Math Meth Oper Res (2009) 70:567–596
DOI 10.1007/s00186-009-0284-7

ORIGINAL ARTICLE

Optimal control of Markovian jump processes with partial information and applications to a parallel queueing model

Ulrich Rieder · Jens Winter

Received: 28 October 2008 / Accepted: 27 January 2009 / Published online: 19 February 2009

Abstract We consider a stochastic control problem over an infinite horizon where the state process is influenced by an unobservable environment process. In particular, the Hidden-Markov-model and the Bayesian model are included. This model under partial information is transformed into an equivalent one with complete information by using the well-known filter technique. In particular, the optimal controls and the value functions of the original and the transformed problem are the same. An explicit representation of the filter process which is a piecewise-deterministic process, is also given. Then we propose two solution techniques for the transformed model. First, a generalized verification technique (with a generalized Hamilton–Jacobi–Bellman equation) is formulated where the strict differentiability of the value function is weaken to local Lipschitz continuity. Second, we present a discrete-time Markovian decision model by which we are able to compute an optimal control of our given problem. In this context we are also able to state a general existence result for optimal controls. The power of both solution techniques is finally demonstrated for a parallel queueing model with unknown service rates. In particular, the filter process is discussed in detail, the value function is explicitly computed and the optimal control is completely characterized in the symmetric case.

Keywords Stochastic control problem with partial information · Markovian jump process · Filter process · Generalized Hamilton–Jacobi–Bellman equation · MDP · Parallel queueing

U. Rieder (✉) · J. Winter
Institute of Optimization and Operations Research,
Universität Ulm, Helmholtzstr. 18, 89069 Ulm, Germany
e-mail: ulrich.rieder@uni-ulm.de

J. Winter
e-mail: jens.winter@uni-ulm.de

1 Introduction

Technological advances, especially in the information technology sector, over the last few years led to a more complex relationship between different systems. To understand the dependencies between systems requires large amount of resources and is hence expensive. Thus very often decisions (in optimization problems) are made on a lack of information. Practitioners deal with this lack of information mostly by their experience. They estimate the unknown parameters somehow and apply a reasonable strategy in order to minimize some cost functionals. But in general, they do not know if their suggested control is optimal. Very often they do not even know how well their policy works compared to the optimal one. In the last few years several optimization problems with incomplete information were studied mathematically and explicit solutions were obtained for some of them, in particular in finance applications (Bäuerle and Rieder 2007).

In this paper, we consider an optimization problem where the expected discounted cost over an infinite horizon should be minimized. The state process which is a Markovian jump process, is influenced by an unobservable environment process and hence the optimization problem has incomplete information (Sect. 2). Hidden-Markov-models and Bayesian models are special cases (Elliott et al. 1997). By using the well-known filter technique (Brémaud 1981; Liptser and Shiryayev 2004), we are able to define an equivalent stochastic control problem under complete information (Sect. 3). For this new model we discuss in Sect. 4 two solution procedures. The first one is an extension of the classical verification technique (with the Hamilton–Jacobi–Bellman equation). In the classical approach the strict differentiability of the value function is needed, which is a very strong condition. With Clarke's gradient (Clarke 1983) we weaken this assumption to local Lipschitz continuity which is fulfilled for our value function, since it is concave. The second approach uses the piecewise-deterministic behaviour of the filter process. The decision time points can be reduced to the jump points and then a control function up to the next time has to be chosen. A discrete-time Markovian Decision Problem (MDP) can be formulated. Solving this discrete-time MDP, we are able to compute an optimal control of the given optimization problem. Moreover, we are able to state a general existence result for optimal controls. Finally, we illustrate in Sect. 5 the power of both procedures for a parallel queueing model. In particular, we compute the value function and characterize the optimal control if the service rates are symmetric. Some interesting properties, e.g. the certainty equivalence principle and the stay-on-the-winner property, are also shown. Further investigations on this topic can be found in Winter (2008).

2 The model

On a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we consider a stochastic process (X_t, Z_t) characterized in (1) and (2). Denote by $\mathcal{F}_t^{X,Z} := \sigma(X_s, Z_s, s \le t)$, $\mathcal{F}_t^X := \sigma(X_s, s \le t)$ and $\mathcal{F}_t^Z := \sigma(Z_s, s \le t)$ the corresponding right-continuous and complete σ-algebras. (Z_t) takes values in a finite state space $S_Z := \{e_1, \dots, e_m\}$ where e_k is the kth unit vector of $\mathbb{R}^m$. Assuming the intensity for a jump from e_k into e_l is given by q_{kl}^Z we get the following semi-martingale representation (Brémaud 1981; Rogers and Williams 2003)

$$dZ_t = Q^Z Z_t \, dt + dM_t^Z \tag{1}$$

where (Q^Z) is the transpose of the generator matrix of (Z_t) and M_t^Z is a $\mathcal{F}_t^Z$-martingale. We will call the process (Z_t) the *environment process*. The so-called *state process* (X_t) takes values in the finite set $S_X := \{e_1, \ldots, e_n\}$. $X_t := (X_t^1, \ldots, X_t^n)$ depends on Z_t, in particular the intensity for a jump of (X_t) from e_i into e_j is given by $q_{ij}^X(k)$ if $Z_t = e_k$. Note that we do not exclude common jumps of X and Z.

Similar to (1) we use the following representation

$$dX_t = Q^X(Z_t) X_t \, dt + dM_t^{X,Z} \tag{2}$$

with a $\mathcal{F}_t^{X,Z}$-martingale $M_t^{X,Z}$. The state process X_t is observable, whereas the environment process Z_t is not observable. Hence the information given at time t is modelled by $\mathcal{F}_t^X$.

Remark 1 In Winter (2008) we assumed that only groups of states of (X_t) are observable and introduced a so-called information structure. An information structure is a partition of S_X. Then we are able to define an observation process (Y_t) and to generalize the construction above. Such models arise for example in inventory models (Bensoussan et al. 2003). Moreover, we discuss several properties of an information structure in view of the following optimization problem.

Assume now that $q_{ij}^X(u, k)$ depends on a *control parameter* $u \in U$ with $U \subset \mathbb{R}^d$ $(d \in \mathbb{N})$ compact and convex such that $q_{ij}^X(u, k)$ is continuous in u. We call a control process $u = (u_t)$ with $u_t : [0, \infty) \to U$ *admissible* if

$$\text{(A)} \begin{cases} (u_t) \text{ is a càdlàg process} \\ u_t \text{ is } \mathcal{F}_t^X\text{-predictable for all } t \geq 0 \\ u_t \in U \text{ for all } t \geq 0. \end{cases}$$

We define the set of admissible controls by

$$\mathcal{U} := \{u = (u_t) \mid u \text{ satisfies (A)}\} .$$

Each control process $u = (u_t)$ determines a state process (X_t^u). In the following we omit the dependency of X_t on u if it is obvious from the context. Since the control process has not be Markovian, (X_t) is not a Markov process. But since the intensity $q_{ij}^X(u, k)$ for the next jump depends only on the current state and the current control, (X_t) is called *Markovian Jump Process*. We assume that the initial distribution of Z_0 is given by μ, i.e. $Z_0 \sim \mu$, and the initial state X_0 is given by x_0. Both, μ and x_0 are given and fixed. Then we are able to introduce the following optimization problem:

$$(P) \begin{cases} \mathbb{E}\left[\int_0^\infty e^{-\beta t} c(X_t, Z_t, u_t)dt\right] \to \min \\ dZ_t = Q^Z Z_t + dM_t^Z,\, Z_0 \sim \mu \\ dX_t = Q^X(u_t, Z_t)X_t dt + dM_t^{X,Z},\, X_0 = x_0 \\ u \in \mathcal{U} \end{cases}$$

where $\beta > 0$ is the discount rate and $c : S_X \times S_Z \times U \to [0, \infty)$ is a given cost function, continuous in u, such that

$$\mathbb{E}\left[\int_0^\infty e^{-\beta t} c(X_t, Z_t, u_t)dt\right] < \infty.$$

Since the environment process (Z_t) is not observable, problem (P) is a stochastic control problem with partial information and hence not solvable directly. We define in the next section an equivalent stochastic control problem with complete information.

3 The transformation

Define the conditional probability $p_t^k := \mathbb{P}(Z_t = e_k \mid \mathcal{F}_t^X)$ with $p_0^k := \mu_k, k = 1, \ldots, m$, and $p_t := (p_t^1, \ldots, p_t^m) \in \triangle^m$. Note that p_t depends on the control process u through the state process $X = (X_t^u)$. It is well-known (Brémaud 1981) that the $\mathcal{F}_t^X$-intensity of (X_t) is given by

$$q_{ij}^X(u, p) := \sum_{k=1}^m q_{ij}^X(u, k) p_k$$

and thus we find a semi-martingale representation for (X_t) with respect to $\mathcal{F}_t^X$ as

$$dX_t = Q^X(u_t, p_t)dt + dM_t^X.$$

Theorem 1 *Let $u = (u_t) \in \mathcal{U}$. Then, p_t is the unique solution of the filter equation*

$$dp_t = b(u_t, X_t, p_t)dt + \sum_{i=1}^n \sum_{j=1}^n \Phi_{ij}(u_t, p_{t-}) X_{t-}^i dN_t^X(i, j), \quad p_0 = \mu \tag{3}$$

with

$$b(u, x, p) := \left(Q^Z - \sum_{i=1}^n \sum_{j=1}^n \Phi_{ij}(u, p) x_i q_{ij}^X(u, p) \right) p$$

$$\Phi_{ij}(u, p) := \frac{1}{q_{ij}^X(u, p)} \begin{pmatrix} q_{ij}^X(u, 1)p_1 \\ \vdots \\ q_{ij}^X(u, m)p_m \end{pmatrix} - p.$$

$N_t^X(i, j)$ counts the jumps of X from e_i to e_j up to time t.

Proof First, we know by Brémaud (1981) that $dp_t = Q^Z p_t dt + dM_t$, where M_t is a $\mathcal{F}_t^X$-martingale. Hence it admits an unique representation $dM_t = \phi_t dM_t^X$. Since X is completely characterized by $N_t^X(i,j)$ we compute $Z_t N_t^X(i,j)$ and $p_t N_t^X(i,j)$ with the help of Itô's formula and the quadratic covariation

$$[Z, X](t) = \sum_{0<s\le t} \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_{ij}^{kl} (e_j - e_i)(e_l - e_k) \triangle N_s^Z(k,l) X_{s-}^i$$

where $\delta_{ij}^{kl} = 1$ if a jump of Z from e_k into e_l induces a jump of X from e_i into e_j and $N_t^Z(k,l)$ counts the jumps of Z from e_k into e_l up to time t. Since the expectation of these both expression has to be equal we get Φ by comparison of coefficients. □

Again, the filter process (p_t) depends on the control process (u_t) through the intensities $q_{ij}^X(u,k)$.

Define

$$(\tilde{P}) \begin{cases} \mathbb{E}\left[\int_0^\infty e^{-\beta t} c(X_t, p_t, u_t) dt\right] \to \min \\ dp_t = b(u_t, X_t, p_t) dt + \sum_{i=1}^{n} \sum_{j=1}^{n} \Phi_{ij}(u_t, p_{t-}) X_{t-}^i dN_t^X(i,j),\ p_0 = \mu \\ dX_t = Q^X(u_t, p_t) X_t dt + dM_t^X,\ X_0 = x_0 \\ u \in \mathcal{U} \end{cases}$$

with $c(x,p,u) := \sum_{k=1}^m c(x, e_k, u) p_k$. All components in $(\tilde{P})$ are measurable with respect to $(\mathcal{F}_t^X)$, hence $(\tilde{P})$ is a stochastic control problem with complete information. The process (X_t, p_t) is piecewise-deterministic, and its state space is given by $S_X \times \triangle^m$. We denote by $J(x,p;u)$ the expected discounted cost under a control u if $X_0 = x$ and $p_0 = p$, and by

$$J(x,p) := \inf_{u\in\mathcal{U}} J(x,p;u), \quad (x,p) \in S_X \times \triangle^m,$$

the minimal cost function. The connection between (P) and $(\tilde{P})$ is given in the next theorem.

Theorem 2 (a) *It holds:* $\mathbb{E}\left[\int_0^\infty e^{-\beta t} c(X_t, Z_t, u_t) dt\right] = \mathbb{E}\left[\int_0^\infty e^{-\beta t} c(X_t, p_t, u_t) dt\right]$ *for all* $u \in \mathcal{U}$.

(b) *The optimal values and the optimal controls of* (P) *and* $(\tilde{P})$ *are equal, in particular*

$$value(P) = J(x_0, \mu).$$

Proof (a) Let $u = (u_t) \in \mathcal{U}$. It is sufficient to prove that

$$\mathbb{E}[c(X_t, Z_t, u_t)] = \mathbb{E}[c(X_t, p_t, u_t)] \quad \forall t \ge 0.$$

From the definition we get immediately

$$\mathbb{E}\left[c(X_t, Z_t, u_t) \mid \mathcal{F}_t^X\right] = c(X_t, p_t, u_t)$$

and hence the statement.

(b) follows from (a). □

The following will be useful in Sect. 4.

Proposition 1 *The value function $p \mapsto J(x, p)$ is concave, $x \in S_X$.*

Proof It is clear that $J(x, p; u) = \sum_{k=1}^{m} J(x, e_k; u) p_k$; hence $p \mapsto J(x, p; u)$ is linear. Then we conclude:

$$\begin{aligned} J(x, \lambda p + (1-\lambda) p) &= \inf_{u \in \mathcal{U}} \{\lambda J(x, p; u) + (1-\lambda) J(x, q; u)\} \\ &\geq \lambda J(x, p) + (1-\lambda) J(x, q). \end{aligned}$$

□

Note that $p \mapsto J(x, p)$ is locally Lipschitz continuous by Proposition 1.

4 Two solution procedures

In this section we present two procedures for solving the stochastic control problem $(\tilde{P})$. The power of these both procedures is illustrated in Sect. 5 for a queueing model.

4.1 Solution via a generalized Hamilton–Jacobi–Bellman equation

For a locally Lipschitz continuous function $f : \mathbb{R}^m \to \mathbb{R}$ define the *upper derivative* at x (in direction y) by

$$f^0(x; y) := \limsup_{\substack{z \to x \\ \varepsilon \to 0}} \frac{f(z + \varepsilon y) - f(z)}{\varepsilon}.$$

Replacing lim sup by lim inf, the *lower derivative* $f_0(x; y)$ is defined. The *Clarke gradient* of f at x is given by

$$\partial f(x) := \left\{\xi \in \mathbb{R}^d \mid f^0(x; y) \geq \xi y \text{ for all } y \in \mathbb{R}^d\right\},$$

which is a nonempty, convex and compact subset of $\mathbb{R}^d$. We want to understand ξ as a row vector. If $f(x)$ is differentiable at x with gradient $\nabla f(x)$ then $\partial f(x) = \{\nabla f(x)\}$. It holds

$$f^0(x; y) = \max_{\xi \in \partial f(x)} \xi y \quad \text{and} \quad f_0(x; y) = \min_{\xi \in \partial f(x)} \xi y. \tag{4}$$

Using the local Lipschitz continuity we conclude that f is differentiable almost everywhere and we can find for every $x \in \mathbb{R}^d$ a sequence (x_n) with $x_n \in \mathbb{R}^d$ such that x_n converges to x and f is differentiable at x_n for all $n \in \mathbb{N}$. Hence $\partial f(x)$ can be written as the closed convex hull of existing limits of sequences of the derivatives $\nabla f(x_n)$, i.e.

$$\partial f(x) = co\left\{\lim_{n\to\infty} \nabla f(x_n) \mid \lim_{n\to\infty} x_n = x\right\}.$$

A locally Lipschitz continuous function f is called *regular* at x if the ordinary directional derivative

$$f'(x; y) := \lim_{\varepsilon\to 0} \frac{f(x+\varepsilon y) - f(x)}{\varepsilon}$$

exists for all y and $f^0(x; y) = f'(x; y)$. Every concave function f (which is even locally Lipschitz) is regular (see Clarke (1983) for more details).

The Hamiltonian is defined by

$$\begin{aligned} &Hv(x, p, u, \xi) \\ &\quad := c(x, p, u) + \xi b(u, x, p) + \sum_{j=1}^{n} \left(v(e_j, p + \Phi_{xj}(u, p)) - v(x, p)\right) q_{xj}^X(u, p) \end{aligned}$$

for $(x, p) \in S_X \times \triangle^m$, $u \in U$ and $\xi \in \mathbb{R}^m$. In what follows, $\partial_p v(x, p)$ is the Clarke gradient of v with respect to p.

Next we prove that the value function of $(\tilde{P})$ is a solution of the *generalized Hamilton–Jacobi–Bellman (HJB) equation*:

$$\beta v(x, p) = \inf_{\substack{\xi\in\partial_p v(x,p)\\ u\in U}} Hv(x, p, u, \xi), \quad (x, p) \in S_X \times \triangle^m. \tag{5}$$

Theorem 3 *The value function $J(x, p)$ is a solution of the generalized HJB-equation* (5).

Proof Denote by T_n the jump times of X. Note that T_n is also a jump time of (p_t). Due to the local Lipschitz continuity there exists for all $0 =: T_0 < T_1 < T_2 < \cdots$ a function $D(e^{-\beta s} J(X_s, p_s))$ such that

$$e^{-\beta T_i -} J(X_{T_i -}, p_{T_i -}) - e^{-\beta T_{i-1}} J(X_{T_{i-1}}, p_{T_{i-1}}) = \int_{T_{i-1}}^{T_i -} D\left(e^{-\beta s} J(X_s, p_s)\right) ds.$$

The function $D\left(e^{-\beta s}J(X_s, p_s)\right)$ may be chosen as the derivative of $e^{-\beta s}J(X_s, p_s)$ with respect to s, which exists almost everywhere on $[0, \infty)$. Hence

$$D\left(e^{-\beta s}J(X_s, p_s)\right) = e^{-\beta s}\left(-\beta J(X_s, p_s) + J_p(X_s, p_s)b(u_s, X_s, p_s)\right).$$

Extending this consideration over the jump time points T_n we can write

$$\begin{aligned} &e^{-\beta t}J(X_t, p_t) \\ &= J(X_0, p_0) + \int_0^t D\left(e^{-\beta s}J(X_s, p_s)\right) ds \\ &\quad + \int_0^t e^{-\beta s}\sum_{j=1}^n \left(J(e_j, p_{s-} + \Phi_{X_{s-}j}(u_s, p_{s-})) - J(X_{s-}, p_{s-})\right) dN_s^X(X_{s-}, e_j). \end{aligned}$$

Denote by τ the first jump time point after time t then we obtain from the Bellman equation for all $u \in \mathcal{U}[t, \tau)$ and $\hat{t} > t$

$$\begin{aligned} &e^{-\beta t}J(x, p) \\ &\le \mathbb{E}\left[\int_t^{\tau\wedge\hat{t}} e^{-\beta s}c(X_s, p_s, u_s)ds + e^{-\beta(\tau\wedge\hat{t})}J(X_{\tau\wedge\hat{t}}, p_{\tau\wedge\hat{t}}) \mid X_t = x, p_t = p\right] \\ &= \mathbb{E}\Bigg[\int_t^{\tau\wedge\hat{t}} e^{-\beta s}c(X_s, p_s, u_s)ds + e^{-\beta t}J(X_t, p_t) + \int_t^{\tau\wedge\hat{t}} D\left(e^{-\beta s}J(X_s, p_s)\right) ds \\ &\quad + \int_t^{\tau\wedge\hat{t}} e^{-\beta s}\sum_{j=1}^n \left(J\left(e_j, p_{s-} + \Phi_{X_{s-}j}(u_s, p_{s-})\right) - J(X_{s-}, p_{s-})\right) \\ &\quad \cdot dN_s^X(X_{s-}, e_j) \mid X_t = x, p_t = p\Bigg]. \end{aligned}$$

Hence

$$0 \le \mathbb{E}\left[\int_t^{\tau\wedge\hat{t}} e^{-\beta s}\tilde{H}J(X_s, p_s, u_s, J_p)ds \mid X_t = x, p_t = p\right]$$

where $\tilde{H}J(x, p, u, \xi) := HJ(x, p, u, \xi) - \beta J(x, p)$. Using $\tilde{u}_s \equiv u$ for $s \in [t, t+\varepsilon)$, $\varepsilon > 0$ we conclude

$$
\begin{aligned}
0 &\le \lim_{\hat{t}\downarrow t} \mathbb{E}\left[\frac{1}{\hat{t}-t}\int_t^{\tau\wedge\hat{t}} e^{-\beta s}\tilde{H}J(X_s, p_s, \tilde{u}_s, J_p)ds \mid X_t = x, p_t = p\right] \\
&= \lim_{\hat{t}\downarrow t} \mathbb{E}\left[\frac{1}{\hat{t}-t}\int_t^{\hat{t}} e^{-\beta s}\tilde{H}J(X_s, p_s, \tilde{u}_s, J_p)ds \mid X_t = x, p_t = p\right] \cdot \mathbb{P}(\hat{t} < \tau) \\
&\quad + \lim_{\hat{t}\downarrow t} \mathbb{E}\left[\frac{1}{\hat{t}-t}\int_t^{\tau} e^{-\beta s}\tilde{H}J(X_s, p_s, \tilde{u}_s, J_p)ds \mid X_t = x, p_t = p\right] \cdot \mathbb{P}(\hat{t} \ge \tau).
\end{aligned}
$$

For $\alpha \ge \sup_{k,i} \sup_{u\in U} \sum_{j\neq i} q_{ij}^X(u, k)$ we get

$$
\mathbb{P}(\hat{t} \ge \tau) \le 1 - e^{-\alpha(\hat{t}-t)} \to 0 \quad \text{for } \hat{t} \downarrow t.
$$

Thus we obtain at points p where $J(x, p)$ is differentiable (with respect to p) that

$$
0 \le e^{-\beta t}\tilde{H}J(X_t, p_t, \tilde{u}_t, J_p) = e^{-\beta t}\tilde{H}J(x, p, u, J_p),
$$

hence

$$
0 \le \inf_{u\in U} \tilde{H}J(x, p, u, J_p).
$$

At points p where $J(x, p)$ is not differentiable at p the generalized gradient can be written as

$$
\begin{aligned}
\partial_p J(x, p) &= co\left\{\lim_{n\to\infty} \nabla J(x, p_{t_n}) \mid p_{t_n} \to p_t = p\right\} \\
&= co\left\{\lim_{n\to\infty} \nabla J(x, p_{t_n}) \mid t_n \to t\right\}.
\end{aligned}
$$

In particular, every $\xi \in \partial_p J(x, p)$ is a convex combination of $\xi^m = \lim_{n\to\infty} \nabla J(x, p_{t_n^m})$ for sequences $t_n^m \to t$, along which $J(x, p)$ is differentiable. Since $p \mapsto J(x, p)$ is locally Lipschitz continuous we obtain with the chain rule that

$$
\begin{aligned}
0 \le e^{-\beta t}\Bigg(& c(x, p, u) - \beta J(x, p) + \xi^m b(u, x, p) \\
& + \sum_{j=1}^{n} \left(J\left(e_j, p + \Phi_{xj}(u, p)\right) - J(x, p)\right) q_{xj}^X(u, p)\Bigg).
\end{aligned}
$$

Dividing by $e^{-\beta t}$ and remembering that ξ is a convex combination of ξ^m we conclude

$$\begin{aligned} 0 \le\, & c(x,p,u) - \beta J(x,p) + \xi b(u,x,p) \\ & + \sum_{j=1}^{n} \big(J(e_j, p + \Phi_{xj}(u,p)) - J(x,p) \big) q_{xj}^X(u,p) \end{aligned}$$

and therefore we get

$$0 \le \inf_{\substack{\xi \in \partial_p J(x,p) \\ u \in U}} \tilde{H} J(x,p,u,\xi).$$

On the other hand for $\varepsilon > 0$ and $0 < t < \hat{t} < \infty$ with $t - \hat{t} > 0$ small enough there exists a strategy u^ε with corresponding state process (X_t, p_t) such that

$$\begin{aligned} & e^{-\beta t} J(x,p) + \varepsilon(\hat{t} - t) \\ & \ge \mathbb{E}\left[\int_t^{\tau \wedge \hat{t}} e^{-\beta s} c(X_s, p_s, u_s^\varepsilon) ds + e^{-\beta(\tau \wedge \hat{t})} J(X_{\tau \wedge \hat{t}}, p_{\tau \wedge \hat{t}}) \mid X_t = x, p_t = p \right]. \end{aligned}$$

As before we obtain

$$\begin{aligned} \varepsilon & \ge \mathbb{E}\left[\frac{1}{\hat{t} - t} \int_t^{\tau \wedge \hat{t}} e^{-\beta s} \tilde{H} J(X_s, p_s, u_s^\varepsilon, \xi) ds \mid X_t = x, p_t = p \right] \\ & \ge \mathbb{E}\left[\frac{1}{\hat{t} - t} \int_t^{\tau \wedge \hat{t}} e^{-\beta s} \inf_{u_s \in U} \tilde{H} J(X_s, p_s, u_s, \xi) ds \mid X_t = x, p_t = p \right] \\ & = \mathbb{E}\left[\frac{1}{\hat{t} - t} \int_t^{\tau \wedge \hat{t}} e^{-\beta s} \tilde{H} J(X_s, p_s, u_s^*, \xi) ds \mid X_t = x, p_t = p \right] \end{aligned}$$

where the existence of u_s^* is guaranteed since $u \mapsto HJ(x,p,u,\xi)$ is continuous and U compact. If $J(x,p)$ is differentiable at p we get as above

$$\begin{aligned} \varepsilon & \ge \lim_{\hat{t} \downarrow t} \mathbb{E}\left[\frac{1}{\hat{t} - t} \int_t^{\tau \wedge \hat{t}} e^{-\beta s} \tilde{H} J(X_s, p_s, u_s^*, J_p) ds \mid X_t = x, p_t = p \right] \\ & = e^{-\beta t} \tilde{H} J(x,p,u_t^*, J_p) = e^{-\beta t} \inf_{u \in U} \tilde{H} J(x,p,u,J_p) \end{aligned}$$

and finally, since ε is arbitrarily,

$$0 \geq \inf_{u \in U} \tilde{H} J(x, p, u, J_p).$$

If $J(x, p)$ is not differentiable at p we get that every $\xi \in \partial_p J(x, p)$ is a convex combination of ξ^m. With the same computations and with the help of the approximating sequence t_n^m as above it follows

$$0 \geq \inf_{\substack{\xi \in \partial_p J(x,p) \\ u \in U}} \tilde{H} J(x, p, u, \xi).$$

Altogether we have shown that

$$\beta J(x, p) = \inf_{\substack{\xi \in \partial_p J(x,p) \\ u \in U}} H J(x, p, u, \xi), \quad (x, p) \in S_X \times \triangle^m,$$

which proves the statement. □

Now we are in a position to formulate a generalized *verification technique* for computing optimal controls.

Theorem 4 *If there exists $u^* = (u_t^*) \in \mathcal{U}$ with corresponding state process (X_t^*, p_t^*) such that for almost all $t \geq 0$ a process $\xi_t^* \in \partial_p J(X_t^*, p_t^*)$ exists such that*

$$\beta J(X_t^*, p_t^*) = H J(X_t^*, p_t^*, u_t^*, \xi_t^*)$$

then $u^* = (u_t^*)$ is an optimal control for $(\tilde{P})$, and moreover, $J(x, p)$ is the unique locally Lipschitz and regular (in p) solution of the HJB-equation (5).

Proof Due to the concavity of $p \mapsto J(x, p)$, $J(x, p)$ is locally Lipschitz and regular. Then we get:

$$\begin{aligned}
\frac{\partial}{\partial t} J(X_t^*, p_t^*) &= \lim_{\varepsilon \to 0} \frac{J(X_t^*, p_{t-\varepsilon}^*) - J(X_t^*, p_t^*)}{-\varepsilon} \\
&= \lim_{\varepsilon \to 0} \frac{J(X_t^*, p_t^* - \varepsilon \cdot b(u_t^*, X_t^*, p_t^*)) - J(X_t^*, p_t^*)}{-\varepsilon} \\
&= -J^0(X_t^*, p_t^*; -b(u_t^*, X_t^*, p_t^*)) \\
&\leq -\xi_t \left(-b(u_t^*, X_t^*, p_t^*)\right) \quad \forall \xi_t \in \partial_p J(X_t, p_t) \\
&= \xi_t b(u_t^*, X_t^*, p_t^*)
\end{aligned}$$

where the third equality is true due to the regularity and the inequality due to the properties of Clarke derivatives. In particular, for ξ_t^* it holds that

$$\begin{aligned}\frac{\partial}{\partial t} J(X_t^*, p_t^*) &\leq \xi_t^* b(u_t^*, X_t^*, p_t^*)\\ &= \beta J(X_t^*, p_t^*) - c(X_t^*, p_t^*, u_t^*)\\ &\quad - \sum_{j=1}^{n} \Big(J(e_j, p_{t-}^* + \Phi_{X_{t-}^* j}(u_t^*, p_{t-}^*)) - J(X_{t-}^*, p_{t-}^*)\Big)\, q_{X_{t-}^* j}^X(u_t^*, p_{t-}^*)\end{aligned}$$

and

$$\frac{\partial}{\partial t}\left(e^{-\beta t} J(X_t^*, p_t^*)\right) = e^{-\beta t}\left(-\beta J(X_t^*, p_t^*) + \xi_t^* b(u_t^*, X_t^*, p_t^*)\right),$$

and by integration from 0 to t we get

$$\begin{aligned}e^{-\beta t} J(X_t^*, p_t^*) \leq J(x,p) &- \int_0^t e^{-\beta s} c(X_s^*, p_s^*, u_s^*)ds\\ &+ \int_0^t e^{-\beta s} \sum_{j=1}^{n} \Big(J(e_j, p_{s-}^* + \Phi_{X_{s-}^* j}(u_s^*, p_{s-}^*)) - J(X_{s-}^*, p_{s-}^*)\Big)\\ &\qquad \cdot \Big(dN_s^X(X_{s-}^*, e_j) - q_{X_{s-}^* j}^X(u_s^*, p_{s-}^*)ds\Big).\end{aligned}$$

Letting $t \to \infty$ then $e^{-\beta t} J(X_t^*, p_t^*) \to 0$ and taking expectation we obtain

$$J(x,p) \geq \mathbb{E}\left[\int_0^\infty e^{-\beta t} c(X_t^*, p_t^*, u_t^*)dt\right].$$

This proves the optimality of u^*. Next we have to prove the uniqueness. Let $\tilde{J}(x,p)$ be another locally Lipschitz and regular solution of the HJB-equation. By interchanging $\tilde{J}(x,p)$ and $J(x,p)$ in the proof above we conclude, that $\tilde{J}(x,p) \geq J(x,p)$. On the other hand we get as in the proof of Theorem 3 that

$$\begin{aligned}&e^{-\beta t}\tilde{J}(X_t, p_t)\\ &\quad = \tilde{J}(x,p) + \int_0^t D\Big(e^{-\beta s}\tilde{J}(X_s, p_s)\Big)ds\\ &\qquad + \int_0^t e^{-\beta s}\sum_{j=1}^{n}\Big(\tilde{J}(e_j, p_{s-} + \Phi_{X_{s-} j}(u_s, p_{s-})) - \tilde{J}(X_{s-}, p_{s-})\Big)\, dN_s^X(X_{s-}, e_j).\end{aligned}$$

At those points p where $\tilde{J}(x, p)$ is differentiable we know from the HJB-equation that

$$D\left(e^{-\beta s}\tilde{J}(X_s, p_s)\right) \geq e^{-\beta s}\Bigg(-c\,(X_s, p_s, u_s) - \sum_{j=1}^{n}\left(\tilde{J}\left(e_j, p_{s-} + \Phi_{X_{s-}j}(u_s, p_{s-})\right) - \tilde{J}(X_{s-}, p_{s-})\right) q^X_{X_{s-}j}(u_s, p_{s-})\Bigg)$$

and since $\tilde{J}(x, p)$ is differentiable almost everywhere we conclude by integration from 0 to t that

$$e^{-\beta t}\tilde{J}(X_t, p_t) \geq \tilde{J}(x, p) - \int_0^t e^{-\beta s} c(X_s, p_s, u_s) ds + \int_0^t e^{-\beta s}\sum_{j=1}^{n}\left(\tilde{J}\left(e_j, p_{s-} + \Phi_{X_{s-}j}(u_s, p_{s-})\right) - \tilde{J}(X_{s-}, p_{s-})\right) \cdot \left(dN^X_s(X_{s-}, e_j) - q^X_{X_{s-}j}(u_s, p_{s-})ds\right).$$

Letting $t \to \infty$ and taking expectation we finally get

$$\tilde{J}(x, p) \leq \mathbb{E}\left[\int_0^\infty e^{-\beta s} c(X_s, p_s, u_s) ds\right],$$

since the second integral is a martingale. This inequality holds for all $u \in \mathcal{U}$, hence $\tilde{J}(x, p) \leq J(x, p)$. □

4.2 Solution via a discrete-time Markov decision problem

Since (X_t, p_t) is a piecewise-deterministic process, we are able to solve the stochastic control problem $(\tilde{P})$ by a discrete-time Markov Decision Problem (MDP). Based on the post-jump state an action $a = (a_t)$ with $a : [0, \infty) \to U$ has to be chosen (depending on the time t elapsed since the last jump).

Define

$$\alpha := \max_{i,k}\max_{u\in U}\sum_{j\neq i} q^X_{ij}(u, k).$$

Then the *uniformized* state process (X_t) is characterized by its jump times T_n with $T_{n+1} - T_n \sim \exp(\alpha)$ and its (uniformized) transition probability

$$p_{ij}^X(u,k) = \begin{cases} \frac{q_{ij}^X(u,k)}{\alpha} & i \neq j \\ 1 - \frac{q_i^X(u,k)}{\alpha} & i = j \end{cases}$$

with $q_i^X(u,k) := \sum_{j \neq i} q_{ij}^X(u,k)$. The process (p_t) jumps if and only if (X_t) jumps and then its jump size is determined by $\Phi_{ij}(u, p)$. Between two jumps under an action $a = (a_t)$ the process (p_t) evolves according to the (deterministic) differential equation $dp_t = b(a_t, x, p_t)dt$, $X_{T_n} = x$. For $t \in [T_n, T_{n+1})$ and $X_{T_n} = x$ it holds that $p_t = \phi_{t-T_n}^a(x, p_{T_n})$ where $\phi_t^a(x, p)$ is the unique solution of

$$dp_t = b(a_t, x, p_t)dt, \quad p_0 = p.$$

The (uniformized) *discrete-time MDP* (S, A, D, q, r, δ) is then defined by

$$\begin{aligned} S &:= S_X \times \triangle^m \\ A &:= \{a : [0, \infty) \to U \mid \text{measurable}\} \\ q\left((x, p), a, (y, B)\right) &:= \int_0^\infty e^{-\alpha t} q_{xy}^X \left(a_t, \phi_t^a(x, p)\right) \mathbb{1}\left\{\phi_t^a(x, p)\right. \\ &\qquad \left. + \Phi_{xy}\left(a_t, \phi_t^a(x, p)\right) \in B\right\} dt, x \neq y \\ q\left((x, p), a, (x, B)\right) &:= 1 - \sum_{y \neq x} q\left((x, p), a, (y, B)\right) \\ \overline{c}(x, p, a) &:= \int_0^\infty \alpha e^{-\alpha t} \left(\int_0^t e^{-\beta s} c\left(x, \phi_s^a(x, p), a_s\right) ds \right) dt \\ \delta &:= \frac{\alpha}{\alpha + \beta} < 1. \end{aligned}$$

A sequence $\pi = (f_n) \in F^\infty$ where $f_n \in F := \{f : S \to A \text{ measurable}\}$ is called a *(Markov) strategy*. The expected cost of such a strategy $\pi = (f_0, f_1, \ldots)$ is defined by

$$V_\pi(x, p) := \mathbb{E}_\pi \left[\sum_{n=0}^\infty \delta^n \, \overline{c}\left(X_{T_n}, p_{T_n}, f_n\left(X_{T_n}, p_{T_n}\right)\right) \mid X_0 = x, p_0 = p \right]$$

where the expectation is taken with respect to $\mathbb{P}_\pi$ which is defined by the transition probabilities q and a strategy π. The expectation is well-defined. The value function of the discrete-time MDP is denoted by

$$V(x, p) := \inf_{\pi \in F^\infty} V_\pi(x, p), \quad (x, p) \in S.$$

Define for $v \in \mathbb{B} := \{v : S \to \mathbb{R}_+ \mid \text{measurable}\}$ the well-known operators

$$
\begin{aligned}
(Lv)(x,p,a) &:= \bar{c}(x,p,a) + \delta \int_{\triangle^m} \sum_{y \in S_X} v(y,\rho) q\left((x,p),a,(y,d\rho)\right) \\
&= \int_0^\infty e^{-(\alpha+\beta)t} \Bigg\{ c\big(x,\phi_t^a(x,p),a_t\big) + v\big(x,\phi_t^a(x,p)\big)\Big(\alpha - q_x^X\left(a_t,\phi_t^a(x,p)\right)\Big) \\
&\quad + \sum_{y \neq x} v\left(y,\phi_t^a(x,p) + \Phi_{xy}\left(a_t,\phi_t^a(x,p)\right)\right) q_{xy}^X\left(a_t,\phi_t^a(x,p)\right) \Bigg\} dt \\
(\mathcal{T}v)(x,p) &:= \inf_{a \in A} (Lv)(x,p,a),\ (x,p) \in S.
\end{aligned}
$$

The relation between $(\tilde{P})$ and the introduced MDP is shown in the following theorem.

Theorem 5 (a) $J(x,p) = V(x,p)$.
(b) *If* $\pi^* = (f_n) \in F^\infty$ *is optimal for the MDP, then* $u^* = (u_t^*) \in \mathcal{U}$ *with*

$$u_t^* := f_n(X_{T_n}^*, p_{T_n}^*)(t - T_n) \quad \textit{for } t \in [T_n, T_{n+1})$$

is optimal for $(\tilde{P})$ *and hence also for* (P).

Proof (a) Let $u = (u_t)$ be a Markovian control, i.e. there exists a strategy $\pi = (f_n)$ with $f_n \in F$ such that

$$u_t = f_n(X_{T_n}, p_{T_n})(t - T_n) \quad \text{for } t \in [T_n, T_{n+1}).$$

Then we obtain

$$
\begin{aligned}
J(x,p;u) &= \mathbb{E}\left[\int_0^\infty e^{-\beta t} c\left(X_t,p_t,u_t\right) dt\right] \\
&= \mathbb{E}\left[\sum_{n=0}^\infty \int_{T_n}^{T_{n+1}} e^{-\beta t} c\left(X_t,p_t,u_t\right) dt\right] \\
&= \sum_{n=0}^\infty \mathbb{E}\left[e^{-\beta T_n} \mathbb{E}\left\{\int_0^{T_{n+1}-T_n} e^{-\beta t} c\left(X_{t+T_n}, p_{t+T_n}, u_{t+T_n}\right) dt \mid X_{T_n}, p_{T_n}\right\}\right] \\
&= \sum_{n=0}^\infty \mathbb{E}_\pi\left[\prod_{k=1}^n e^{-\beta(T_k - T_{k-1})} \overline{c}\left(X_{T_n}, p_{T_n}, f_n(X_{T_n}, p_{T_n})\right)\right] \\
&= \sum_{n=0}^\infty \mathbb{E}_\pi\left[\delta^n \overline{c}(X_{T_n}, p_{T_n}, f_n(X_{T_n}, p_{T_n}))\right].
\end{aligned}
$$

Hence

$$J(x, p; u) = V_\pi(x, p). \tag{6}$$

Now, let $u = (u_t) \in \mathcal{U}$ be arbitrary. Then there exists a sequence (f_n) with $f_n : S^{n+1} \to A$ such that

$$u_t = f_n(X_{T_0}, p_{T_0}, \ldots, X_{T_n}, p_{T_n})(t - T_n) \quad \text{for } t \in [T_n, T_{n+1}).$$

Due to the Markovian structure of the state process, we obtain (see Bertsekas and Shreve 1978, p. 216)

$$J(x, p) = \inf_{u \text{ Markovian}} J(x, p; u) = V(x, p).$$

(b) follows from (a) and (6). □

To state an existence result for an optimal policy for the MDP (and hence for $(\tilde{P})$ and (P)) we have to extend deterministic controls to *relaxed controls* $r \in \mathcal{R}$ where

$$\mathcal{R} := \{r : [0, \infty) \to \mathbb{P}(U) \mid \text{measurable}\}.$$

For a relaxed control $r = (r_t) \in \mathcal{R}$ define

$$\overline{r}_t := \int_U u\, r_t(du) \in U.$$

Instead of choosing at each time point t a fixed control parameter $u \in U$ we randomize now over the set of possible values in U. The case of deterministic controls is always included by choosing the Dirac-measure on U. Because of the relaxation we have to consider our state process (X_t, p_t) now with respect to relaxed controls. We define for a relaxed control $r = (r_t)$ the corresponding state process (X_t^r, p_t^r) by its (relaxed) intensities

$$q_{ij}^X(r, k) := \int_U q_{ij}^X(u, k) r(du).$$

$\phi_t^r(x, p)$ is the unique solution of

$$dp_t = b(\overline{r}_t, x, p_t)dt, \quad p_0 = p.$$

Note that the drift component is not relaxed. Hence we conclude

$$\phi_t^r(x, p) = \phi_t^{\overline{r}}(x, p). \tag{7}$$

The jump times are again $\exp(\alpha)$-distributed and the jump size of (p_t) under a relaxed control is given by

$$\Phi_{ij}(r, p) := \int_U \Phi_{ij}(u, p) r(du).$$

The process (X_t^r, p_t^r) is well-defined (Davis 1993). Finally, the cost rate under a relaxed control r is given by

$$c(x, p, r) := \int_U c(x, p, u) r(du).$$

Proposition 2 (a) *$\mathcal{R}$ is compact.*
(b) *$(p, r) \mapsto \phi_t^r(x, p)$ is continuous on $\Delta^m \times \mathcal{R}$.*

Proof (a) Davis (1993, Proposition 43.3).
(b) Let $r_n \to r$ and $p_n \to p$ and let $\phi_t^{r_n}(x, p_n)$ and $\phi_t^r(x, p)$ be the unique solutions of

$$dp_t = b(\bar{r}_t^n, x, p_t)dt, \;\; p_0 = p_n \quad \text{and} \quad dp_t = b(\bar{r}_t, x, p_t)dt, \;\; p_0 = p.$$

Then

$$\begin{aligned}
&|\phi_t^{r_n}(x, p_n) - \phi_t^r(x, p)| \\
&\quad \le |p_n - p| + \int_0^t \left| b(\bar{r}_s^n, x, \phi_s^{r_n}) - b(\bar{r}_s^n, x, \phi_s^r) \right| ds \\
&\qquad + \int_0^t \left| b(\bar{r}_s^n, x, \phi_s^r) - b(\bar{r}_s, x, \phi_s^r) \right| ds \\
&\quad \le |p_n - p| + \int_0^t L|\phi_s^{r_n} - \phi_s^r| ds + \int_0^t \left| b(\bar{r}_s^n, x, \phi_s^r) - b(\bar{r}_s, x, \phi_s^r) \right| ds
\end{aligned}$$

where the last inequality is true due to the Lipschitz continuity of $p \mapsto b(r, x, p)$. Since $r_n \to r$ and $p_n \to p$ the statement follows if the second integral tends to 0. Remember that $b(r, x, p)$ is continuous in r, and $p \mapsto b(r, x, p) - b(\tilde{r}, x, p)$ is Lipschitz continuous on the compact set Δ^m and bilinear-quadratic. Therefore the maximum point $p^*(r)$ is a continuous function of r and we conclude that

$$|b(r, x, p) - b(\tilde{r}, x, p)| \le |h(r, x) - h(\tilde{r}, x)|$$

for a continuous function $h(\cdot, x)$. Hence we are able to apply the Grönwall-inequality and obtain

$$\left|\phi_s^{r^n} - \phi_s^r\right| \leq e^{Ls}\left|p^n - p\right| + e^{Ls}\int_0^s e^{-Lt}\left|h(\overline{r}_t^n, x) - h(\overline{r_t}, x)\right| dt.$$

Therefore, $|\phi_s^{r^n} - \phi_s^r| \to 0$ and finally $|\phi_t^{r^n}(x, p^n) - \phi_t^r(x, p)| \to 0$. □

Using results of MDP-theory we are in a position to state the following main existence result.

Theorem 6

(a) *There exists an optimal stationary relaxed strategy* $\pi^* = (f^*, f^*, \dots)$ *for the MDP, i.e.* $f^*(x, p) \in \mathcal{R}$ *and* $V_{\pi^*} = V$.
(b) *If* $u \mapsto F(x, p, u)$ *is convex, where*

$$F(x, p, u) := c(x, p, u) + \sum_{y \neq x} V(y, p + \Phi_{xy}(u, p)) q_{xy}^X(u, p) + V(x, p)\left(\alpha - \sum_{y \neq x} q_{xy}^X(u, p)\right)$$

then

$$\overline{\pi}^* = \left(\overline{f}^*, \overline{f}^*, \dots\right)$$

is an optimal stationary deterministic strategy for the MDP, where

$$\overline{f}^*(x, p) := \int_U u f^*(x, p)(du) \in A.$$

Proof (a) It is sufficient to prove that there exists f^* with

$$(LV)(x, p, f^*(x, p)) = \inf_{r \in \mathcal{R}} (LV)(x, p, r).$$

We know from Proposition 2 that $\mathcal{R}$ is compact. Since $r \mapsto \phi_t^r(x, p)$, $r \mapsto q_{ij}^X(r, p)$ and $r \mapsto c(x, p, r)$ are continuous and $p \mapsto V(x, p)$ is concave, we conclude that

$$r \mapsto (LV)(x, p, r)$$

is continuous. Hence there exists a measurable minimizer $f^*(x, p)$.

(b) Since U is convex, $\overline{f}^*(x,p)(t) \in U$ for all $t \geq 0$. It holds

$$\mathcal{T}V(x,p) = \inf_{a \in A}(LV)(x,p,a) \geq \inf_{r \in \mathcal{R}}(LV)(x,p,r).$$

On the other hand we get in view of (7)

$$\begin{aligned}(LV)(x,p,r) &= \int_0^\infty \int_U e^{-(\alpha+\beta)t} F\left(\phi_t^r(x,p),x,u\right) r(du)dt \\ &= \int_0^\infty \int_U e^{-(\alpha+\beta)t} F\left(\phi_t^{\overline{r}}(x,p),x,u\right) r(du)dt \\ &\geq \int_0^\infty e^{-(\alpha+\beta)t} F\left(\phi_t^{\overline{r}}(x,p),x,\overline{r}\right) dt = (LV)(x,p,\overline{r})\end{aligned}$$

and

$$\inf_{r \in \mathcal{R}}(LV)(x,p,r) \geq \inf_{r \in \mathcal{R}}(LV)(x,p,\overline{r}) = \inf_{a \in A}(LV)(x,p,a) = \mathcal{T}V(x,p).$$

In total, $\mathcal{T}V(x,p) = \inf_{r \in \mathcal{R}}(LV)(x,p,r)$. Therefore,

$$\mathcal{T}V(x,p) = (LV)(x,p,f^*(x,p)) \geq (LV)(x,p,\overline{f}^*(x,p)) \geq \mathcal{T}V(x,p).$$

Since $\overline{f}^*(x,p)$ is a measurable minimizer, the stationary deterministic strategy $(\overline{f}^*, \overline{f}^*, \ldots)$ is optimal for the MDP. □

The following existence result follows from Theorems 5 and 6. The function $u \mapsto F(x,p,u)$ is convex if the following conditions are satisfied:

$$\text{(C)} \begin{cases} u \mapsto c(x,p,u) \text{ is convex} \\ u \mapsto q_{ij}^X(u,p) \text{ is linear} \\ \Phi_{ij}(u,p) \text{ is independent of } u. \end{cases}$$

Theorem 7 *Assume* (*C*). *Then there exists an optimal control process for* $(\tilde{P})$ *and hence also for* (*P*).

5 A parallel queueing model

Queueing models appear in different and various fields, for example in data flow of the internet, in machinery productions or in call centers. They are treated carefully in several publications and also over the last years with the restriction of partial information (Altman et al. 2003; Honhon and Seshadri 2007; Lin and Ross 2003). But most of them are not able to characterize the optimal control completely, whereas we can do.

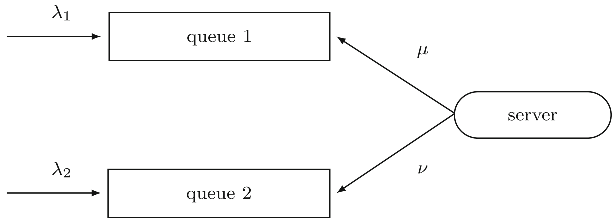

Fig. 1 Parallel queueing model

We consider now a parallel queueing model (Fig. 1) described as follows. There are two queues with infinite buffer, where customers arrive at queue k corresponding to a Poisson process with arrival rate λ_k, $k = 1, 2$. Additionally, one server is available, who has to decide at each time point t which of the two queues (to be more precisely: which one of the first customers waiting in each queue) is served. The service time of a customer in queue 1 is $\exp(\mu)$-distributed, in queue 2 $\exp(\nu)$-distributed. We assume that arrivals and service times are independent. For each waiting customer a cost at rate c_k, $k = 1, 2$, occurs and we want to find a service strategy for the server minimizing the expected discounted waiting costs over an infinite horizon.

It is well-known that in the case of complete information the μc*-rule* is optimal (Asmussen 2003). That means under $c_1\mu \geq c_2\nu$ it is optimal to serve queue 1 if there is a customer waiting, if not then serve queue 2.

We assume now that the service rates are *Bayesian*. This means in the context of the previous sections $Z_t = (Z_t^1, Z_t^2)$ with $Z_t^1 \in \{\mu_1, \mu_2\}$, $Z_t^2 \in \{\nu_1, \nu_2\}$ with $Q^Z \equiv 0$, $\mathbb{P}(Z_0^1 = \mu_1) = p_0, \mathbb{P}(Z_0^2 = \nu_1) = q_0$. This situation may occur if two types of customers are in the system and the server can not differ which group is waiting in which queue.

The server now can spend his service capacity simultaneously to both queues. Therefore the control set U is defined by $U = [0, 1]$ in contrast to the pure service restriction in the classical complete information model. There we have seen that it is never optimal to split service. We interpret $u \in U$ as the service rate spent to queue 1. Hence the service rate $1 - u$ is spent to queue 2.

The extension of the theory developed in Sects. 2, 3 and 4 to countable state spaces is straightforward if the generator of the two-dimensional Markovian jump process $X_t = (X_t^1, X_t^2)$ is conservative. We understand X_t^k as the number of customers waiting in queue k. Hence $S_X = \mathbb{N}_0 \times \mathbb{N}_0$ with $x = (x_1, x_2) \in S_X$.

Define the filter process for the both service rates by $p_t := \mathbb{P}(Z_t^1 = \mu_1 | \mathcal{F}_t^X)$ and by $q_t := \mathbb{P}(Z_t^2 = \nu_1 | \mathcal{F}_t^X)$. We will consider in the following only p_t. The results can be analogously derived for q_t.

The $\mathcal{F}_t^X$-generator of (X_t) is given by $Q^X(u, p, q) = (q_{ij}^X(u, p, q))$ with $q_{xy}^X(u, p, q)$

$$= \begin{cases} \mu(p)u\mathbb{1}(x_1 > 0) & (y_1, y_2) = (x_1 - 1, x_2) \\ \nu(q)(1 - u)\mathbb{1}(x_2 > 0) & (y_1, y_2) = (x_1, x_2 - 1) \\ \lambda_1 & (y_1, y_2) = (x_1 + 1, x_2) \\ \lambda_2 & (y_1, y_2) = (x_1, x_2 + 1) \\ -\lambda_1 - \lambda_2 - \mu(p)u\mathbb{1}(x_1 > 0) - \nu(q)(1 - u)\mathbb{1}(x_2 > 0) & y = x \end{cases}$$

where $\mu(p) := \mu_1 p + \mu_2(1-p)$ is the estimated service rate (the conditional mean) at queue 1, analogously $\nu(q) := \nu_1 q + \nu_2(1-q)$.

5.1 The filter process

Lemma 1 *The filter process* (p_t) *is the unique solution of*

$$dp_t = u_t\big(\mu_2-\mu_1\big)p_t(1-p_t)\mathbb{1}\Big(X_t^1>0\Big)\,dt + \Phi_1(p_{t-})dN_t^1\big(X_{t-}^1, X_{t-}^1-1\big) \tag{8}$$

where the jump-size $\Phi_1(p)$ *is given by*

$$\Phi_1(p) = \frac{1}{\mu(p)}\mu_1 p - p.$$

Note that $\Phi_1(p)$ is independent of the control $u=(u_t)$.

Proof The assertions are a direct consequence of Theorem 1. □

We see that p_t jumps if and only if new information about the unknown service rate is available. This is the case if and only if the service of a customer in queue 1 is finished. In this case the new estimate p_t is proportional to the possible intensities μ_1 and μ_2 with respect to the pre-jump-estimate p_{t-}, that means

$$p_t = p_{t-} + \Phi_1(p_{t-}) = \frac{1}{\mu(p_{t-})}\mu_1 p_{t-}.$$

We omit in the following $\mathbb{1}(X_t^1>0)$ since it is obvious that new information is only available through service, which is reasonable only if customers are waiting in queue 1. Between the jumps p_t is described by the deterministic part of (8)

$$dp_t = u_t(\mu_2-\mu_1)p_t(1-p_t)dt$$

and can be calculated explicitly.

Lemma 2 *Denote by* T_n *the nth jump time of* (p_t). *For* $t\in[T_n, T_{n+1})$ *and* $p_{T_n}=p$, *it holds that* $p_t = \phi_{t-T_n}^u(p)$ *where*

$$\phi_t^u(p) = \frac{p\exp\{(\mu_2-\mu_1)\int_0^t u_s ds\}}{p\exp\{(\mu_2-\mu_1)\int_0^t u_s ds\} - p + 1}. \tag{9}$$

Remark 2 (1) If $p=1$ then $\phi_t^u(p)\equiv 1$. On the other hand if $p=0$ then $\phi_t^u(p)\equiv 0$. If we have complete information about Z_t^1, then the estimator does not change between the jumps. It will be constant over time and the complete information is not destroyed.

(2) If only queue 2 is served for $s \in [t, t+h]$ then $\phi_s^u(p) \equiv \phi_t^u(p)$ for all $s \in [t, t+h]$. Thus the estimator p_t is updated only if queue 1 is served.
(3) p_t is independent of the length of the queues $X_t = (X_t^1, X_t^2)$ as long as $X_t^1 > 0$.

We assume now without loss of generality

$$\mu_1 < \mu_2. \tag{10}$$

This assumptions induces that service under μ_2 tends to an earlier completion than under μ_1. We investigate next the behaviour of $\phi_t^u(p)$ in dependence of t and p.

Lemma 3 (a) *$t \mapsto \phi_t^u(p)$ is monotone increasing for all $p \in [0, 1]$.*
(b) *$t \mapsto \phi_t^u(p)$ is Lipschitz continuous for all $p \in [0, 1]$.*
(c) *$p \mapsto \phi_t^u(p)$ is monotone increasing for all $t \geq 0$.*
(d) *$p \mapsto \phi_t^u(p)$ is concave for all $t \geq 0$.*
(e) *$p \mapsto \phi_t^u(p)$ is Lipschitz continuous, uniformly in u for all $t \geq 0$.*

Proof (a)–(d) follow directly by differentiation.
(e) We first note that for the denominator of $\phi_t^u(p)$ in (9) holds, that

$$p \exp\left\{(\mu_2 - \mu_1)\int_0^t u_s ds\right\} - p + 1 \geq 1$$

and we conclude with $p_1, p_2 \in [0, 1]$:

$$\begin{aligned} |\phi_t^u(p_1) - \phi_t^u(p_2)| &\leq \left|(p_1 - p_2)\exp\left\{(\mu_2 - \mu_1)\int_0^t u_s ds\right\}\right| \\ &\leq \exp\{(\mu_2 - \mu_1)t\}\,|p_1 - p_2|. \end{aligned}$$

□

By part (a) if $u_s > 0$ for $s \in [t, t+h]$ we have strong monotonicity for $p \in (0, 1)$. But if $u_s \equiv 0$ then $\phi_s^u(p)$ is constant in $[t, t+h]$. In other words: if one serves queue 1, then the parameter μ_1 becomes more likely. This is reasonable since $\mu_2 > \mu_1$. The greater the prior probability p, the greater is the estimate $\phi_t^u(p)$.

After we have discussed the behaviour of p_t between the jumps we analyze now the jump behaviour of p_t (where the jump-size is independent of the control). It can be proven that a jump reduces the probability that μ_1 is the true parameter.

Lemma 4

(a) $p + \Phi_1(p) = \frac{1}{\mu(p)}\mu_1 p < p$ *for* $p \in (0, 1)$*; in particular* $\Phi_1(p) < 0$.
(b) $p \mapsto p + \Phi_1(p)$ *is monotone increasing on* $(0, 1)$.
(c) $p \mapsto \phi_t^u(p) + \Phi_1(\phi_t^u(p))$ *is Lipschitz continuous.*

Proof (a)–(b) follow directly from the form of $\Phi_1(p)$.

(c) Since $\mu(p) \geq \mu_1$ and

$$\mu\left(\phi_t^u(p_2)\right)\phi_t^u(p_1) - \mu\left(\phi_t^u(p_1)\right)\phi_t^u(p_2) = \mu_2\left(\phi_t^u(p_1) - \phi_t^u(p_2)\right)$$

we conclude

$$\begin{aligned}
&\left|\phi_t^u(p_1) + \Phi_1(\phi_t^u(p_1)) - \phi_t^u(p_2) - \Phi_1(\phi_t^u(p_2))\right| \\
&= \left|\frac{\mu_1\phi_t^u(p_1)}{\mu\left(\phi_t^u(p_1)\right)} - \frac{\mu_1\phi_t^u(p_2)}{\mu\left(\phi_t^u(p_2)\right)}\right| \\
&= \mu_1\left|\frac{\mu\left(\phi_t^u(p_2)\right)\phi_t^u(p_1) - \mu\left(\phi_t^u(p_1)\right)\phi_t^u(p_2)}{\mu\left(\phi_t^u(p_1)\right)\mu\left(\phi_t^u(p_2)\right)}\right| \\
&\leq \frac{\mu_2}{\mu_1}\left|\phi_t^u(p_1) - \phi_t^u(p_2)\right|.
\end{aligned}$$

□

Note that $\Phi_1(0) = \Phi_1(1) = 0$. Thus if we have complete information before a jump, we have complete information after a jump or in other words: new information (due to jumps) gives no update. But if $p \in (0, 1)$ the estimator is updated due to a finished service. If $\mu_1 = 0$ we have $p_t = p_{t-} + \Phi_1(p_{t-}) = \frac{1}{\mu(p_{t-})}\mu_1 p_{t-} = 0$. Hence after a jump (which is impossible under the hypothesis $Z_t^1 = \mu_1 = 0$) the conditional probability $p_t \equiv 0$ for all t.

From (b) we obtain: the greater the prior probability p_{t-}, the greater the posterior probability p_t after a jump. In particular we saw in the proof that under the hypothesis $Z_t^1 = \mu_1 = 0$ the function $p \mapsto p + \Phi_1(p)$ is constant.

5.2 General statements

Since the cost function $c(x, p, u) := c_1x_1 + c_2x_2$ is independent of u and the intensities are linear in u we conclude immediately from Theorem 7 the existence of an optimal deterministic control. Before characterizing the optimal control more precisely, we compute the value function $J(x, p, q) = V(x, p, q)$.

Theorem 8 *Let $x = (x_1, x_2) \in \mathbb{N}^2$. The value function is given by*

$$J(x, p, q) = \frac{c_1x_1 + c_2x_2}{\beta} + \frac{c_1\lambda_1 + c_2\lambda_2}{\beta^2} + g(p, q) \qquad (11)$$

where $g(p, q)$ is the unique solution of

$$
\begin{aligned}
&g(p,q)\\
&\quad=\inf_{a\in A}\left\{\int_0^\infty e^{-(\alpha+\beta)t}\left\{\left(g\left(\phi_t^a(p)+\Phi_1(\phi_t^a(p)),\varphi_t^a(q)\right)-g\left(\phi_t^a(p),\varphi_t^a(q)\right)-\frac{c_1}{\beta}\right)\right.\right.\\
&\qquad\times\mu(\phi_t^a(p))a_t+\left(g\left(\phi_t^a(p),\varphi_t^a(q)+\Phi_2(\varphi_t^a(q))\right)-g\left(\phi_t^a(p),\varphi_t^a(q)\right)-\frac{c_2}{\beta}\right)\\
&\qquad\left.\left.\times\ \nu(\varphi_t^a(q))(1-a_t)+g(\phi_t^a(p),\varphi_t^a(q))\alpha\right\}dt\right\}.
\end{aligned}
$$

$\varphi_t^a(q)$ is the analogon to $\phi_t^a(p)$ for q_t.

Proof By induction on n we prove for $J_n(x,p,q):=\mathcal{T}J_{n-1}(x,p,q)$ (with $J_0:=0$) that

$$J_n(x,p,q)=(c_1x_1+c_2x_2)\,K_n+(c_1\lambda_1+c_2\lambda_2)\,L_{n-1}+g_{n-1}(p,q).$$

Here, $K_n:=\frac{1}{\alpha+\beta}\sum_{k=0}^{n-1}(\frac{\alpha}{\alpha+\beta})^k$, $L_{n+1}:=\frac{1}{\alpha+\beta}(K_{n+1}+\alpha L_n)$ (with $L_0:=0$) and further $g_{n+1}(p,q):=\mathcal{A}g_n(p,q)$ (with $g_0:=0$), where $\mathcal{A}g(p,q)$ denotes the right hand side of the characterization of $g(p,q)$ in the theorem. Since the value iterations holds, i.e. $J(x,p,q)=\lim_{n\to\infty}J_n(x,p,q)$, the statement follows. □

We see that the value function has a separation property and is also increasing in the number of waiting customers x_1 and x_2. Formula (11) has the following interpretation: The first term are the discounted costs for waiting customers, the second one are the expected discounted costs due to new arrivals and the third one are the expected reduced costs due to the finished service of a waiting customer.

Corollary 1 *Denote by $J^C(x)$ the value functions of the complete information model. Then it holds:*

$$0\geq J^C(x)-J(x,p,q)=-\frac{c_1\mu}{\beta^2}-g(p,q).$$

Proof The value function for the complete information model reads due to Theorem 8 as

$$J^C(x)=\frac{c_1x_1+c_2x_2}{\beta}+\frac{c_1\lambda_1+c_2\lambda_2}{\beta^2}-\frac{c_1\mu}{\beta^2}.$$

Since every control of the partial information model is admissible for the complete information model, we get $J(x,p,q)\geq J^C(x)$. □

Let us now study the optimal control process in more detail. From Theorem 3 we know that the value function $J(x,p,q)$ is a solution of the generalized HJB-equation given by

$$\beta v(x,p,q) = \inf_{\substack{\xi_p \in \partial_p v(x,p,q) \\ \xi_q \in \partial_q v(x,p,q) \\ u \in [0,1]}} Hv(x,p,q,u,\xi_p,\xi_q) \tag{12}$$

where the generalized Hamiltonian for $x \in \mathbb{N}^2$ is defined as

$$\begin{aligned} Hv(x,p,q,u,\xi_p,\xi_q) &:= c_1 x_1 + c_2 x_2 \\ &+ \xi_p(\mu_2-\mu_1)p(1-p)u + \xi_q(\nu_2-\nu_1)q(1-q)(1-u) \\ &+ (v(x_1+1,x_2,p,q) - v(x_1,x_2,p,q))\lambda_1 \\ &+ (v(x_1,x_2+1,p,q) - v(x_1,x_2,p,q))\lambda_2 \\ &+ (v(x_1-1,x_2,p+\Phi_1(p),q) - v(x_1,x_2,p,q))\mu(p)u \\ &+ (v(x_1,x_2-1,p,q+\Phi_2(q)) - v(x_1,x_2,p,q))\nu(q)(1-u). \end{aligned}$$

By Theorem 4 it is sufficient to compute the minimum points (u^*, ξ_p^*, ξ_q^*) of the generalized Hamiltonian. We note first that $u \mapsto Hv(\cdot, u)$ is linear. Consequently the minimum point will be (if it is unique) equal to 0 or 1. Hence the optimal control will serve one queue exclusively. If the minimum point u^* is not unique, i.e. in cases where $Hv(\cdot, u)$ does not depend on u or $Hv(\cdot, 0) = Hv(\cdot, 1)$, we choose u^* in such a way that p_t and q_t keep constant between the jumps.

Theorem 9 *There exists an optimal control $u^* = (u_t^*) \in \mathcal{U}$ with $u_t^* = u^*(X_{t-}^*, p_{t-}^*, q_{t-}^*)$ where*

$$u^*(x,p,q) = \begin{cases} 0 & x_1 = 0 \\ 1 & x_2 = 0 \\ u^*(p,q) & x \in \mathbb{N}^2. \end{cases}$$

In particular, if the minimum point of (12) *is unique, only one queue is served exclusively, i.e. $u^*(p,q) \in \{0,1\}$.*

Proof It is clear that it is never optimal to serve an empty queue, hence we only have to consider the case $x \in \mathbb{N}^2$. For this we apply the Verification Theorem 4. If we can prove that there exists a minimum point (u^*, ξ_p^*, ξ_q^*) of the generalized Hamiltonian, then the statement follows. Due to the linearity of $HJ(x,p,q,u,\xi_p,\xi_q)$ in u and by using

$$\begin{aligned} F(x,p,q) := c_1x_1 + c_2x_2 &+ (J(x_1+1,x_2,p,q) - J(x_1,x_2,p,q))\lambda_1 \\ &+ (J(x_1,x_2+1,p,q) - J(x_1,x_2,p,q))\lambda_2 \end{aligned}$$

the HJB-equation can be written as

$$\begin{aligned}
&\beta J(x,p,q) - F(x,p,q) \\
&\quad = \min \Bigg\{ \inf_{\xi_p \in \partial_p J(x,p,q)} \big\{ \xi_p (\mu_2 - \mu_1) p(1-p) \\
&\qquad\qquad + \left(J\left(x_1 - 1, x_2, p + \Phi_1(p), q\right) - J\left(x_1, x_2, p, q\right)\right) \mu(p) \big\}, \\
&\qquad\qquad \inf_{\xi_q \in \partial_q J(x,p,q)} \big\{ \xi_q (\nu_2 - \nu_1) q(1-q) \\
&\qquad\qquad + \left(J\left(x_1, x_2 - 1, p, q + \Phi_2(q)\right) - J\left(x_1, x_2, p, q\right)\right) \nu(q) \big\} \Bigg\} \\
&\quad = \min \big\{ J_{0,p}(x,p,q;1)(\mu_2 - \mu_1) p(1-p) \\
&\qquad\qquad + \left(J\left(x_1 - 1, x_2, p + \Phi_1(p), q\right) - J\left(x_1, x_2, p, q\right)\right) \mu(p), \\
&\qquad\qquad J_{0,q}(x,p,q;1)(\nu_2 - \nu_1) q(1-q) \\
&\qquad\qquad + \left(J\left(x_1, x_2 - 1, p, q + \Phi_2(q)\right) - J\left(x_1, x_2, p, q\right)\right) \nu(q) \big\}
\end{aligned}$$

where we used (4) and the definition of the lower Clarke derivative $J_{0,p}(x,p,q;1)$ with respect to p. Since $J(x,p,q)$ is regular in p we conclude that $J_{0,p}(x,p,q;1)$ exists and is equal to the right derivative. Analogously for $J_{0,q}(x,p,q;1)$. □

5.3 The symmetric case

Assume now that $\mu_1 = \nu_2$, $\mu_2 = \nu_1$ and $c_1 = c_2 = 1$. Additionally we assume that if $Z_t^1 = \mu_1$, then $Z_t^2 = \mu_2$ and vice versa. This will be called the *symmetric case*. In particular, we assume again $\mu_2 > \mu_1$ and hence we see that if the true value $Z_t^1 = \mu_1$ then queue 1 is the "bad" queue and an optimal controller prefers according to the $c\mu$-rule always queue 2. If on the other hand $Z_t^1 = \mu_2$ then the optimal decision is vice versa.

Since we are in the symmetric case it is sufficient to consider only the filter process

$$p_t := \mathbb{P}\left(Z_t^1 = \mu_1 \mid \mathcal{F}_t^X\right) = \mathbb{P}\left(Z_t^2 = \mu_2 \mid \mathcal{F}_t^X\right)$$

which is the solution of

$$\begin{aligned}
dp_t &= (\mu_2 - \mu_1)(2u_t - 1) p_t (1 - p_t) dt \\
&\quad + \Phi_1(p_{t-}) dN_t^1 \left(X_{t-}^1, X_{t-}^1 - 1\right) + \Phi_2(p_{t-}) dN_t^2 \left(X_{t-}^2, X_{t-}^2 - 1\right)
\end{aligned}$$

with

$$\Phi_1(p) := \frac{1}{\mu(p)} \mu_1 p - p \quad \text{and} \quad \Phi_2(p) := \frac{1}{\mu(1-p)} \mu_2 p - p.$$

Here we set again $\mu(p) := \mu_1 p + \mu_2(1-p)$. From this stochastic differential equation we see that between the jump times T_n and T_{n+1}

$$p_t = \phi^u_{t-T_n}(p_{T_n}) \text{ is } \left\{ \begin{matrix} \text{monotone increasing} \\ \text{constant} \\ \text{monotone decreasing} \end{matrix} \right\} \quad \text{if } u_t \left\{ \begin{matrix} > \\ = \\ < \end{matrix} \right\} \frac{1}{2}.$$

The interpretation of this result is as in Sect. 5.1. But now the filter process depends on both queues. The completion of a service in both queues leads to jumps and therefore to updates of the filter process. Since $\mu_2 > \mu_1$ it follows

$$\Phi_1(p) \leq 0 \quad \text{and} \quad \Phi_2(p) \geq 0. \tag{13}$$

By Theorem 8 the value function is given by

$$J(x, p) = \frac{x_1 + x_2}{\beta} + \frac{\lambda_1 + \lambda_2}{\beta^2} + g(p)$$

where $g(p)$ is the unique solution of

$$g(p) = \inf_{a \in A} \left\{ \int_0^\infty e^{-(\alpha+\beta)t} \left\{ \left(g\left(\phi^a_t(p) + \Phi_1(\phi^a_t(p))\right) - g\left(\phi^a_t(p)\right) - \frac{1}{\beta} \right) \mu\left(\phi^a_t(p)\right) a_t \right. \right.$$
$$+ \left(g\left(\phi^a_t(p) + \Phi_2\left(\phi^a_t(p)\right)\right) - g\left(\phi^a_t(p)\right) - \frac{1}{\beta} \right) \mu\left(1 - \phi^a_t(p)\right)(1 - a_t)$$
$$\left. \left. + g(\phi^a_t(p))\alpha \right\} dt \right\}.$$

Obviously it holds

$$g(p) = g(1-p).$$

The function $p \mapsto g(p)$ is also concave, monotone increasing for $p < \frac{1}{2}$ and decreasing for $p > \frac{1}{2}$. Therefore we get for an element ξ of the generalized Clarke gradient $\partial_p g(p) = co\left\{\limsup_{n\to\infty} \nabla g(p_n) \mid \lim_{n\to\infty} p_n = p\right\}$ that $\xi \geq 0$ if $p < \frac{1}{2}$, $\xi \leq 0$ for $p > \frac{1}{2}$ and $\{0\} \in \partial_p g(\frac{1}{2})$. Additionally we have

$$\partial_p g(p) = -\partial_p g(1-p),$$

hence $\partial_p g(\frac{1}{2})$ is a symmetric interval (with respect to 0). The following Theorem shows that the optimal control is also "symmetric".

Theorem 10 *For the existing optimal control from Theorem 9 it holds:* $u^*(p)=1-u^*(1-p)$.

Proof From the proof of Theorem 3 we conclude that the value function $J(x, p)$ and the optimal control (u_t^*) with corresponding state process (X_t^*, p_t^*) fulfils the generalized HJB-equation for almost all $t \geq 0$. For $x \in \mathbb{N}^2$ and with Theorem 8 this can be written as

$$\inf_{\xi \in \partial_p g(p_t^*)} Hg\left(p_t^*, u^*(p_{t-}^*), \xi\right) = 0,$$

where

$$\begin{aligned} Hg(p, u, \xi) := & \left(g(p + \Phi_1(p)) - g(p) - \frac{1}{\beta}\right) \mu(p)u \\ & + \left(g(p + \Phi_2(p)) - g(p) - \frac{1}{\beta}\right) \mu(1-p)(1-u) \\ & + \xi(p) \cdot (\mu_2 - \mu_1) p(1-p)(2u-1) - \beta g(p) \\ =: & \; v(p)u + w(p). \end{aligned}$$

Due to $p - \Phi_1(1-p) = p + \Phi_2(p)$ and the symmetry of $g(p)$ and $\partial_p g(p)$ we conclude that

$$Hg(1-p, u, \xi) = -v(p)u + \tilde{w}(p).$$

By the linearity of $Hg(p, u, \xi)$ in u and the symmetry of the coefficients $v(p)$ of u it follows that if $u^*(p) = 1$ is a minimum point of $u \mapsto Hg(p, u, \xi)$, then $u^*(1-p) = 0$ is a minimum point of $u \mapsto Hg(1-p, u, \xi)$. Analogously for $u^*(p) = 0$. If $p = \frac{1}{2}$ we will see later that 0 and 1 are minimum points of the generalized Hamiltonian. Thus $u^*\left(\frac{1}{2}\right) = \frac{1}{2}$ can be chosen as optimal decision. Hence the statement is proven. □

The next theorem states that it is always optimal to serve the queue where the better service rate is assumed. It shows the optimality of a threshold-strategy with threshold $p^* = \frac{1}{2}$. In other words the *certainty equivalence principle* for the $c\mu$-rule holds true, since

$$\mu(p) \geq \mu(1-p) \iff p \leq \frac{1}{2}.$$

Theorem 11 *Let $(u_t^*) \in \mathcal{U}$ be the existing optimal control from Theorem 9. Then $u^*(p)$ is given by*

$$u^*(p) = \begin{cases} 1 & p < \frac{1}{2} \\ \frac{1}{2} & p = \frac{1}{2} \\ 0 & p > \frac{1}{2}. \end{cases}$$

In particular, $u^*(p)$ is a threshold control with threshold $p^* = \frac{1}{2}$.

Proof Choose ξ^* as the lower derivative of $g(p)$ for $p < \frac{1}{2}$ (see the proof of Theorem 9). Due to the separation property of the value function $J(x, p)$ we only have to compute the minimum points of the Hamiltonian $Hg(p, \cdot, \xi^*)$. For this we have to show that

$$v(p) \begin{cases} <0 & p < \frac{1}{2} \\ =0 & p = \frac{1}{2} \end{cases}$$

(see the proof of Theorem 10). Define

$$h(p) := 2\xi^*(\mu_2 - \mu_1)p(1-p) - \frac{\mu(p)}{\beta} + \frac{\mu(1-p)}{\beta}.$$

Then

$$v(p) = \left[\left(g\left(\frac{\mu_1 p}{\mu(p)}\right) - g(p)\right)\mu(p) - \left(g\left(\frac{\mu_2 p}{\mu(1-p)}\right) - g(p)\right)\mu(1-p) + h(p)\right]u.$$

It holds that $h(p) < 0$ for $p < \frac{1}{2}$. Since $p \mapsto g(p)$ is concave, we know that

$$g(p) \geq \frac{p - p_1}{p_2 - p_1} g(p_2) + \frac{p_2 - p}{p_2 - p_1} g(p_1)$$

for all $0 \leq p_1 \leq p \leq p_2 \leq 1$, $p_2 \neq p_1$. Choosing now $p_1 := \frac{\mu_1 p}{\mu(p)}$ and $p_2 := \frac{\mu_2 p}{\mu(1-p)}$, we conclude

$$\begin{aligned}(p_2 - p_1) g(p)\mu(p)\mu(1-p) &= (\mu_2 p \mu(p) - \mu_1 p \mu(1-p))\, g(p) \\ &\geq (\mu(p)p - \mu_1 p)\,\mu(1-p) g\left(\frac{\mu_2 p}{\mu(1-p)}\right) \\ &\quad + (\mu_2 p - \mu(1-p)p)\mu(p) g\left(\frac{\mu_1 p}{\mu(p)}\right)\end{aligned}$$

and after rearranging terms

$$\begin{aligned}0 \geq &\left[\left(g\left(\frac{\mu_1 p}{\mu(p)}\right) - g(p)\right)\mu(p) - \left(g\left(\frac{\mu_2 p}{\mu(1-p)}\right) - g(p)\right)\mu(1-p)\right] \\ &\times (\mu_2 - \mu_1)(1-p).\end{aligned}$$

Since $(\mu_2 - \mu_1)(1-p) > 0$ it follows that $v(p) < 0$ for $p < \frac{1}{2}$.

If $p = \frac{1}{2}$ then $(u^*, \xi^*) = (1, g_p^0(\frac{1}{2}; 1))$ and $(u^*, \xi^*) = (0, g_{0,p}(\frac{1}{2}; 1))$ are minimum points of the Hamiltonian. Hence both allocations are optimal. We choose $u^*\left(\frac{1}{2}\right) = \frac{1}{2}$ which is a minimum point with corresponding gradient $\xi^* = \frac{1}{2}(g_{0,p}(\frac{1}{2}; 1) + g_p^0(\frac{1}{2}; 1)) = 0 \in \partial_p g(\frac{1}{2})$. In this case, the filter p_t^* remains constant ($= \frac{1}{2}$) between the jumps. □

Remark 3 From the proof we conclude the so-called *stay-on-a-winner property*: If the server finishes the service of a customer at queue k, then the server will continue serving queue k (assuming queue k is not empty). The stay-on-a-winner property follows directly from (13) and Theorem 11. A similar results was obtained in Donchev (1998) and Donchev (1999) in the context of bandit problems.

References

Altman E, Jimenez T, Nunez-Queija R, Yechiali U (2003) Optimal routing among ./M/1 queues with partial information. INRIA Research Report No. 4985
Asmussen S (2003) Applied probabilities and queues. Springer, Berlin
Bäuerle N, Rieder U (2007) Portfolio optimization with jumps and unobservable intensity process. Math Finance 17(2):205–224
Bensoussan A, Cakanyildirim M, Sethi S (2003) Partially observed inventory systems. In: Proceedings of the 44th IEEE conference on decision and control, pp 1023–1028
Bertsekas D, Shreve S (1978) Stochastic optimal control: the discrete time case. Academic Press, Dublin
Brémaud P (1981) Point processes and queues. Springer, Berlin
Clarke F (1983) Optimization and nonsmooth analysis. Wiley, New York
Davis D (1993) Markov models and optimization. Chapman & Hall, London
Donchev D (1998) On the two-armed bandit problem with non-observed Poissonian switching of arms. Math Methods Oper Res 47:401–422
Donchev D (1999) Exact solution of the Bellman equation for a β-discounted reward in a two-armed bandit with switching arms. J Appl Math Stoch Anal 12(2):151–160
Elliott R, Aggoun R, Moore J (1997) Hidden Markov models: estimation and control. Springer, Berlin
Honhon D, Seshadri S (2007) Admission control with incomplete information to a finite buffer queue. Probab Eng Inform Sci 21(1):19–46
Lin K, Ross S (2003) Admission control with incomplete information of a queueing system. INFORMS Oper Res 51:645–654
Liptser R, Shiryayev A (2004) Statistics of random processes. Springer, Berlin
Rogers L, Williams D (2003) Diffusions, Markov processes and martingales. Cambridge University Press, Cambridge
Winter J (2008) Optimal control of Markovian jump processes with different information structures. PhD-Thesis, Universität Ulm