLR policy which is independent of $\beta_m(x)$. Using Lemma A.4 and the special structure of the LLR policy it is easily seen that $\lim_{m\to\infty} f^{\beta_m(x)}(x_m(x)) = f(x)$. Since $y \to \phi_t(y)$ is continuous, f induces a weakly continuous kernel $p_t(x, f; \cdot)$ and due to Lemma 3.10 we have $G_f = J_f$. Applying Theorem 3.6 we obtain our result. □

B) Deterministic Two-Buffer Case

In this section the linear holding costs are not necessarily equal at the buffers. However, we will restrict to two buffers and deterministic input, i.e. $\lambda(z) \equiv \lambda$ for all $z \in Z$. The optimal policy in this case is given by a feedback switching control.

Theorem 6.14:

Let $N = 2$. Suppose that $\mu_1, \mu_2 > \lambda$ and w.l.o.g. $c_1 \le c_2$. Then the following feedback switching control $g : \mathbb{R}_+^2 \to U$ is optimal

$$g(y_1, y_2) = \begin{cases} ((1,0), v^*(y)), & \text{if } y_2 > S(y_1) \\ ((0,1), v^*(y)), & \text{if } y_2 \le S(y_1) \end{cases}$$

where $S(y_1) := -\frac{\mu_2 - \lambda}{\beta} log\left(1 - \frac{c_1}{c_2}(1 - e^{-\beta\frac{y_1}{\mu_1}})\right)$.

Proof: Due to the assumption $\mu_1, \mu_2 > \lambda$, g is obviously optimal if $y_1 = 0$ or $y_2 = 0$. Therefore, suppose $y > 0$. As long as $y_t > 0$ the adjoint functions in the maximum principle (Theorem 4.5) are of the form

$$p_j(t) = \frac{1}{\beta}(c_j + C_j e^{\beta t}),\ j = 1, 2$$

with constants $C_1, C_2 < 0$. From part (i) of the maximum principle we know that it is optimal in state y to route to 2 if $p_1(0) \geq p_2(0)$ which is equivalent to $c_1 + C_1 \geq c_2 + C_2$. Since $c_1 \leq c_2$ this implies $C_2 < C_1$. It is easy to see that the adjoint functions p_1 and p_2 have at most one intersection and if $p_1(0) \geq p_2(0)$, then $p_1(t) \geq p_2(t)$ for all $t \geq 0$. Moreover, the adjoint functions hit zero if and only if the buffer empties, i.e. if we denote $t_j = \inf\{t \geq 0 \mid p_j(t) = 0\}$ $j = 1, 2$, then $y_1 - \mu_1 t_1 = 0,\ y_2 - (\mu_2 - \lambda)t_2 = 0$ which implies $C_1 = -c_1 e^{-\frac{\beta y_1}{\mu_1}},\ C_2 = -c_2 e^{-\frac{\beta y_2}{\mu_2 - \lambda}}$. The condition $c_1 + C_1 \geq c_2 + C_2$ now reads

$$c_1 \left(1 - e^{-\beta \frac{y_1}{\mu_1}}\right) \geq c_2 \left(1 - e^{-\beta \frac{y_2}{\mu_2 - \lambda}}\right)$$

which gives our switching curve. □

Remark 6.6:
If we let $\beta \downarrow 0$ the switching curve $S(y_1)$ converges to $\frac{\mu_2 - \lambda}{\mu_1} \frac{c_1}{c_2} y_1$, which is the switching curve in the undiscounted case (cf. Avram (1997)).

7 Asymptotic Optimality of Tracking-Policies

In this section we want to shed some light on the important role of deterministic fluid problems which we obtain as a special case of our SFP, by allowing only one environment state. In this case we obtain the following deterministic control problem

$$(F)\begin{cases} \int\limits_0^\infty e^{-\beta t} c(y_t, a_t)\, dt \to \min \\ y_t = y_0 + \int\limits_0^t b(a_s)\, ds \\ y_t \geq 0, \\ a_t \in U,\ t \geq 0 \end{cases}$$

which we will refer to as the *fluid problem*. These problems are not only interesting for themselves, but also serve as an approximation for stochastic queueing networks. For example, it has been shown recently that there is a close connection between the stability of a stochastic network and the corresponding fluid model (see e.g. Dai (1995), Bramson (1996), Maglaras (1998a)). Since in examples it often turned out that the optimal control in the fluid problem and the optimal policy in the stochastic network coincide (see Section 6), the question arises, whether there is also a connection between the control problem in the stochastic network and the fluid problem (cf. Avram et al. (1995), Avram (1997), Atkins/Chen (1995), Meyn (1997)). This is an important issue, since control problems in stochastic networks are difficult to solve. Although the dynamic programming technique which is the most common solution method, is well developed, only a few special networks allow for an explicit solution (cf. also Stidham/Weber (1993)). Due to the enormous state space of the problems, a numerical solution is often intractable. On the other hand, several authors have shown that the optimization problem in the fluid setting is often easy to solve: in Avram et al. (1995) one can find numerous scheduling problems which have been solved explicitely using Pontryagins maximum principle. In addition, the authors give an efficient approximation algorithm to solve the deterministc control problems which arise when the fluid model is considered (cf. Section 5.1). Weiss (1995, 1996, 1997) solved several scheduling problems in fluid re-entrant lines, showing that modifications of the 'Last-Buffer-First-Served'-policy are optimal.
In the literature we can find several results concerning the relation between the control problem for stochastic queueing networks and the corresponding fluid problem. Meyn (1997) for example proved in Theorem 5.2 that in the average cost model, the policy iteration if initialized with a stable policy for the fluid model, yields a sequence of relative value functions which converge when properly normalized against the value function in the fluid model. Chen/Meyn (1998) used this fact to suggest that the value iteration can be speed up when initialized with the value function of the fluid model. In Atkins/Chen (1995) one can find a large numerical study, where the optimal control from the fluid model has been used as a heuristic

for the policy in the stochastic network. The performance of this implementation has been compared to simple priority rules. It turned out that the fluid heuristic was often slightly better than the priority rules but not always. Alanyali/Hajek (1998) considered a special routing problem and proved that the Load-Balancing policy which is optimal in the associated fluid model is asymptotically optimal in the stochastic network. However, the crucial question in general is how to translate the optimal fluid control in an admissible policy for the stochastic network. A first proposal came from Maglaras (1998a,b, 1999) who used the BIGSTEP idea of Harrison (1996) to construct a class of policies which he called discrete-review policies. These policies are asymptotically optimal under fluid scaling in multi-class queueing networks, i.e. when the intensity of the process increases by factor γ and the jump height decreases by the same factor. The idea is to review the state of the system at discrete time points and compute from linear programs the actions that have to be carried out over the next planning period. The information about the fluid model is here put into the LP. Safety stock requirements ensure that the plans can be processed properly.

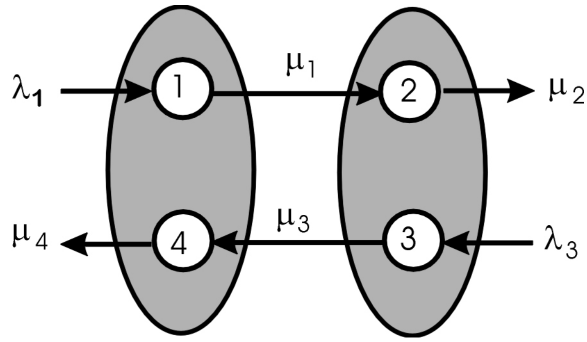

fig.7.1 : Rybko-Stolyar network

We will now propose a new class of policies which can be constructed from the optimal fluid control directly and are very intuitive. We will call these policies *Tracking-policies*. They are asymptotically optimal under fluid scaling in the same sense as in Maglaras (1998b) and work for a general class of network problems. The name Tracking-policy is chosen, since the scaled state process converges to the optimal fluid trajectory. In fact it is possible to use this approach to track every arbitrary chosen fluid trajectory. Hence this method works for a large class of objectives. The approach relies on the observation that in fluid problems the optimal control is usually piecewise constant (see Theorem 5.1). As a numerical example we have taken the Rybko-Stolyar network in figure 7.1 (cf. Rybko/Stolyar (1992), Maglaras (1998b)): queue 1 and 4 are processed by server 1, while queue 2 and 3 are processed by server 2. The service times of jobs are independent and exponentially distributed with rate $\mu_1 = \mu_3 = 3$ and $\mu_2 = \mu_4 = 1.5$. Queue 1 and 3 receive jobs from outside according to Poisson processes with rate $\lambda_1 = \lambda_3 = 1$.

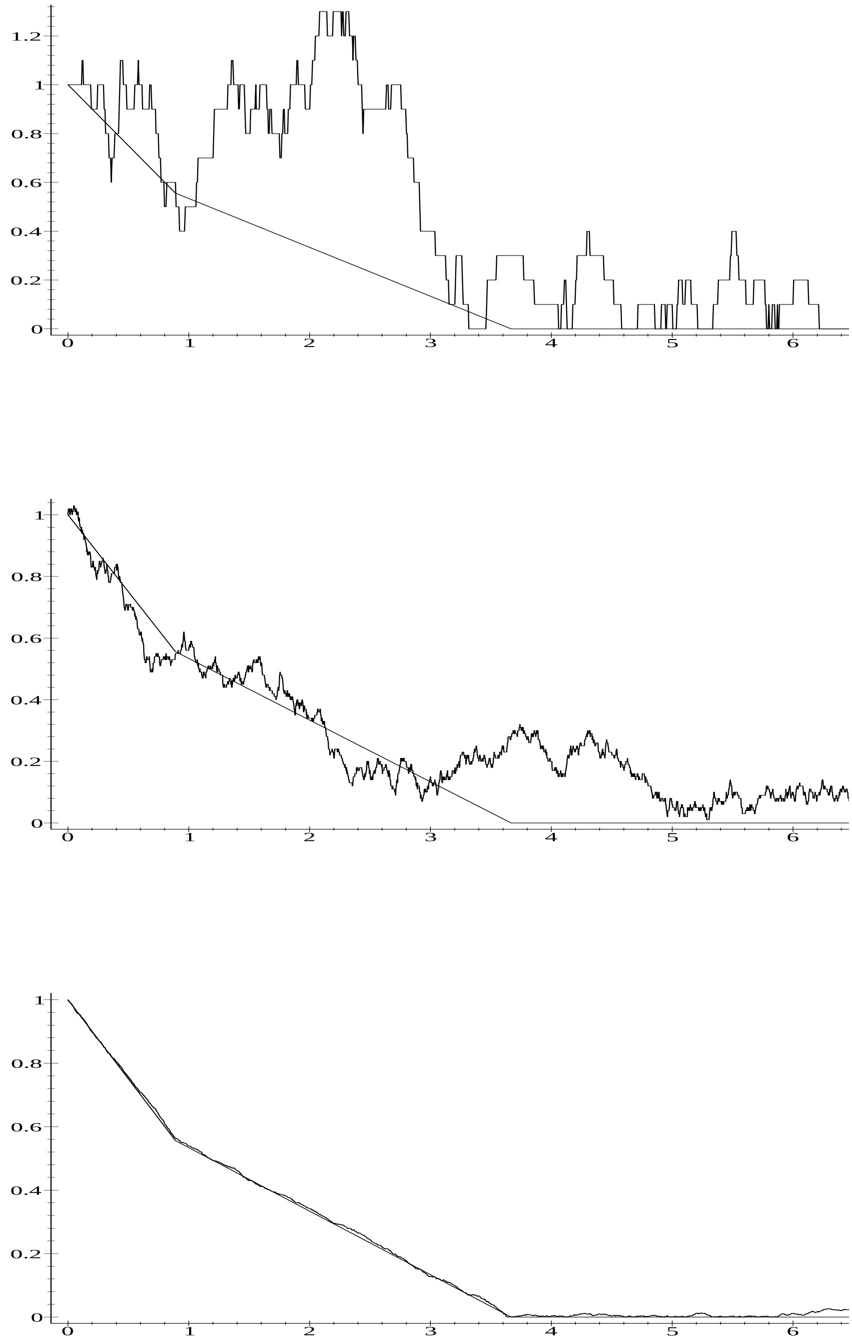

fig.7.2 : Trajectories for buffer 1 in the Rybko-Stolyar network with $\gamma = 10, 10^2, 10^4$

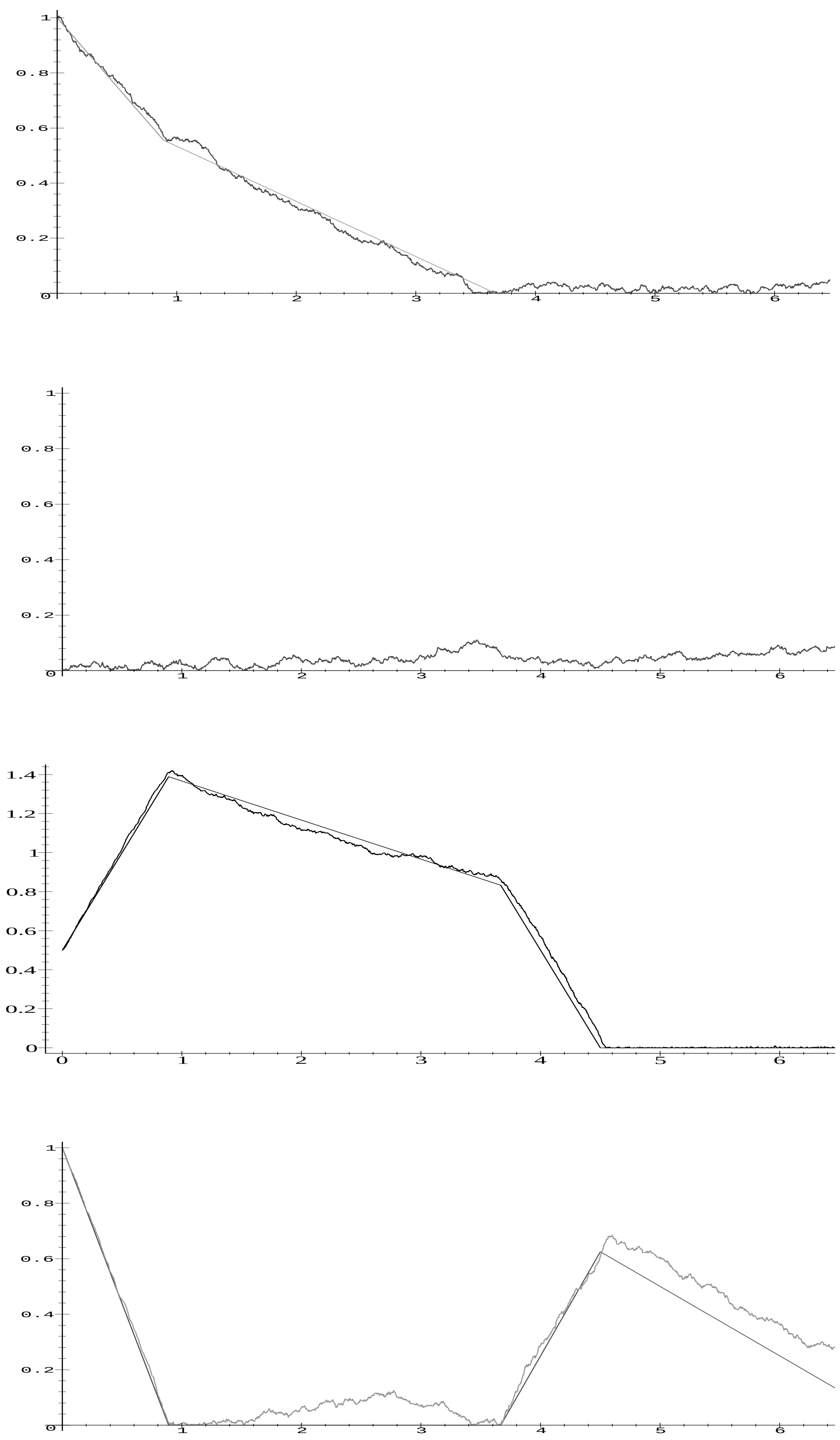

fig.7.3 : Trajectories for buffer 1-4 in the Rybko-Stolyar network with $\gamma = 10^3$

The initial state is $y_0 = (1, 0, 0.5, 1)$ and we assume linear holding cost $c_1 = \ldots = c_n = 1$ for the jobs. The optimization problem is to schedule the servers in order to minimize the expected discounted cost. In figure 7.2 we see simulation results of the trajectory of queue 1 under the Tracking-policy for scaling parameter $\gamma = 10, 100$ and 10000. The solid line is the optimal trajectory in the fluid model. In Section 7.3 we will prove that the trajectories and value functions of the stochastic network under the Tracking-policy converge against the optimal ones of the fluid model when γ tends to ∞. Figure 7.3 shows a simulation result for the trajectory of queue 1-4 respectively, with scaling parameter $\gamma = 1000$. The solid line is again the optimal trajectory.
In the next table we find the value function $V^{\gamma}_{\sigma^{\gamma}}$ for different scaling parameter γ under the Tracking-policy σ^{γ}. The optimal cost in the fluid problem are 7.22. From the simulation we can see that the Tracking-policy performs well, when we are close to the limit, i.e. in queueing systems with a large initial state and high intensity.

γ	10^4	10^5	10^7	∞
$V^{\gamma}_{\sigma^{\gamma}}(y_0)$	7.5189	7.3522	7.2287	7.2222

tab. 7.1: Value function under Tracking-policy σ^{γ} for different γ

7.1 Control Problems in Stochastic Networks

In this section we present a rather general model for a stochastic queueing network. The state process is formulated as a continuous-time Markov chain (Y_t) in $I\!N_0^N$, where the j-th component of (Y_t) gives the number of jobs at queue j at time t. To keep the formulation simple we assume that arrival and service times are independent and exponentially distributed. The model allows to control the transition rates of the process at each point in time in a non-anticipating fashion. However, it is known that in this case the optimal policy can be found among the discrete-time policies, where decisions have to be taken at state changes only. The formulation includes admission control, routing, service control and scheduling. This leads to the following Markov decision process (see e.g. Sennott (1998), Tijms (1986)): we assume that there are N queues, hence the state space is $S = I\!N_0^N$. The action space $U \subset I\!R_+^K$ has to be compact and convex. The generator $Q = (q(y, u, y'))$ of (Y_t) should satisfy the following conditions for all $y, y' \in S$:

(i) $D(y) := \{u \in U \mid q(y, u, y') = 0, \text{ if } y' \notin S\} \neq \emptyset$.

(ii) there exists a linear function $b : U \to I\!R^N$ such that for all $u \in D(y)$

$$\sum_{y' \in S} (y' - y) q(y, u, y') = b(u).$$

(iii) there exists a $q \in \mathbb{R}_+$ with $\sup_{u\in U} \sup_{y,y'\in S} |q(y,u,y')| < q$.

The set $D(y)$ is the set of admissible actions in state y. As usual define $q(y,u) := \sum_{y'\neq y} q(y,u,y')$.
In the Rybko-Stolyar network of the introduction we have for example $U = \{u \in [0,1]^4 \mid u_1 + u_4 \leq 1,\ u_2 + u_3 \leq 1\}$ and $D(y) = \{u \in U \mid y_j = 0 \Rightarrow u_j = 0, j = 1,2,3,4\}$. For $u \in D(y)$ the generator is

$$\begin{aligned}
&q(y,u,y+e_1) = \lambda_1\\
&q(y,u,y+e_3) = \lambda_3\\
&q(y,u,y-e_2) = \mu_2 u_2\\
&q(y,u,y-e_4) = \mu_4 u_4\\
&q(y,u,y+e_2-e_1) = \mu_1 u_1\\
&q(y,u,y+e_4-e_3) = \mu_3 u_3
\end{aligned}$$

The cost rate function $c : S \times U \to \mathbb{R}_+$ of the general model should satisfy

(i) $c(y,u) = c_1(y) + c_2 u$ with $c_2 \in \mathbb{R}_+^K,\ c_1 : \mathbb{R} \to \mathbb{R}_+$.
(ii) c_1 is lower semicontinuous.

Denote by (T_n), $T_0 := 0$ the sequence of jump times of the Markov process (Y_t). A policy $\pi = (f_0, f_1, \ldots)$ for the Markov decision process is a sequence of decision rules $f_n : S \to U$ with $f_n(y) \in D(y)$, where f_n is applied at time T_n. For a fixed policy π and initial state $y \in S$, there exists a family of probability measures P_y^π on a measurable space $(\Omega, \mathcal{F})$ and stochastic processes (Y_t) and (π_t) such that for $0 =: T_0 < T_1 < T_2 < \ldots$

$$Y_t = Y_{T_n},\ T_n \leq t < T_{n+1}$$

$$\pi_t = f_n(Y_{T_n}),\ T_n \leq t < T_{n+1}$$

and

(i) $P_y^\pi(Y_0 = y) = P_y^\pi(T_0 = 0) = 1$ for all $y \in S$.

(ii) $P_{y_0}^\pi(T_{n+1} - T_n > t \mid T_0, Y_{T_0}, \ldots, T_n, Y_{T_n} = y) = e^{-q(y,f_n(y))t}$ for all $y \in S, t \geq 0$.

(iii) $P_{y_0}^\pi(Y_{T_{n+1}} = y' \mid T_0, Y_{T_0}, \ldots, T_n, Y_{T_n} = y, T_{n+1}) = \frac{q(y,f_n(y),y')}{q(y,f_n(y))}$ for $y, y' \in S, y \neq y'$ and zero, if $y = y'$.

In this section we are interested in the discounted cost criterion and define

$$V_\pi(y) = E_y^\pi \left[\int_0^\infty e^{-\beta t} c(Y_t, \pi_t)\, dt \right]$$

The optimization problem is

$$V(y) = \inf_\pi V_\pi(y).$$

In the sequel we assume that $D(y)$ is compact for all $y \in S$ and the mapping $u \to q(y,u,y')$ is continuous for all $y, y' \in S$. Under these assumptions, there exists an optimal stationary policy for the β-discounted problem. Moreover, this policy is optimal among all non-anticipating policies. The value iteration is of the form

$$V_{n+1}(y) = \min_{u \in U} \left[\frac{1}{\beta + q(y,u)} c(y,u) + \frac{1}{\beta + q(y,u)} \left(\sum_{y' \neq y} q(y,u,y') V_n(y') \right) \right].$$

Although problems of this type can in principle be solved by policy iteration, the size of the state space, even for simple examples makes this procedure intractable. Hence we would be satisfied with a policy which is in some sense "good" and easily computable. Let us now introduce a scaling parameter $\gamma > 0$ for the stochastic process as follows: let $\{y^\gamma\}$ be a sequence of initial states such that $\lim_{\gamma \to \infty} \frac{y^\gamma}{\gamma} = y$ for $y \in S$. To ease notation we will assume for our problem that $y^\gamma = \gamma y$ for all $\gamma \in \mathbb{N}$, though the proofs are in a more general setting. Denote by $(\hat{Y}_t^\gamma)$ the state process with initial state y^γ under a fixed policy $\pi^\gamma = (f_n^\gamma)$ and define by

$$Y_t^\gamma := \frac{1}{\gamma} \hat{Y}_{\gamma t}^\gamma$$

the scaled state process. Note that $(\hat{Y}_t^\gamma)$ is a process on the state space $S = \mathbb{N}_0^N$, whereas (Y_t^γ) is a process on the state space $\frac{1}{\gamma} S$. If we define the policy $\tilde{\pi}^\gamma = (\tilde{f}_n^\gamma)$ on the state space $\frac{1}{\gamma} S$ by $\tilde{f}_n^\gamma(\frac{1}{\gamma} y) = f_n^\gamma(y)$ and the generator $\tilde{Q}^\gamma$ by $\tilde{q}(\frac{1}{\gamma} y, u, \frac{1}{\gamma} y') = \gamma q(y,u,y')$, then the corresponding process $(\tilde{Y}_t^\gamma)$ is in distribution equal to the process (Y_t^γ). The scaled action process is defined by

$$\pi_t^\gamma := f_n^\gamma(\hat{Y}_{T_n}^\gamma), \quad \text{if } T_n \leq \gamma t < T_{n+1}$$

where (T_n) are the jump times of process $(\hat{Y}_t^\gamma)$. As γ tends to ∞ the intensity of the scaled process increases by factor γ, while the jump heights decrease by the same rate. This scaling is refered to as *fluid scaling*. The scaled value function is then defined by

$$V_{\pi^\gamma}^\gamma(y) = E_y^{\pi^\gamma} \left[\int_0^\infty e^{-\beta t} c(Y_t^\gamma, \pi_t^\gamma) \, dt \right].$$

The optimization problem is as before, where we now write $V_{\pi^\gamma}^\gamma(y)$ and $V^\gamma(y)$ respectively, to make the dependence on γ explicit.
Associated with the discounted stochastic network optimization problem is the following deterministic control problem

$$(F) \begin{cases} \int\limits_0^\infty e^{-\beta t} c(y_t, a_t) \, dt \to \min \\ y_t = y + \int\limits_0^t b(a_s) \, ds \\ y_t \geq 0 \\ a_t \in U, \ t \geq 0 \end{cases}$$

We will call (F) the fluid problem. The value function of this problem will be denoted by $V^F(y)$ and the optimal control (which exists due to Theorem 2.5) and state trajectory by a_t^* and y_t^* respectively.

7.2 An Asymptotic Lower Bound on the Value Function

In this section we will show that the value function V^F of the fluid problem (F) provides an asymptotic lower bound on the value function $V_{\pi^\gamma}^\gamma$ of the β-discounted stochastic network, irrespective of the chosen sequence of policies (π^γ). We denote by (Y_t^γ), for $\gamma \in I\!N$, the state process under fixed policy $\pi^\gamma = (f_n^\gamma)$ and initial state y. For the convergence results which follow, the processes (Y_t^γ) are defined on a common probability space $(\Omega', \mathcal{F}', P_y)$. Such a probability space can be constructed. As usual, we denote by $(Y_t^\gamma) \Rightarrow (Y_t)$ the weak convergence of the processes as $\gamma \to \infty$. We understand the processes (Y_t^γ) as random elements with values in $D^N[0,\infty)$, which is the space of $I\!R^N$-valued functions on $[0,\infty)$ that are right continuous and have left-hand limits and all endowed with the Skorokhod topology. For the following Lemma and Theorem 7.2 we suppose that $\pi^\gamma = (f^\gamma)^\infty$ is a stationary policy and define the process

$$M_t^\gamma = Y_t^\gamma - y - \int_0^t b(\pi_s^\gamma)\, ds.$$

We will first show

Lemma 7.1: $(M_t^\gamma) \Rightarrow 0$ as $\gamma \to \infty$.

Proof: Let $\pi^\gamma = (f^\gamma, f^\gamma, \ldots)$ and thus

$$\pi_t^\gamma = f^\gamma(\hat{Y}_{T_n}^\gamma), \quad T_n \le \gamma t < T_{n+1}.$$

Denote by $\mathcal{F}_t^\gamma = \sigma(Y_t^\gamma)$ the σ-algebra generated by the process (Y_t^γ). From the Dynkin formula we can conclude that $(M_j^\gamma(t)),\ j = 1, \ldots, N$ is a martingale w.r.t. the filtration $(\mathcal{F}_t^\gamma)$. This follows, since by definition the generator $\mathcal{A}$ of the process (Y_t^γ) is

$$\mathcal{A}\, g(\frac{1}{\gamma}y) = \sum_{y'} \left(g(\frac{1}{\gamma}y') - g(\frac{1}{\gamma}y) \right) \gamma q(y, f^\gamma(y), y')$$

where $g : \frac{1}{\gamma}S \to I\!R$. Setting $g(y) = y_j,\ j = 1, \ldots, N$ it follows with Proposition 14.13 in Davis (1993) and Assumption (ii) on the generator that $(M_j^\gamma(t))$ is a martingale. Define $\tau_n := \inf\{t \ge 0 \mid M_j^\gamma(t) \ge n\},\ n \in N$. Since M_j^γ has jumps of size $\frac{1}{\gamma}$, $M_j^\gamma(t \wedge \tau_n)$ is bounded and hence a square integrable martingale. Using the Lemma of Fatou we obtain

$$E_y\left[(M_j^\gamma(t))^2\right] \leq \liminf_{n\to\infty} E_y\left[(M_j^\gamma(t\wedge\tau_n))^2\right] = \liminf_{n\to\infty} E_y\left[< M_j^\gamma(t\wedge\tau_n) >\right]$$
$$\leq \frac{1}{\gamma^2} E_y\,[\text{ number of jumps in } [0,t]] \leq \frac{1}{\gamma^2} q\gamma t = O(\frac{1}{\gamma})$$

where $< M_j^\gamma(t) >$ is the quadratic variation of $M_j^\gamma(t)$. Applying Doob's inequality gives us

$$E_y\left[\sup_{0\leq s\leq t} (M_j^\gamma(s))^2\right] \leq 4E_y\left[(M_j^\gamma(t))^2\right] \leq O(\frac{1}{\gamma}).$$

Hence we have that $(M_t^\gamma) \Rightarrow 0$ for $\gamma \to \infty$ on compact intervals. Applying Theorem VI.16 in Pollard (1984) we obtain $(M_t^\gamma) \Rightarrow 0$ for $\gamma \to \infty$. □

Theorem 7.2:
Every sequence $(Y_t^\gamma, \pi_t^\gamma)$ has a further subsequence $(Y_t^{\gamma_n}, \pi_t^{\gamma_n})$ such that $(Y_t^{\gamma_n}, \pi_t^{\gamma_n}) \Rightarrow (Y_t, R_t)$ and the limit satisfies with $\pi_t := \int_U uR_t(du)$

(i) $Y_t = y + \int_0^t b(\pi_s)\, ds$.
(ii) $Y_t \in I\!R_+^N$.
(iii) $\pi_t \in U$.

Proof: Let us interpret (π_t^γ) as a random element $(R_t^\gamma) \in \mathcal{R}$, hence $\pi_t^\gamma = \int_U uR_t^\gamma(du)$ for all $t \geq 0$. The first step is to show that the sequence (Y_t^γ, R_t^γ) is tight. Due to Proposition 3.2.4 in Ethier/Kurtz (1986) we can do this separately. As far as (R_t^γ) is concerned, it is trivially tight, since $\mathcal{R}$ is compact. For (Y_t^γ) we use the conditions given in Kushner (1990) Theorem 4.4. That is we have to check

(i) $\lim_{m\to\infty} \sup_\gamma P_y(\|Y_t^\gamma\| \geq m) = 0$ for all $t \geq 0$.
(ii) $\lim_{\delta\to 0} \limsup_{\gamma\to\infty} \sup_{\tau\leq T} E_y\left[\min\{1, \|Y_{\tau+\delta}^\gamma - Y_\tau^\gamma\|\}\right] = 0$.

We make now use of the fact that $(\|Y_t^\gamma - \frac{1}{\gamma}y^\gamma\|)$ is stochastically dominated by a Poisson process (Λ_t^γ) with parameter $q\gamma$ and jumps of height $\frac{1}{\gamma}$. With the Chebychev inequality we obtain

$$P_y(\|Y_t^\gamma\| \geq m) \leq \frac{1}{m^2} E_y\left[\|Y_t^\gamma\|^2\right] \leq \frac{1}{m^2}\left((qt)^2 + \frac{qt}{\gamma} + 2qt\frac{\|y^\gamma\|}{\gamma} + \frac{\|y^\gamma\|^2}{\gamma^2}\right)$$

which implies (i). For (ii) we note that $E_y\left[\min\{1, \|Y_{\tau+\delta}^\gamma - Y_\tau^\gamma\|\}\right] \leq \delta q$. Therefore, (Y_t^γ, R_t^γ) is tight, which gives us a subsequence $(Y_t^{\gamma_n}, R_t^{\gamma_n})$ weakly converging to a limit (Y_t, R_t). By Skorokhod's Theorem (Ethier/Kurtz (1986) Theorem 3.1.8) the process can be constructed on the same probability space such that the convergence is almost sure. Since U is convex, we can define $\pi_t := \int_U uR_t(du) \in U$ for all $t \geq 0$. Using Lemma A.4 we know that almost sure

$$\int_0^t \int_U uR_s^{\gamma_n}(du)\, ds \to \int_0^t \pi_s\, ds.$$

Together with Lemma 7.1, (i) and (iii) follow. Because of $Y_t^\gamma \in I\!R_+^N$ for all γ we obtain (ii) and the proof is complete. $\square$

Now we are able to prove the main theorem of this section

Theorem 7.3:
For all sequences of policies (π^γ) and initial states $y \in S$ we obtain

$$\liminf_{\gamma\to\infty} V_{\pi^\gamma}^\gamma(y) \geq V^F(y).$$

Proof: Suppose first that $\pi^\gamma = (f^\gamma, f^\gamma, \ldots)$ is a stationary policy. Let $(Y_t^{\gamma_n}, \pi_t^{\gamma_n})$ be a subsequence such that $(Y_t^{\gamma_n}, \pi_t^{\gamma_n}) \Rightarrow (Y_t, \pi_t)$ and $y^{\gamma_n} = \gamma_n y$ for all $n \in I\!N$. Due to the assumption on the cost function we have

$$E_y\left[\int_0^\infty e^{-\beta t} c(Y_t^{\gamma_n}, \pi_t^{\gamma_n})\, dt\right] = E_y\left[\int_0^\infty e^{-\beta t} c_1(Y_t^{\gamma_n})\, dt\right] + E_y\left[\int_0^\infty e^{-\beta t} c_2 \pi_t^{\gamma_n}\, dt\right]$$

Let us first look at the second term. Define the mapping $\hat{c}_2 : \mathcal{R} \to I\!R_+$ by

$$\hat{c}_2(r) := c_2 \int_0^\infty e^{-\beta t} \int_U u r_t(du) dt.$$

It is easy to see that $\hat{c}_2$ is continuous (cf. Lemma A.4) and since U is compact, $\hat{c}_2$ is bounded on $\mathcal{R}$. Hence we have

$$\lim_{n\to\infty} E_y\left[\int_0^\infty e^{-\beta t} c_2 \pi_t^{\gamma_n}\, dt\right] = E_y\left[\int_0^\infty e^{-\beta t} c_2 \pi_t\, dt\right]$$

Now define $\hat{c}_1^m : D^N[0,\infty) \to I\!R_+$ by

$$\hat{c}_1^m(y) := \int_0^m e^{-\beta t} c_1^m(y_t) dt,$$

where $c_1^m \uparrow c_1$ and $c_1^m : I\!R^N \to I\!R_+$ is continuous (see Lemma 7.14 in Bertsekas/Shreve (1978)). Hence $\hat{c}_1^m$ is continuous and thus $\hat{c}_1^m(Y_t^{\gamma_n}) \Rightarrow \hat{c}_1^m(Y_t)$. Therefore, we obtain with the Lemma of Fatou and since the convergence $c_1^m \uparrow c_1$ is monotone

$$\liminf_{n\to\infty} E_y\left[\int_0^\infty e^{-\beta t} c_1(Y_t^{\gamma_n})\, dt\right] = \liminf_{n\to\infty} \lim_{m\to\infty} E_y\left[\hat{c}_1^m(Y_t^{\gamma_n})\right] \geq$$

$$\lim_{m\to\infty} \liminf_{n\to\infty} E_y \left[\hat{c}_1^m(Y_t^{\gamma_n})\right] \geq \lim_{m\to\infty} E_y \left[\hat{c}_1^m(Y_t)\right] = E_y \left[\int_0^\infty e^{-\beta t} c_1(Y_t)dt\right].$$

From Theorem 7.2 we know that the limit (Y_t, π_t) of every converging subsequence is for almost all ω an admissible state-action-trajectory for the fluid problem (F). Hence we have in particular

$$E_y \left[\int_0^\infty e^{-\beta t} c(Y_t, \pi_t)\, dt\right] \geq V^F(y)$$

and thus $\liminf_{\gamma\to\infty} V_{\pi^\gamma}^\gamma(y) \geq V^F(y)$. Since for arbitrary policies $\pi^\gamma = (f_0^\gamma, f_1^\gamma, \ldots)$ it holds that $V_{\pi^\gamma}^\gamma \geq V_{f^\gamma}^\gamma(y)$, where $(f^\gamma)^\infty$ is the optimal policy, the statement follows.

□

7.3 β-Discounted Asymptotic Optimality

For the β-discounted problem we will show that it is possible at least for some network models to construct a policy in such a way that the lower bound of the last section is achieved in the limit. We will call a policy with this property β-discounted asymptotically optimal. This notion coincides with the ones used by Meyn (1997) and Maglaras (1998b). A crucial observation for this construction is that the optimal control a^* in problem (F) is often piecewise constant (see Theorem 5.1). Otherwise, it is possible to construct for every $\varepsilon > 0$ a piecewise constant policy which is ε-optimal (so-called 'Chattering Theorem' see e.g. Kushner/Dupuis (1992) Section 4.6). The implementation of our policy is a direct translation of the fluid solution. The policy itself is instationary, i.e. the decision depends also on the current time. A state (y, t) consists now of the queue length and the time at which the jump occurs. The policy is defined in the following way: suppose that $a_t^* = u^*(\nu)$ on the interval $[t_\nu, t_{\nu+1}),\ \nu = 0, 1, \ldots, m,\ t_0 := 0$ and use the decision rule

$$f^\gamma(y, t) = u^*(\nu),\ \text{if } \gamma t_\nu \leq t < \gamma t_{\nu+1},$$

irrespective of the state Y_t the network is in. This of course may lead to unfeasible allocations where we want to serve a job though there is none there. In such cases we reduce the service rate to zero. We will call a policy of this type *Tracking-policy*. Obviously these policies are instationary and the only necessary information about the state is which components are zero. At first glance this policy may look very inefficient, however under fluid scaling it becomes much more important to catch the right drift of the process instead of being locally optimal. We will show that Tracking-policies are asymptotically optimal for two important classes of control problems in stochastic networks. To do this, we need a further assumption on the cost rate function

Assumption 7.1:

(i) $y \mapsto c_1(y)$ is increasing and convex.

(ii) There exist constants $C_0 \in \mathbb{R}_+, k \in \mathbb{N}$ such that for all $y \in \mathbb{R}^N$

$$c_1(y) \leq C_0(1 + \|y\|^k)$$

Multiclass Queueing Networks (cf. Dai (1995))
In the literature the multiclass queueing network is defined as follows: there are d single-server stations $k = 1, \ldots, d$ and server k is responsible for the jobs at queue $j \in K_k \subset \{1, \ldots, N\}$. Each queue j has exogenous arrivals at rate λ_j. The potential service rate of sever k is μ_k. Upon completion of service of a job at queue j, it is routed to queue i with probability p_{ji}, independent of all previous history. The optimization problem is to schedule the severs among their queues in order to minimize the discounted expected cost of the system. We obtain this network as a special case of our general model in the following way: denote by $\{K_1, \ldots, K_d\}$, $d < N$ a partition of the set $\{1, \ldots, N\}$. The action space is given by $U = \{u \in [0,1]^N \mid \sum_{j \in K_k} u_j \leq 1,\ k = 1, \ldots, d\}$, where u_j is the fraction of the k-th server that is allocated to queue $j \in K_k$. As in Section 6.2 define the matrix $A = D(I - P)$, where $P = (p_{ji})$ is transient, i.e. $\sum_{n=0}^{\infty} P^n < \infty$ and D is an N-dimensional diagonal matrix with elements $\mu_j \geq 0$ on the diagonal. The linear function b is of the form $b(u) = \lambda - A^T u$ with $\lambda \in \mathbb{R}_+^N$. The set of admissible actions in state $y \in S$ is $D(y) = \{u \in U \mid y_j = 0 \Rightarrow u_j = 0,\ j = 1, \ldots, N\}$. Suppose that a_t^* is the optimal control in the corresponding fluid model and $a_t^* = u^*(\nu)$ on $[t_\nu, t_{\nu+1})$, $\nu = 0, 1, \ldots, m$. The Tracking-policy $\sigma^\gamma = (f^\gamma, f^\gamma, \ldots)$ is formally defined by

$$f^\gamma(y,t) = u^*(\nu) \wedge \delta(y), \quad \text{if } \gamma t_\nu \leq t < \gamma t_{\nu+1},$$

where $\delta(y) = (\delta_1(y), \ldots, \delta_N(y))$ is given by

$$\delta_j(y) = \begin{cases} 0, & \text{if } y_j = 0 \\ 1, & \text{if } y_j > 0 \end{cases}.$$

Note that $f^\gamma(y,t) \in D(y)$ for all $t \geq 0$. We will now show

Theorem 7.4:
Under Assumption 7.1, the Tracking-policy σ^γ in the multiclass queueing network satisfies for $y \in S$

$$\lim_{\gamma \to \infty} V_{\sigma^\gamma}^\gamma(y) = V^F(y)$$

and hence σ^γ is β-discounted asymptotically optimal.

Proof: Let us first consider a continuously defined policy π_t^γ with corresponding scaled process (Y_t^γ) which is given by

$$\pi_t^\gamma = u^*(\nu) \wedge \delta(Y_t^\gamma), \quad \text{if } \gamma t_\nu \le t < \gamma t_{\nu+1}.$$

Denote by $(\bar{Y}_t^\gamma)$ the scaled process, where we use the Tracking-policy σ^γ. The difference between these two processes is the duration of the time intervals on which the actions $u^*(\nu) \wedge \delta(y)$ are taken. If (T_n^γ) is the sequence of jump times of process $(\bar{Y}_t^\gamma)$ and $N^\gamma(t) := \inf\{n \in I\!N \mid T_n^\gamma > t\}$ then we obtain for $\gamma \to \infty$

$$T_{N^\gamma(t)}^\gamma \to t \quad \text{a.s.}$$

This means that the change points, where we use a different server allocation in the processes (Y_t^γ) and $(\bar{Y}_t^\gamma)$ converge together a.s. Hence (Y_t^γ) and $(\bar{Y}_t^\gamma)$ have the same limit. Therefore, it suffices to prove the statement for the policy π_t^γ. Define $Y_0^\gamma = y \in S$ for all γ.
On time interval $[t_\nu, t_{\nu+1})$ we can think of the process (Y_t^γ) as a Jackson-network with N servers and fixed service rates $\mu_1 u_1^*(\nu), \ldots, \mu_N u_N^*(\nu)$, $\nu = 1, \ldots, m$. In this network server k is only idle when there is no job at queue k. This queueing discipline is called work-conserving. We will now look at the process on time interval $[0, t_1)$ only. Under the Tracking-policy we have $\pi_t^\gamma = U_1^* \hat{\pi}_t^\gamma$ where $U_1^* = diag(u^*(1))$ and $\hat{\pi}_t^\gamma \in [0,1]^N$ and our process fulfills for all $t \in [0, t_1)$

$$Y_t^\gamma = \frac{1}{\gamma} y^\gamma + \int_0^t \left(\lambda - A^T U_1^* \hat{\pi}_s^\gamma\right) ds - M_t^\gamma \ge 0 \tag{7.16}$$

$$\hat{\pi}_t^\gamma \in [0,1]^N \tag{7.17}$$

$$\int_0^\infty Y_t^\gamma (1\!\!1 - \hat{\pi}_t^\gamma)\, dt = 0 \tag{7.18}$$

As before we can show that every sequence $(Y_t^\gamma, \hat{\pi}_t^\gamma)$ has a further subsequence $(Y_t^{\gamma_n}, \hat{\pi}_t^{\gamma_n})$ such that $(Y_t^{\gamma_n}, \hat{\pi}_t^{\gamma_n}) \Rightarrow (Y_t, \hat{\pi}_t)$ and the limit satisfies for all $t \in [0, t_1)$ a.s. (see Dai (1995) for the convergence of (7.18))

$$Y_t = y + \int_0^t \left(\lambda - A^T U_1^* \hat{\pi}_s\right) ds \ge 0 \tag{7.19}$$

$$\hat{\pi}_t \in [0,1]^N \tag{7.20}$$

$$\int_0^\infty Y_t (1\!\!1 - \hat{\pi}_t)\, dt = 0 \tag{7.21}$$

From Chen (1995) (p. 641) we know that the solution $(Y_t, \hat{\pi}_t)$ of (7.19)-(7.21) is unique on the interval $[0, t_1)$ up to sets of measure zero. However, we know by definition that $u^*(1)$ is admissible for the fluid problem (F) on $[0, t_1)$. Thus, we get that $(y_t^*, 1\!\!1)$ is the unique solution of (7.19)-(7.21) on $[0, t_1)$. Since the limit is independent of ω, this implies (up to sets of measure zero)

$$(Y_t^\gamma, \hat{\pi}_t^\gamma) \Rightarrow (y_t^*, 1\!\!1) \text{ on } [0, t_1)$$

Thus, in particular $Y_{t_1}^{\gamma} \to y_{t_1}^*$ a.s. Inductively we obtain in this way that the convergence holds for all $t \geq 0$.
Now it remains to show that $\lim_{\gamma\to\infty} V_{\pi^\gamma}^{\gamma}(y) = V^F(y)$. Due to the proof of Theorem 7.3 it is left to show that

$$\lim_{\gamma\to\infty} E_y \left[\int_0^\infty e^{-\beta t} c_1(Y_t^\gamma)\, dt\right] = \int_0^\infty e^{-\beta t} c_1(y_t^*)\, dt.$$

First, since $c_1 \geq 0$, it holds that

$$E_y \left[\int_0^\infty e^{-\beta t} c_1(Y_t^\gamma)\, dt\right] = \int_0^\infty e^{-\beta t} E_y\left[c_1(Y_t^\gamma)\right]\, dt.$$

Using Assumption 7.1 (i) we obtain that $c_1(Y_t^\gamma)$ is stochastically dominated by $c_1(\Lambda_t^\gamma, \dots, \Lambda_t^\gamma)$ for all $t \geq 0$, where Λ_t^γ is a Poisson process with parameter $q\gamma$ and jump heights $\frac{1}{\gamma}$. From Bäuerle (1998a) Lemma 1 it follows that for all $t \geq 0$ and $\gamma \geq \gamma'$

$$\Lambda_t^\gamma \leq_{cx} \Lambda_t^{\gamma'}$$

where $\leq_{cx}$ is the convex ordering. Thus, we obtain with Assumption 7.1

$$E_y\left[c_1(Y_t^\gamma)\right] \leq E_y\left[c_1(\Lambda_t^\gamma, \dots, \Lambda_t^\gamma)\right] \leq E_y\left[c_1(\Lambda_t, \dots, \Lambda_t)\right] < \infty$$

Moreover, since c_1 is also continuous, we obtain $c_1(Y_t^\gamma) \Rightarrow c_1(y_t^*)$ for all $t \geq 0$. Applying dominated covergence we obtain

$$\lim_{\gamma\to\infty} E_y[c_1(Y_t^\gamma)] = c_1(y_t^*).$$

Using Assumption 7.1 (ii) we obtain $\int_0^\infty e^{-\beta t} E_y\left[c_1(\Lambda_t, \dots, \Lambda_t)\right] dt < \infty$ and applying again dominated convergence yields

$$\lim_{\gamma\to\infty} E_y \left[\int_0^\infty e^{-\beta t} c_1(Y_t^\gamma)\, dt\right] = \int_0^\infty e^{-\beta t} c_1(y_t^*)\, dt$$

and the statement is proven. □

Admission and Routing Problems
Under an admission and routing problem, we understand the following model: there are d external streams of jobs arriving with intensity λ_k, $k = 1, \dots, d$ and jobs of type k can be routed to the queues $j \in K_k \subset \{1, \dots N\}$. Each queue j has a server with potential service rate μ_j. The optimization problem is to decide upon admission/rejection of jobs and in case of admission, to decide upon the routing of the jobs in order to minimize the discounted expected cost of the system. Our general model specializes to an admission and routing problem in the following way: let $K_1, \dots, K_d,\ d < N$ be subsets of the set $\{1, \dots, N\}$. The action space is given by $U = \{(u, v) \in [0,1]^{d\times N} \times [0,1]^N \mid u_{kj} = 0,$ if $j \notin K_k, \sum_{j\in K_k} u_{kj} \leq 1,\ k =$

$1,\dots,d, 0 \le v_j \le 1,\ j = 1,\dots,N\}$, where u_{kj} is the fraction of jobs of type k which is routed to queue j. v_j is the activation level of server j. Let $\lambda \in \mathbb{R}^d_+$ and let D be an N-dimensional diagonal matrix with elements $\mu_j \ge 0$ on the diagonal. Thus, the linear function b is of the form $b(u) = \lambda u - Dv$. The set of admissible actions in state $y \in S$ is $D(y) = \{(u,v) \in U \mid y_j = 0 \Rightarrow (\lambda u - Dv)_j \ge 0,\ j = 1,\dots,N\}$. Suppose that a^*_t is the optimal control in the corresponding fluid model and $a^*_t = u^*(\nu)$ on $[t_\nu, t_{\nu+1})$, $\nu = 0,1,\dots,m$. The Tracking-policy is exactly defined as before. Hence we obtain

Theorem 7.5:
Suppose Assumption 7.1 is valid.

a) The Tracking-policy σ^γ in the admission and routing problem satisfies for $y \in S$

$$\lim_{\gamma\to\infty} V^\gamma_{\sigma^\gamma}(y) = V^F(y)$$

and hence σ^γ is β-discounted asymptotically optimal.

b) $V^\gamma_{\sigma^\gamma}(y)$ is decreasing in γ for all $y \in S$. In particular, $V^F(y)$ is a lower bound for all $V^\gamma_{\sigma^\gamma}(y)$.

Proof: a) The idea is the same as in the proof of Theorem 7.4. Let us first look at time interval $[0,t_1)$. Since there is no rerouting, each queue separately is an M/M/1-queue with input rates λu and output rates Dv. From the theory of large deviation (see e.g. Shwartz/Weiss (1995) chapter 11) we know that

$$Y^\gamma_t \Rightarrow y + (\lambda u^*(1) - Dv^*(1))t$$

on $[0,t_1)$. Using the same arguments as before we can complete the first part of the proof.
b) Denote by $\xi^\gamma_j(t) = A^\gamma_j(t) - B^\gamma_j(t),\ j = 1,\dots,N$, the difference between a Poisson process $A^\gamma_j(t)$ with parameter $\gamma\sum_k \lambda_k u_{kj}$ and jump heights $\frac{1}{\gamma}$ and a Poisson process $B^\gamma_j(t)$ with parameter $\gamma\mu_j$ and jump heights $\frac{1}{\gamma}$ which are independent. From Bäuerle (1998a) it can be deduced that for $\gamma \ge \gamma'$ and $0 \le t_1 < t_2 < \dots t_n < \infty$

$$(\xi^\gamma_1(t_1),\dots,\xi^\gamma_1(t_n),\dots,\xi^\gamma_N(t_1),\dots,\xi^\gamma_N(t_n)) \le_{cx}$$

$$(\xi^{\gamma'}_1(t_1),\dots,\xi^{\gamma'}_1(t_n),\dots,\xi^{\gamma'}_N(t_1),\dots,\xi^{\gamma'}_N(t_n))$$

where $\le_{cx}$ denotes the convex ordering. Now it holds that

$$Y^\gamma_j(t) = y_j + \xi^\gamma_j(t) + \sup_{0\le s\le t}\left(-\xi^\gamma_j(s)\right).$$

Since this is a convex functional, we obtain for all $\gamma \geq \gamma'$ and $0 \leq t_1 < t_2 < \ldots < t_n < \infty$

$$(Y_1^\gamma(t_1), \ldots, Y_1^\gamma(t_n), \ldots, Y_N^\gamma(t_1), \ldots, Y_N^\gamma(t_n)) \leq_{icx}$$
$$(Y_1^{\gamma'}(t_1), \ldots, Y_1^{\gamma'}(t_n), \ldots, Y_N^{\gamma'}(t_1), \ldots, Y_N^{\gamma'}(t_n))$$

where $\leq_{icx}$ denotes the increasing convex ordering. Using the assumptions on c_1 we obtain

$$\hat{c}_1(Y_t^\gamma) \leq_{icx} \hat{c}_1(Y_t^{\gamma'})$$

for $\gamma \geq \gamma'$ and the statement follows. □

Corollary 7.6:
In the multiclass queueing network as well as in the admission and routing problem we have for $y \in S$ under Assumption 7.1

$$\lim_{\gamma \to \infty} V^\gamma(y) = V^F(y).$$

Proof: From the previous theorems we obtain

$$V^F(y) = \limsup_{\gamma\to\infty} V_{\sigma^\gamma}^\gamma(y) \geq \limsup_{\gamma\to\infty} V^\gamma(y) = \limsup_{\gamma\to\infty} V_{\hat{\pi}^\gamma}^\gamma(y) \geq \liminf_{\gamma\to\infty} V_{\hat{\pi}^\gamma}^\gamma(y) \geq V^F(y),$$

where $\hat{\pi}^\gamma$ is the optimal policy for scaling parameter γ, which exists due to our assumption and the proof is complete. □

Remark 7.1:

a) Of course Theorems 7.4 - 7.5 are asymptotic statements, which means that the Tracking-policy is only good when the system is close to the limit. This situation occurs when the initial state is large and the system is operating with a high intensity.

b) Theorems 7.4 - 7.5 can be extended to the case where the interarrival times and service times are i.i.d. but arbitrary (cf. Dai (1995)).

c) If the cost rate function satisfies $c(\frac{1}{\gamma}y, u) = \frac{1}{\gamma}c(y, u)$, then the value function V_π^γ can be expresses with the help of the original value function V_π. An easy substitution gives us

$$V_\pi^\gamma(y) = \frac{1}{\gamma^2} V_\pi^{\frac{\beta}{\gamma}}(\gamma y),$$

where $V_\pi^{\frac{\beta}{\gamma}}$ is the original value function ($\gamma = 1$) with interest rate $\frac{\beta}{\gamma}$.

d) There are several alternatives for the implementation of the Tracking-policy. We explain the procedures here in the setting of the multiclass queueing network. The only thing one has to make sure is that the fraction of the server allocation to buffer j is in the long run equal to $u_j^*(\nu)$ on the time interval $[t_\nu, t_{\nu+1})$. If we are not allowed to split the server, there are two possibilities:

   (i) we interpret $u_j^*(\nu)$ as a *randomized decision*, i.e. we do a random experiment for each buffer independent of the history, where $u_j^*(\nu)$ is the probability that the k-th server is assigned to queue $j \in K_k$.

   (ii) when we can write $u_j^*(\nu) = \frac{\alpha_j}{\sum_{i \in K_k} \alpha_i}$, with $\alpha_j \in \mathbb{N}_0, j = 1, \ldots, N$, then we can follow a so-called *generalized round-robin* policy (cf. Dai (1998)): assign the k-th server in a cyclic fashion α_{j_1}-times to queue $j_1 \in K_k$, then α_{j_2}-times to queue $j_2 \in K_k$ and so on.

7.4 Average Cost Asymptotic Optimality

Let us finally look at the average cost case in our general network model. Here we define for a stationary policy $\pi = f^\infty$

$$G_f(y) = \limsup_{t \to \infty} \frac{1}{t} E_y^\pi \left[\int_0^t c(Y_s, \pi_s) \, ds \right]$$

and we want to solve

$$G(y) = \inf_f G_f(y).$$

The scaled processes are defined as before and we write

$$G_f^\gamma(y) = \limsup_{t \to \infty} \frac{1}{t} E_y^\pi \left[\int_0^t c(Y_s^\gamma, \pi_s^\gamma) \, ds \right].$$

We will assume here for the cost rate function that $c_1(y) = 0$ if $y = 0$ and $c_2 = 0$. A policy $\pi = f^\infty$ is called *stable* for the stochastic network, if the corresponding fluid model is stable (for a definition see e.g. Dai (1995) Definition 4.1).

Theorem 7.7:
If $\pi = f^\infty$ is a stable policy for our model and $\int c_1 d\nu < \infty$, where ν is the stationary distribution of the state process under policy π, then for $y \in S$

$$\lim_{\gamma \to \infty} G_f^\gamma(y) = 0$$

and π is hence average asymptotically optimal.

Proof: Since π is stable, we obtain from Dai (1995) Theorem 4.2 and Remark 2 that there exists a stationary distribution ν^γ on $\frac{1}{\gamma}S$ such that for $\gamma > 0$

$$G_f^\gamma(y) = E[c_1(Y_\infty^\gamma)] = \int c_1 \, d\nu^\gamma < \infty.$$

Obviously we have $\nu^\gamma(y) = \nu(\gamma y)$ for $y \in \frac{1}{\gamma}S$ and thus $\nu^\gamma \Rightarrow \delta_0$ for $\gamma \to \infty$. Since the $c_1(Y_\infty^\gamma)$ are uniformly integrable, we obtain the statement. $\square$

Appendix

A Sets and Functions

For the following definitions and theorems we suppose that the sets E, A, U, D and $\mathcal{R}$ are defined as in Section 2. For a proof of Theorem A.1 see Schäl (1975). Lemma A.3 and A.4 can be found in appendix B of Forwick (1998).

Definition A.1:
Let $v : E \to \bar{I\!R}$ be such that $v(x) < \infty$ for at least one point, then v is called *lower semicontinuous* if for all $x \in E$ and sequences (x_n) with $x_n \to x$

$$\liminf_{n\to\infty} v(x_n) \geq v(x).$$

Definition A.2:
Suppose $\psi : E \to D$ is defined by $\psi(x) = D(x)$. ψ is called *upper semicontinuous* if for all closed sets $B \subset D$

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

is again closed.

Theorem A.1: *(Measurable selection Theorem)*
Suppose $\psi : E \to D$ is defined by $\psi(x) = D(x)$. If ψ is compact-valued and upper semicontinuous and $v : D \to I\!R$ lower semicontinuous and bounded below on D, then there exists an $f^* \in F$ such that

$$v^*(x) := v(x, f^*(x)) = \min_{a\in D(x)} v(x, a)$$

and v^* is lower semicontinuous and bounded below on E.

Theorem A.2:
Let $v_n, v : D \to I\!R$ be lower semicontinuous functions bounded below. If $D(x)$ is compact for all $x \in E$ and $v_n \uparrow v$, then

$$\lim_{n\to\infty} \min_{a\in D(x)} v_n(x, a) = \min_{a\in D(x)} v(x, a)$$

for all $x \in E$.

Proof: For $x \in E$ define

$$l(x) = \lim_{n\to\infty} \min_{a \in D(x)} v_n(x,a) \quad \text{and} \quad u^*(x) = \min_{a\in D(x)} v(x,a).$$

Since $v_n \uparrow v$ we have $l \le u^*$. For the converse, fix $x \in E$. Let $a_n \in D(x)$ be such that $v_n(x,a_n) = \min_{a\in D(x)} v_n(x,a)$. Since $D(x)$ is compact we have a convergent subsequence (a_{n_k}) with $a_{n_k} \to a^* \in D(x)$ for $k \to \infty$. Using that the v_n are increasing we obtain for all $n_k \ge n$

$$v_{n_k}(x, a_{n_k}) \ge v_n(x, a_{n_k}).$$

Letting $k \to \infty$, the lower semicontinuity of v_n yields $l(x) \ge v_n(x,a^*)$. With $n \to \infty$ we finally obtain $l(x) \ge v(x,a^*) \ge u^*(x)$ and the proof is complete. □

Lemma A.3:

a) $r : I\!\!R_+ \to I\!\!P(U)$ is measurable if and only if

$$t \mapsto \int_U \psi(u) r(t,du)$$

is measurable for all continuous functions $\psi : U \to I\!\!R$.

b) $f : E \to \mathcal{R}$ is measurable if and only if

$$x \mapsto \int_0^\infty \int_U \psi(t,u) f(x,t,du)\, dt$$

is measurable for all measurable functions $\psi : I\!\!R_+ \times U \to I\!\!R$ such that $u \mapsto \psi(t,u)$ is continuous for all $t \ge 0$ and $\int_0^\infty \sup_{u\in U} |\psi(t,u)| dt < \infty$.

There are more general and elegant ways to define measurability (cf. Rieder (1975)), however Lemma A.3 is sufficient for our purpose.

Lemma A.4:
Let $r^n, r \in \mathcal{R}$. The following statements are equivalent

(i) $r^n \to r$ for $n \to \infty$.

(ii)
$$\int_0^\infty \int_U \psi(t,u) r_t^n(du)dt \to \int_0^\infty \int_U \psi(t,u) r_t(du)dt$$
for all measurable functions $\psi : I\!\!R_+ \times U \to I\!\!R$ such that $u \mapsto \psi(t,u)$ is continuous for all $t \ge 0$ and $\int_0^\infty \sup_{u\in U} |\psi(t,u)| dt < \infty$.

B Markov Chains

This part of the appendix is essentially taken from Meyn/Tweedie (1993). In the sequel, (X_n) is a homogeneous, discrete-time Markov chain with state space E and transition kernel p. For $A \in \mathfrak{B}(E)$ define $\tau_A := \min\{n \mid X_n \in A\}$.

Definition B.1:
A measure ϕ on $\mathfrak{B}(E)$ with $\phi(A) > 0$ implies $P_x(\tau_A < \infty) > 0$ for all $x \in E$ is called *irreducibility measure*.

There always exists an essentially unique irreducibility measure ψ on $\mathfrak{B}(E)$ such that

(i) for all $x \in E$, $\psi(A) > 0$ implies $P_x(\tau_A < \infty) > 0$.
(ii) $\psi(A) = 0$ implies $\psi(\{y|P_y(\tau_A < \infty) > 0\}) = 0$.
(iii) $\psi(A^c) = 0$ implies $A = A_0 \cup N$, where $\psi(N) = 0$ and $p(x, A_0) = 1$ for all $x \in A_0$.

Let us assume now that (X_n) is ψ-irreducible and denote by $\mathfrak{B}^+(E) := \{A \in \mathfrak{B}(E) \mid \psi(A) > 0\}$.

Definition B.2:

a) (X_n) is called *Harris recurrent*, if for all $A \in \mathfrak{B}^+(E)$

$$P_x(\tau_A < \infty) = 1$$

whenever $x \in A$.
b) (X_n) is called *positive Harris recurrent*, if it is Harris recurrent and has an invariant probability measure μ, i.e.

$$\mu(A) = \int_E p(x, A)\mu(dx)$$

C Viscosity Solutions

Suppose the situation of Section 4 is given. We will provide a definition of a constrained viscosity solution in terms of test functions. Usually viscosity solutions are defined with the help of sub- and superdifferentials of V and then shown afterwards that this is equivalent to the test function approach. However, since it is much more convenient to work with test functions we decided to use the following definition. For details about viscosity solutions see e.g. Bardi/Capuzzo-Dolcetta (1997), Sethi/Zhang (1994).

Definition C.1:
$y \mapsto V(y,z)$ is a constrained viscosity solution of

$$(\beta + q_z)V(x) = \min_{u \in U} \left[l(y,u) + b^z(u)V_y(x) \right]$$

if

a) $y \mapsto V(y,z)$ is continuous and $|V(x)| \leq C_0(1 + \|y\|^k)$.

b) for all continuously differentiable $\psi^1 : S \to I\!R$, where $V(x) - \psi^1(y)$ has a local maximum at $y = y_0 \in \overset{\circ}{S}$, relative to $\overset{\circ}{S}$, it holds that

$$(\beta + q_z)V(x_0) - \min_{u \in U} \left[\mathfrak{c}_V(y_0,u) + b^z(u)\psi^1_y(y_0) \right] \leq 0.$$

c) for all continuously differentiable $\psi^2 : S \to I\!R$, where $V(x) - \psi^2(y)$ has a local minimum at $y = y_0 \in S$, relative to S, it holds that

$$(\beta + q_z)V(x_0) - \min_{u \in U} \left[\mathfrak{c}_V(y_0,u) + b^z(u)\psi^1_y(y_0) \right] \geq 0.$$

References

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

[2] Alanyali M and Hajek B (1998) Analysis of simple algorithms for dynamic load balancing. *Math. Operations Res.* **22** 840-871.

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

[4] Atkins D and Chen H (1995) Performance evaluation of scheduling control of queueing networks: fluid model heuristics. *Queueing Systems* **21** 391-413.

[5] Avram F (1997) Optimal control of fluid limits of queueing networks and stochasticity corrections. In *Mathematics of Stochastic Manufacturing Systems* Lectures in Applied Mathematics, eds. Yin GG and Zhang Q, American Mathematical Society, Providence **33** 1-36.

[6] Avram F, Bertsimas D and Ricard M (1995) Fluid models of sequencing problems in open queueing networks: an optimal control approach. In *Stochastic Networks* IMA Volumes in Mathematics and its Applications, eds. Kelly FP and Williams RJ, Springer-Verlag, New York 199-234.

[7] Bäuerle (1998a) The advantage of small machines in a stochastic fluid production process. *Math. Meth. Operations Res.* **47**, 83-97.

[8] Bäuerle (1998b) Discounted stochastic fluid programs. Preprint.

[9] Bäuerle N, Brüstl G and Rieder U (1998) Optimal scheduling in heterogeneous two station queueing networks. *Math. Meth. Operations Res.* **48**, 337-347.

[10] Bäuerle N and Rieder U (2000) Optimal control of single-server fluid networks. To appear in *Queueing Systems.*

[11] Bardi M and Capuzzo-Dolcetta I (1997) *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations.* Birkhäuser, Boston.

[12] Bertsekas DP (1995) *Dynamic programming and optimal control, vol. 2.* Athena Scientific, Belmont Massachusetts.

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

[14] Bramson M (1996) Convergence to equilibria for fluid models of FIFO queueing networks. *Queueing Systems* **22** 5-45.

[15] Chen H (1995) Fluid approximations and stability of multiclass queueing networks: work-conserving disciplines. *Ann. Appl. Probab.* **5** 637-665.

[16] Chen H and Mandelbaum A (1994) Hierarchical modeling of stochastic networks part 1: fluid models. In *Stochastic modeling and analysis of manufacturing systems*, ed. D.D. Yao, Springer-Verlag, New York 47-105.

[17] Chen H and Yao DD (1993) Dynamic scheduling of a multiclass fluid network. *Operations Res.* **41** 1104-1115.

[18] Chen RR and Meyn S (1998) Value iteration and optimization of multiclass queueing networks. Preprint.

[19] Costa OLV (1990) Stationary distributions for piecewise deterministic Markov processes. *J. Appl. Probab.* **27** 60-73.

[20] Dai JG (1995) On positive Harris recurrence of multiclass queueing networks: a unified approach via fluid limit models. *Ann. of Appl. Probab.* **5** 49-77.

[21] Dai JG (1998) Stability of fluid and stochastic processing networks. Lecture Note at the University of Aarhus. *http://www.isye.gatech.edu/faculty/dai/*

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

[23] Ethier SN and Kurtz TG (1986) *Markov processes.* John Wiley & Sons, New York.

[24] Fleming WH and Rishel RW (1975) *Deterministic and stochastic optimal control.* Springer-Verlag, New York.

[25] Fleming WH and Soner HM (1992) *Controlled Markov processes and viscosity solutions.* Springer-Verlag, New York.

[26] Forwick L (1998) *Optimale Kontrolle stückweise deterministischer Prozesse.* PhD Dissertation at the University of Bonn.

[27] Glynn P and Sigman K (1992) Uniform Cesaro limit theorems for synchronous processes with applications to queues. *Stoch. Proc. Applns.* **40**, 29-43.

[28] Harrison JM (1996) The BIGSTEP approach to flow management in stochastic processing networks. In *Stochastic networks: stochastic control theory and applications.*, Royal Statistical Society Lecture Notes Series, eds. Kelly F, Zachary S and Ziedins I., Clarendon Press, Oxford 57-90.

[29] Hartman P (1982) *Ordinary Differential Equations.* 2nd Edition, Birkhäuser, Boston.

[30] Hernández-Lerma O and Lasserre JB (1996) *Discrete-time Markov control processes.* Springer-Verlag, New York.

[31] Hinderer K (1984) *Grundbegriffe der Wahrscheinlichkeitstheorie.* Springer-Verlag, Berlin Heidelberg.

[32] Hordijk A and Van der Duyn Schouten FA (1983) Average optimal policies in Markov decision drift processes with applications to a queueing and a replacement model. *Adv. Appl. Probab.* **15** 274-303.

[33] Kitaev M and Rykov V (1995) *Controlled queueing systems.* CRC-Press, Boca Raton.

[34] Klimov GP (1974) Time-sharing service systems I. *Theory Probab. Appl.* **19** 532-551.

[35] Kushner HJ (1990) Numerical methods for stochastic control problems in continuous time. *SIAM J. Control and Optim.* **5** 999-1048.

[36] Kushner HJ and Dupuis PG (1992) *Numerical methods for stochastic control problems in continuous time.* Springer-Verlag, New York.

[37] Maglaras C (1998a) Dynamic scheduling in multiclass queueing networks: stability under discrete-review policies. Preprint.

[38] Maglaras C (1998b) Discrete-review policies for scheduling stochastic networks: fluid-scale asymptotic optimality. Preprint.

[39] Maglaras C (1999) Discrete-review policies for scheduling stochastic networks: trajectory tracking and fluid-scale asymptotic optimality. Preprint.

[40] Meyn SP (1997) The policy iteration algorithm for average reward Markov decision processes with general state space. *IEEE Trans. Autom. Control* **42** 1663-1679.

[41] Meyn SP (1997) Stability and optimization of queueing networks and their fluid models. In *Mathematics of Stochastic Manufacturing Systems* Lectures in Applied Mathematics, eds. Yin GG and Zhang Q, American Mathematical Society, Providence Mathem. **33** 175-199.

[42] Meyn SP and Tweedie RL (1993a) *Markov chains and stochastic stability.* Springer-Verlag, London.

[43] Meyn SP and Tweedie RL (1993b) Stability of Markovian processes II: Continuous time processes and sampled chains. *Adv. Appl. Probab.* **25** 487-517.

[44] Meyn SP and Tweedie RL (1993c) Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes. *Adv. Appl. Probab.* **25** 518-548.

[45] Pollard D (1984) *Convergence of stochastic processes.* Springer-Verlag, New York.

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

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

[48] Presman E, Sethi S and Zhang Q (1995) Optimal feedback production planning in a stochastic N-machine flowshop. *Automatica* **31** 1325-1332.

[49] 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.

[50] Rajagopal S, Kulkarni VG and Stidham S (1995) Optimal flow control of stochastic fluid-flow systems. *IEEE Journal on selected areas in Communications* **13** 1219-1228.

[51] Rieder U (1994) *Stochastische Modelle und Optimierung.* Lecture Note at the University of Ulm.

[52] Rieder U (1975) Bayesian dynamic programming. *Adv. Appl. Probab.* **7** 330-348.

[53] Rockafellar RT (1970) *Convex analysis.* Princeton University Press, Princeton.

[54] Ross SM (1970) *Applied probability models with optimization applications.* Holden-Day, San Francisco.

[55] Rybko AN and Stolyar AL (1992) Ergodicity of stochastic processes describing the operation of open queueing networks. *Problems of Information Transmission* **28** 199-220.

[56] Schäl M (1975) Conditions for optimality and for the limit of n-stage optimal policies to be optimal. *Z. Wahrscheinlichkeitstheorie und verw. Geb.* **32** 179-196.

[57] Schäl M (1993) Average optimality in dynamic programming with general state space. *Math. Operations Res.***18** 163-172.

[58] Seierstad A and Sydsæter K (1987) *Optimal control theory with economic applications*, North-Holland, Amsterdam.

[59] Sennott LI (1989a) Average cost optimal stationary policies in infinite state Markov decision processes with unbounded costs. *Operations Res.* **37** 626-633.

[60] Sennott LI (1989b) Average cost semi-Markov decision processes and the control of queueing systems. *Probab. Engineering and Inform. Sci.* **3** 247-272.

[61] Sennott LI (1998) *Stochastic dynamic programming and the control of queues.* John Wiley & Sons, New York.

[62] Serfozo R (1982) Convergence of Lebesgue integrals with varying measures. *Sankhyā Ser. A* **44** 380-402.

[63] Sethi SP, Soner HM, Zhang Q and Jiang J (1992) Turnpike sets and their analysis in stochastic production planning problems. *Math. Operations Res.* **17** 932-950.

[64] Sethi SP, Suo W, Taksar MI and Yan H (1998) Optimal production planning in a multi-product stochastic manufacturing system with long-run average cost. *Discrete Event Dynamic Systems: Theory and Applications* **8** 1-18.

[65] Sethi SP, Suo W, Taksar MI and Zhang Q (1997) Optimal production planning in a stochastic manufacturing system with long-run average cost. *J. Optim. Theory Appl.* **92** 161-188.

[66] Sethi SP and Zhang Q (1994) *Hierarchical decision making in stochastic manufacturing systems.* Birkhäuser, Boston.

[67] Shwartz A and Weiss A (1995) *Large deviations for performance analysis.* Chapman & Hall, London.

[68] Stidham S and Weber R (1993) A survey of Markov decision models for control of networks of queues. *Queueing Systems* **13** 291-314.

[69] Sznajder R and Filar JA (1992) Some comments on a theorem of Hardy and Littlewood. *J. Optim. Theory and Appl.* **75** 201-208.

[70] Tijms HC (1986) *Stochastic modelling and analysis: a computational approach.* John Wiley & Sons, Chichester.

[71] Walrand J (1988) *An introduction to queueing networks.* Prentice-Hall, Englewood Cliffs, New Jersey.

[72] Weiss G (1995) On optimal draining of re-entrant fluid lines, in *Stochastic Networks* IMA Volumes in Mathematics and its Applications, eds. Kelly FP and Williams RJ, Springer-Verlag, New York 91-103.

[73] Weiss G (1996) Optimal draining of fluid re-entrant lines: Some solved examples, In *Stochastic networks: stochastic control theory and applications.*, Royal Statistical Society Lecture Notes Series, eds. Kelly F, Zachary S and Ziedins I., Clarendon Press, Oxford 19-34.

[74] Weiss G (1997) Algorithm for minimum wait draining of two-station fluid re-entrant line. Preprint.

[75] Winston W (1977) Optimality of the shortest-line discipline. *J. Appl. Probab.* **14** 181-189.

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