Optimal Control of Markovian Jump Processes with Different Information Structures

ulm university universität
uulm

Optimal Control of Markovian Jump Processes with Different Information Structures

Dissertation
zur Erlangung des Doktorgrades Dr. rer. nat.
der Fakultät für Mathematik und Wirtschaftswissenschaften
der Universität Ulm

vorgelegt von
Jens Thorsten Winter

2008

Amtierender Dekan:	Prof. Dr. Frank Stehling
Erstgutachter:	Prof. Dr. Ulrich Rieder
Zweitgutachter:	Prof. Dr. Dieter Kalin
Tag der Promotion:	15. Oktober 2008

Contents

1 Introduction

In this opening section we motivate this thesis and point out the impact of this area of research. Then we summarize our main results and compare this work with the present literature. At the end we give an outline of this thesis.

1.1 Motivation

Technological advances, especially in the information technology sector, in the last few years led to a more complex relationship between different systems. One may think how the computer revolutionized the banking sector or how the internet brought together people all over the world. In the moment one billion computers are installed worldwide, compared to nearly 600 million units in 2001. In 2014 the number of installed PC will surpass two billion units. To understand the dependencies between systems requires large amount of resources and is hence expensive. Thus very often decisions are made on a lack of information. This is not only true in large and abstract systems, it even arises in the everyday life of everyone. For example most discount stores do not distinguish at checkout between the different types of yoghurt one buy. The sales clerks only count the number of yoghurts and then scan one of them. This procedure saves time and hence money, but for the inventory system is not clear anymore how many yoghurts of each type are in store. Therefore the store manager has to do his reorder based on incomplete information. Due to these inaccuracy a retailer loses roughly estimated 10% of its current profit. On the other hand this loss is at least compensated by the reduction of the costs due to the simplified scanning procedure (see Raman et al. (2001)).

Bensoussan et al. (2003) consider an one-product-inventory. In their paper, the inventory manager does not observe the inventory level, except the store is empty. Such models are called zero balance walk models and arise very often, since counting all unsold product is expensive compared to this strategy.

Another example is the following parallel queueing system:

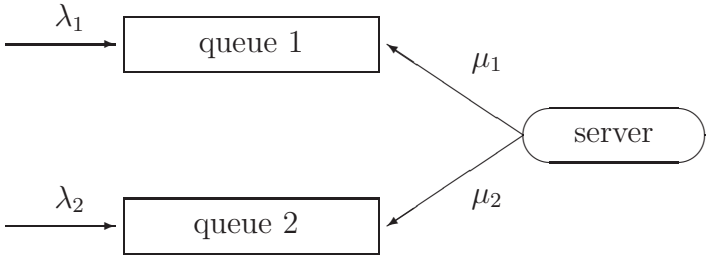


Figure 1: Parallel Queueing Model

There are two types of customer, one at each queue, which arrive at the queues randomly with rate λ_i, $i = 1, 2$. For each waiting customer a cost at rate c_i arises. The server has to decide, how to split his service capacity to the queues. His goal is to minimize the expected

discounted total cost, where the random service time depends on an intensity μ_i. If the server knows which type of customer is waiting in each queue the optimal decision is to serve always the queue with higher value $c_i\mu_i$. But if the server does not know which kind is waiting in which queue the optimal decision is not clear anymore. We will come back to this example in section 5.2.4, referred there as the symmetric case.

Such queueing models appear for example in data flow of the internet, they arise in machinery productions and in call centers. Of course, there are other applications in finance, physics and biology with incomplete information. Practitioners deal with this lack of information mostly by their experience. They estimate the unknown parameters somehow and apply a reasonable strategy to minimize the costs. But then 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 problems with incomplete information were studied mathematically and for some an explicit solution was obtained. There are usually two cases of information lack. In the first case, one is not able to observe a background process which influences the randomness of the system. In the second case, one is not able to observe every state completely. We combine both aspects in this work. To our knowledge nobody has considered a model in which only groups of states are observable, that means where the observation of the states are coarsened to an observation of groups of states.

The central questions of this thesis are:

- How to model such group observations?
- How does the optimal value of the optimization problem with incomplete information depend on the available observations?
- How to transform control problems with incomplete information to problems with complete information?
- How to solve them?

1.2 Contributions of this Thesis

In this dissertation we consider a three-component process: an environment process, a state process and an observation process. The environment process influences the state process in two ways: first, it influences the randomness of it, second changes in the environment may lead to immediate changes in the state process. Thus common jumps of both processes are explicitly possible. This fact was mostly excluded in the research. Notice that not every state of the state process is completely observable, but only groups of states are. For this purpose we introduce the notion of an information structure, which is a disjoint partition of the state space. This is done for the first time. According to this information structure the observation process is modelled. The idea that groups of states are observable has been only used by Bensoussan et al. (2003) in a very special way. Usually, the

observation is given by a process whose intensities depend on the unobservable process (see e.g. Liptser and Shiryayev (2004a), Liptser and Shiryayev (2004b), Brémaud (1981), Borisov and Stefanovich (2005), Borisov (2007), Ceci and Gerardi (2000)). Note that our setup contains the Bayesian and the Hidden-Markov-Model (e.g. Elliott et al. (1997)) too.

We want to control our system such that the expected costs, depending on this three-component-state process, are minimized. We discuss first the impact of information to the minimal cost. Then we derive the filter equation for the unobservable part of the state process. With this result we transform the optimization problem with incomplete information into one with complete information. This transformed model is a piecewise-deterministic control problem. The equivalence between these two models will be shown in the reduction theorem, which is often neglected in the literature.

Based on the reduced model we derive solution procedures for piecewise-deterministic models. First we extend the Hamilton-Jacobi-Bellman equation (HJB) to a generalized version, including the Clarke derivative. This idea is based on Clarke (1983) and Davis (1993). We state here sufficient and necessary conditions for the optimality of a control, in particular we extend the classical verification technique. The advantage of this generalization is that the strong differentiability condition can be weakened to local Lipschitz continuity and regularity of the value function, which will be fulfilled for our value function. For a second solution technique we define a time-discrete Markovian-Decision-Process (MDP), whose value function coincides with the one of the control problem. Additionally one can construct from an optimal policy of the MDP an optimal control for the origin model. Here we proof the existence of an optimal policy and we extend and link the ideas of Davis (1993), Dempster (1989) and Forwick (1997) to the uniformization technique and to discounted problems, which they did not consider. The benefit of this reduction is the opportunity of using all results from the classical MDP-theory.

Using all our developed results, we investigate a parallel queueing model under incomplete information (see the illustrating example in section 1.1). Under complete information it is well-known that the $c\mu$-rule is optimal. We prove that this strategy is also optimal if the information structure is fine enough. If the service rates are Bayesian we show the separation property of the value function and prove the existence of an optimal control, which serves one queue exclusively almost surely. Further we prove in the symmetric case that the certainty equivalence principle holds or in other words, the optimal strategy is a control limit rule with threshold $\frac{1}{2}$. If only one service rate is unknown, the optimal control serves always one queue exclusively and we state additionally sufficient conditions for the optimal control. As a by-product we extend results of the time-continuous bandit problem theory (e.g. ElKaroui and Karatzas (1997) and Kaspi and Mandelbaum (1998)) if one arm is completely observable. In contrast to Lin and Ross (2003), Honhon and Seshadri (2007) and Hordijk and Koole (1992) we do not only propose well performing strategies or prove the optimality with numerical methods (e.g. Altman et al. (2003), Altman et al. (2004)), our results are all proven rigorously. Numerical studies are done for the case, where the number of waiting customers is not observable completely.

1.3 Structure of this Thesis

We start in **section 2** by defining our state process consisting of three components: the environment process (Z_t), the state process (X_t) itself and the observation process (Y_t). All three processes are pure (Markovian) jump processes and are strongly connected to each other. In particular the environment process influences the intensities of the state process and changes in the environment may lead to immediate changes in the state process. Observable are only groups of states of the state process. Thus the notion of an information structure is introduced in definition 2.1. We construct explicitly the intensities matrices and martingales representations for (Z_t, X_t, Y_t) in (2.2), (2.5) and (2.8). As in Elliott et al. (1997) and Miller et al. (2005) our state process takes values in a finite set in contrast to Ceci and Gerardi (2000) where the state process takes values in $\mathbb{R}^d$, which is more applicable for financial applications. Examples illustrate this construction and special cases. In the following section 2.2 we define the set of admissible controls for our optimization problem stated in section 2.3. Controls are only allowed to depend on the available information and not on the unobservable parts of the processes. The optimization problem (P) is based on this three-component process, minimizing the expected discounted cost over an infinite horizon. We clarify in this context how the optimal value of this problem depends on available information and discuss the value of information. But since not all processes are observable the problem is not solvable directly.

Thus we formulate in **section 3** an optimization problem equivalent to (P) in that way, that optimal values and optimal strategies are the same, which will be stated in the reduction theorem 3.13. The benefit of these efforts are that the reduced problem is one with complete information, since the state process there is measurable with respect to the available information and so the reduced optimization problem (P_{red}) is solvable directly. The reduction is done by estimating the unobservable environment and state process with the help of conditional probabilities p_t. We compute in theorem 3.5 an explicit representation for these conditional probabilities and point out the connection to filter theory. We see that the conditional probabilities are piecewise-deterministic processes. Additionally we discuss the behaviour of them between two jumps, yielding from a change in the observation, and at jump points. In section 3.2 we state the connection between the original problem (P) under incomplete information and the reduced problem with complete information, summarized in the above mentioned reduction theorem. Finally we prove properties of the value function of the reduced problem, in particular the concavity of the value function in p.

After finding an equivalent directly solvable optimization problem we discuss in **section 4** two solution methods. We prove in theorem 4.3 that the value function is a solution of the generalized HJB-equation. Here the strict differentiability condition is weakened to differentiability in the sense of Clarke. Due to its concavity the value function is differentiable and regular in the meaning of Clarke. Then we generalize in theorem 4.4 the verification procedure to the context of Clarke derivative. Whereas the verification technique does not make use of the piecewise-deterministic behaviour of the conditional probabilities we do

this in section 4.2. There we formulate a (time-discrete) MDP whose value function coincides with the value function of (P_{red}). This will be proven in theorem 4.7. Additionally we show there, that one can construct an optimal control of (P_{red}) from an optimal policy of the MDP. Finally we answer the question about the existence of optimal controls in theorem 4.14.

In **section 5** we consider a parallel queueing model with one server as introduced in section 1.1. First we consider the complete information case in section 5.1 where the optimality of the $c\mu$-rule is proven. This control is also optimal if the information structure is fine enough. We will then apply the theory developed in section 3 and 4 to solve this problem for different information structures. In section 5.2 we consider the case of Bayesian service rates, where each service rate is unknown between two values. We derive an explicit representation for the conditional probabilities and discuss the monotonicity behaviour and technical characteristics of the estimator process in section 5.2.1. Then we define the corresponding complete information MDP in section 5.2.2 and find a closed formula for the value function in section 5.2.3. This representation is quite similar to the one under complete information (see theorem 5.13). Additionally we prove that it is always optimal to serve one queue exclusively almost everywhere. These results are specified to two models: the symmetric case in section 5.2.4 and the case, where one service rate is known in section 5.2.5. In theorem 5.20 we prove that the optimal control in the symmetric case is a control limit rule with control limit $p^* = \frac{1}{2}$. If only one service rate is unknown the optimal control serves one queue exclusively all the time as proven in theorem 5.22. In both cases the stay-on-a-winner property, famous in bandit models, is observed. In section 5.2.6 we consider a reward criterion. There the optimal (pure) control is an index-strategy. In section 5.3 the information structure is given as a 0-1-observation, in particular the server can distinguish at queue 1 only if there are more than two customers waiting or not. We investigate this case numerically.

2 The Incomplete Information Model

In this section we introduce our state process and define our optimization problem (P). We do this in a constructive way. The state process (Z_t, X_t, Y_t) consists of three components: the environment process (Z_t), which influences the parameters and the random behaviour of the system, the state process (X_t), which depends on the system parameters and the observation process (Y_t), which is connected to the further ones. The last component is the only observable component. The other two are only observable via the observation process, with the help of the then defined information structure. After introducing this three-component-process in section 2.1, we define the class of admissible controls in section 2.2. Apart from technical assumptions, the main requirement on admissible controls is that they are only functions of the observation process and hence do not depend on the unobservable parts of the system. In the last section 2.3 we introduce our optimization problem under partial information (P). We state some first properties of the optimal value in dependence on the information structure.

2.1 Construction of the Processes

Our state process consists of three components: the environment process, the state process and the observation process. All three processes exist on a given measurable space $(\Omega, \mathcal{F})$ and are strongly connected to each other. The environment process $Z = (Z_t)$ takes values in a finite set (this means d different values), where we identify for mathematical reasons each state z_μ with a unit-vector g_μ of $\mathbb{R}^d, \mu = 1, \dots, d$. Consequently the state space of $Z = (Z_t)$ with $Z_t = (Z_t^1, \dots, Z_t^d)^\top$ is given by

$$S_Z := \{g_1, \dots, g_d\}.$$

Let $N_t^Z(\mu, \nu)$ count the number of jumps of Z in $[0, t]$ from g_μ to g_ν (with $\mu \neq \nu$), which occur with (predictable) intensity $q_{\mu\nu}^Z Z_{t-}^\mu \geq 0$. Then Z_t (starting at time 0 in $z_0 \in S_Z$) is uniquely defined by

$$Z_t^\mu := z_0^\mu + \sum_{\substack{\nu=1 \\ \nu\neq\mu}}^{d} N_t^Z(\nu, \mu) - \sum_{\substack{\nu=1 \\ \nu\neq\mu}}^{d} N_t^Z(\mu, \nu), \quad \mu = 1, \dots, d. \tag{2.1}$$

Equivalent to Brémaud (1981), Elliott et al. (1997) and Rogers and Williams (2003) we can give the following martingale representation for the environment process (Z_t):

$$dZ_t = Q^Z Z_t dt + dM_t^Z, \tag{2.2}$$

where $(Q^Z)^\top = \left(q_{\mu\nu}^Z\right)$ with $q_\mu^Z := -q_{\mu\mu}^Z := \sum_{\substack{\nu=1 \\ \nu\neq\mu}}^{d} q_{\mu\nu}^Z$ is called the generator or intensity matrix of Z_t. M_t^Z is a d-dimensional martingale, consisting of the compensated counting processes

$N_t^Z(\mu,\nu)$, in detail

$$M_t^Z(\mu) := \sum_{\nu=1}^{d} \Big(N_t^Z(\nu,\mu) - N_t^Z(\mu,\nu) - \int_0^t q_{\nu\mu}^Z Z_s^\nu ds \Big), \quad \mu = 1,\dots,d.$$

Z_t characterizes the state of the environment at time t. One may think of Z_t as the economic or political situation or an external influence to the state process $X = (X_t)$, the second component of the whole state process, which takes values in the finite set

$$S_X := \{e_1,\dots,e_n\},$$

where e_i is the i-th unit vector of $\mathbb{R}^n$ (like above we model the finite set of values by the set of unit vectors for mathematical reasons).

The environment Z_t influences the state process in two ways. First, changes in the environment may lead directly to changes in the state of the process X_t. For example one may think of the environment as the economic situation and of the state process as the credit rating of a company, both rated on a scale of good, medium and bad. If the environment falls from good to medium or bad the credit rating of the company also drops by one category (if possible). This connection is now modelled as follows. Assume that $Z_{\tau-} = g_\mu$ and $X_{\tau-} = e_i$ and that the process Z jumps at time τ to state g_{ν^*}. If this jumps leads to a jump of X to state $e_{j^*} \neq e_i$, where j^* is unique, in particular $Z_\tau = g_{\nu^*}$ and $X_\tau = e_{j^*}$, then we set

$$\delta_{ij^*}^{\mu\nu^*} := 1.$$

On the other hand, if X remains in e_i (thus the jump of Z from g_μ to g_ν^* has no influence to X) then we set

$$\delta_{ij}^{\mu\nu^*} := 0 \quad \forall j \in \{1,\dots,n\}.$$

Before continuing the construction let us state that for fixed i and μ due to the uniqueness of ν^* and j^*

$$\sum_{j=1}^{n}\sum_{\nu=1}^{d} \delta_{ij}^{\mu\nu} \in \{0,1\}.$$

It is 1 if and only if there exists ν^* and j^* such that $\delta_{ij^*}^{\mu\nu^*} = 1$.

With this function we are able to model now the second component X_t of our state process (Z_t, X_t, Y_t). As above we characterize $X = (X_t)$ in terms of $N_t^X(i,j)$, the process counting the jumps of X in $[0,t]$ from e_i to e_j for $i \neq j$. That means

$$X_t^i := x_0^i + \sum_{\substack{i=1\\ i\neq j}}^{n} N_t^X(i,j) - \sum_{\substack{i=1\\ i\neq j}}^{n} N_t^X(j,i) \tag{2.3}$$

where

$$N_t^X(i,j) := \tilde{N}_t^X(i,j) + \sum_{\mu=1}^{d}\sum_{\nu=1}^{d} \delta_{ij}^{\mu\nu} N_t^Z(\mu,\nu) X_{t-}^i, \quad i \neq j. \tag{2.4}$$

Here we assume that $\tilde{N}_t^X(i,j)$ and $N_t^Z(\mu,\nu)$ do not jump at the same time $\forall i,j,\mu,\nu$. With $\tilde{N}_t^X(i,j)$ we model the jumps of X_t which happen independent of a jump of Z_t, although the intensity of $\tilde{N}_t^X(i,j)$ depends on Z_t. This is the second, the indirect, influence of Z_t on X_t. In particular the intensity of $\tilde{N}_t^X(i,j)$ is defined by

$$\tilde{q}_{ij}^X(Z_t) := \sum_{r=1}^{d} \tilde{q}_{ij,r}^X Z_{t-}^r X_{t-}^i \geq 0.$$

That means if $Z_{t-} = g_\mu$ the intensity of $\tilde{N}_t^X(i,j)$ is given by $\tilde{q}_{ij,\mu}^X X_{t-}^i$. Since $\tilde{N}_t^X(i,j)$ and Z_t do not jump at the same time it is immediately true that

$$\tilde{q}_{ij}^X(Z_t) = \sum_{r=1}^{d} \tilde{q}_{ij,r}^X Z_t^r X_{t-}^i.$$

The interpretation of $\sum_{\mu=1}^{d}\sum_{\nu=1}^{d} \delta_{ij}^{\mu\nu} N_t^Z(\mu,\nu) X_{t-}^i$ in (2.4) is the following: If $X_{t-} = e_i$ and $Z_{t-} = g_\mu$ and if Z now jumps to g_{ν^*} and this jump has some direct influence to X, meaning that X_t directly jumps to e_{j^*}, then we set $\delta_{ij^*}^{\mu\nu^*} = 1$. If the jump of Z_t has no direct influence on X (hence $X_t = X_{t-} = e_i$), then $\delta_{ij}^{\mu\nu} = 0$ for all $\nu \in \{1,\dots,d\}$ and the second term in (2.4) vanishes.

Summarizing, the intensity of $N_t^X(i,j)$ is given by:

$$q_{ij}^X(Z_t) X_{t-}^i := \Big(\sum_{r=1}^{d} \big\{ \tilde{q}_{ij,r}^X + \sum_{\nu=1}^{d} \delta_{ij}^{r\nu} q_{r\nu}^Z \big\} Z_{t-}^r \Big) X_{t-}^i.$$

Note that the intensity is predictable by construction. As in (2.2) we get the following martingale representation for X_t :

$$dX_t = Q^X(Z_t) X_t dt + dM_t^X, \tag{2.5}$$

where $(Q^X(Z))^\top = (q_{ij}^X(Z))$ with $q_i^X(Z) := -q_{ii}^X(Z) := \sum_{j\neq i} q_{ij}^X(Z)$ is the intensity matrix of X_t. With the following abbreviation

$$q_{ij,\mu}^X := \tilde{q}_{ij,\mu}^X + \sum_{\nu=1}^{d} \delta_{ij}^{\mu\nu} q_{\mu\nu}^Z =: \tilde{q}_{ij,\mu}^X + \tilde{q}_{ij,\mu}^Z$$

we can write:

$$Q^X(Z) = \Big((\tilde{Q}_1^X, \ldots, \tilde{Q}_d^X) + (\tilde{Q}_1^Z, \ldots, \tilde{Q}_d^Z) \Big) Z,$$

where $(\tilde{Q}_\mu^X)^\top = (\tilde{q}_{ij,\mu}^X)$ with $\tilde{q}_{i,\mu}^X := -\tilde{q}_{ii,\mu}^X := \sum_{j \neq i} \tilde{q}_{ij,\mu}^X$ and $(\tilde{Q}_\mu^Z)^\top := (\tilde{q}_{ij,\mu}^Z)$ with $\tilde{q}_{ij,\mu}^Z := \sum_{\nu=1}^{d} \delta_{ij}^{\mu\nu} q_{\mu\nu}^Z$ and $\tilde{q}_{i,\mu}^Z := -\tilde{q}_{ii,\mu}^Z := \sum_{j \neq i} \tilde{q}_{ij,\mu}^Z$. M_t^X is a n-dimensional martingale defined by

$$M_t^X(i) := \sum_{j=1}^{n} \Big(N_t^X(j,i) - N_t^X(i,j) - \int_0^t q_{ji}^X(Z_s) X_s^j ds \Big), \quad i = 1, \ldots n.$$

From the construction we see, that we formally have to write $M_t^{X,Z}(i)$, but for simplicity we drop this dependence in our notation.

As mentioned before not every state of X_t may be observable, only groups of states of X_t are. This situation can be observed very often in the real world. For example a server can not count the customers waiting in the queue, he only knows, if there are waiting less than 10, less than 20 or more than 20. Or in inventory systems the storekeeper only knows if the store is empty or not. This fact is modelled in the following definition.

Definition 2.1 *Let $m \in \mathbb{N}$. We call $(I(k), k = 1, \ldots, m)$ an information structure, if*

(i) $\emptyset \neq I(k) \subset \{1, \ldots, n\}$ for all k and

(ii) $I(k) \cap I(l) = \emptyset$ for all $k \neq l$ and

(iii) $\bigcup_{k=1}^m I(k) = \{1, \ldots, n\}$.

(i) means that each element of an informations structure contains at least one state of X_t, from (ii) and (iii) we have $(I(k), k = 1, \ldots, m)$ is a disjoint partition of the state space S_X. An element $I(k)$ of an information structure is identified with f_k and f_k is called a representative of e_i, if $i \in I(k)$, where f_k denotes the k-th unit vector of $\mathbb{R}^m$. By the definition of the information structure, each e_i has exactly one representative and $m \leq n$ is finite.

Remark 2.2

a) If one assumes additionally that some states of (Z_t) are observable directly (and not only via (X_t)) then one has to define the extended state process $\tilde{X}_t := (X_t, Z_t) \in S_X \times S_Z$.

b) *The information structure can also be modelled as a function*

$$g : \{e_1, \ldots, e_n\} \to \{f_1, \ldots, f_m\},$$

where each state e_i is assigned to its representative f_k. $g(x) = x$ is equivalent with complete information. $g(x) \equiv f_1$ is equivalent with no information.

c) *If we skip the disjointness assumption (ii) we possibly have more than one representative f_k of a state e_i. In this case a decision has to be made, at which time state e_i is represented by f_k and at which time by f_l. This decision could be modelled by a random variable. In this case the following construction remains true, but gets more complicate. In our opinion it is more realistic to assume the uniqueness of the representative.*

The observation process $Y = (Y_t)$, the third component of (Z_t, X_t, Y_t) is strongly connected to the information structure and is modelled as a pure jump process on the same measurable space $(\Omega, \mathcal{F})$. Y, taking values in $S_Y := \{f_1, \ldots, f_m\}$, is characterized by the processes $N_t^Y(k,l)$ $(k \neq l)$ (compare with (2.1) and (2.3)), where $N_t^Y(k,l)$ counts the jumps of Y in $[0,t]$ from f_k to f_l, in detail

$$Y_t^k := y_0^k + \sum_{\substack{l=1 \\ l\neq k}}^{m} N_t^Y(l,k) - \sum_{\substack{l=1 \\ l\neq k}}^{m} N_t^Y(k,l)$$

and $N_t^Y(k,l)$ is defined by

$$N_t^Y(k,l) := \sum_{i\in I(k)} \sum_{j\in I(l)} N_t^X(i,j) Y_{t-}^k. \tag{2.6}$$

Thus the (predictable) intensity of $N_t^Y(k,l)$ is given by

$$\begin{aligned} q_{kl}^Y(Z_t, X_t) Y_{t-}^k &:= \Big(\sum_{i\in I(k)} \sum_{j\in I(l)} q_{ij}^X(Z_{t-}) X_{t-}^i \Big) Y_{t-}^k \\ &= \Big(\sum_{i\in I(k)} \sum_{j\in I(l)} \big(\sum_{r=1}^{d} \{\tilde{q}_{ij,r}^X + \sum_{\nu=1}^{d} \delta_{ij}^{r\nu} q_{r\nu}^Z\} Z_{t-}^r \big) X_{t-}^i \Big) Y_{t-}^k. \end{aligned} \tag{2.7}$$

The martingale representation for Y_t is then given by

$$dY_t = Q^Y(Z_t, X_t) Y_t dt + dM_t^Y, \tag{2.8}$$

where $(Q^Y(Z,X))^\top = (q_{kl}^Y(Z,X))$ with $q_k^Y(Z,X) := -q_{kk}^Y(Z,X) := \sum_{\substack{l=1 \\ l\neq k}}^{m} q_{kl}^Y(Z,X)$ and

$$M_t^Y(k) := \sum_{l=1}^{m} \Big(N_t^Y(l,k) - N_t^Y(k,l) - \int_0^t q_{lk}^Y(Z_s, X_s) Y_s^l ds \Big), \quad k = 1, \ldots, m. \tag{2.9}$$

As for M_t^X we skip here the dependence of M_t^Y on X and Z in the notation.

By construction of the process (Z_t, X_t, Y_t) it is obvious, that this process is a Markov-process (although this statement is a little bit imprecise, since we have not introduced any probability measure, what we will do for the controlled processes in section 2.2).

After the construction of this three-component process we present five special models in the following.

Complete-Observation-Model
In the following three models we skip the environment-process (Z_t) for simplicity in the notations. But the models can be developed similar with the environment-process. The best case is that (X_t) is completely observable. Thus everything in the model is observable. This is attained by

$$I(i) = \{e_i\} \quad \forall i \in \{1, \dots, n\} \quad \text{and hence} \quad m = n.$$

By the construction of the processes we get immediately $Y_t \equiv X_t$.

Group-Observation-Model
In most cases only groups of states of (X_t) are observable, which leads to a coarsening of the observation. For this purpose we get

$$|I(k)| \geq 2 \quad \text{for at least one } k \quad \text{and hence} \quad m < n.$$

A very special information structure is the one where one state of X_t is observed completely and all others are represented by one common representative. This type of model arises for example in inventory systems, see e.g. Bensoussan et al. (2003) where the authors consider a zero-walk-inventory model. In such a model the storekeeper only knows if the inventory system is empty or not and in the second case the exact number of products on store is not known. We will call this model 0-1-observation and it is attained by

$$I(1) = \{e_1\} \quad \text{and} \quad I(2) = \{e_2, \dots, e_n\} \quad \text{and hence} \quad m = 2.$$

No-Observation-Model
Assume now that the observer can not differ between any state of (X_t), consequently his only information is that (X_t) takes values in S_X and starts in x_0. This extreme case is achieved by

$$I(1) = \{e_1, \dots, e_n\} \quad \text{and hence} \quad m = 1.$$

Hidden-Markov-Model
In the Hidden-Markov-Model (HMM) the generator of the state process (X_t) takes values in a finite set and is changing over time according to the unobservable environment process (Z_t). This model is considered in Elliott et al. (1997). In the classical setting (X_t) is

completely observable and (Z_t) and (X_t) do not have common jumps. Setting in our model

$$I(k) = \{e_k\} \quad \text{and} \quad \delta_{ij}^{\mu\nu} \equiv 0$$

we get $Y_t \equiv X_t$ and $dX_t = Q^X(Z_t)X_t dt + dM_t^X$ and we obtain this case.

Bayesian-Model
The setting in a Bayesian-Model is similar to the one in the Hidden-Markov-Model. Here, the generator of (X_t) is also unknown in a finite set. The only thing known is its initial distribution (sometimes denoted as a-priori-distribution) p_0. In contrast to the HMM the generator does not change over time. Again the state process (X_t) is assumed to be observable. Hence in our model we have to define

$$I(i) = \{e_i\}\ , \quad \delta_{ij}^{\mu\nu} \equiv 0 \quad \text{and} \quad Q^Z \equiv 0.$$

Remark 2.3

a) *Until now, we did not include some noise $\tilde{N}_t^Y(k,l)$ in our model. This noise leads to some incorrect observations induced by changes in the observation state since this changes are not induced by changes in the unobserved process (X_t). But all the following calculation are possible with slight modifications, if we define*

$$N_t^Y(k,l) := \tilde{N}_t^Y(k,l) + \sum_{i \in I(k)} \sum_{j \in I(l)} N_t^X(i,j) Y_{t-}^k \tag{2.10}$$

under the assumption that $\tilde{N}_t^Y(k,l)$ and $N_t^X(i,j)$ do not jump at the same time.

b) *All previous and following calculations remain true, if we allow time-dependent intensities.*

c) *By the construction of the stochastic differential equations (2.2), (2.5) and (2.8) it is clear, that all these stochastic differential equations admit a unique solution.*

d) *The construction can be extended to countable state space, e.g. $\mathbb{N}_0$ under some technical assumptions. These are the conservativeness of the generators and the condition on finite diagonal elements of the intensity matrices.*

Definition 2.4 *We denote by $\mathcal{F}_t^{Z,X,Y} := \sigma(Z_s, X_s, Y_s, s \le t)$ the (augmentation of the) filtration generated by the process $(Z_s, X_s, Y_s)_{s\in[0,t]}$. Similarly we define $\mathcal{F}_t^Y := \sigma(Y_s, s \le t)$ the (augmentation of the) filtration generated by the observation process Y_s up to time t and we call this filtration the information available at time t. Since a filtration generated by a point process is right-continuous (see Brémaud (1981) and Last and Brandt (1995)) all filtrations satisfy the usual conditions.*

By construction of the process (Z_t, X_t, Y_t) it is clear, that there is a one-to-one-relation between the processes itself and the counting processes $\big(N_t^Z(\cdot,\cdot), N_t^X(\cdot,\cdot), N_t^Y(\cdot,\cdot)\big)$. Hence we have

$$\mathcal{F}_t^Y = \sigma(N_s^Y(k,l), s \le t,\ k,l = 1,\dots,m). \tag{2.11}$$

Remark 2.5

a) *Obviously it holds:* $\mathcal{F}_t^Y \subset \mathcal{F}_t^{Z,X,Y}$.

b) *To keep things simple we identify the given* σ*-algebra* $\mathcal{F}$ *by* $\mathcal{F} = \sigma(Z_t, X_t, Y_t, t \ge 0)$.

c) *To model the available information as a filtration seems natural, since a filtration is a monotone increasing family of* σ*-algebras, thus the longer the observation horizon, the more information is available.*

In contrast to Miller et al. (2005) and Ceci and Gerardi (2000) our observation process is more than a with the unobservable process correlated process or the number of jumps of the unobservable process. We introduce here a much more general framework for systems with unobservable components including unobservable parameters. With this modelling we also cover the Bayesian and the classical Hidden-Markov cases as mentioned in the examples on page 11.

Definition 2.6 *We call an information structure* $(I(k), k = 1,\dots,m)$ *finer than another information structure* $(I(k), k = 1,\dots,m)$ *if*

(i) for all $k = 1,\dots,m$ *there exists one* $k' \in \{1,\dots,m'\}$ *with* $I(k) \subset I'(k')$

(ii) there exists at least one $k \in \{1,\dots,m\}$ *with* $I(k) \neq I'(k')$ *for all* $k' = 1,\dots,m'$.

As an immediate consequence we get $m > m'$. The following theorem gives the connection to the corresponding filtrations.

Theorem 2.7 *Let* $(I(k), k = 1,\dots,m)$ *be a finer information structure than* $(I(k), k = 1,\dots,m)$ *and denote by* (Y_t) *and* (Y_t') *the corresponding observation processes, then*

$$\mathcal{F}_t^{Y'} \subset \mathcal{F}_t^Y \quad \forall t \ge 0.$$

Proof: By definition 2.6, the fact that $m > m'$ and the previous construction of the observation process Y_t there exists at least one additional basic process $N_t^Y(k,l)$ of Y_t compared to $Y'(t)$. Hence we get with (2.11)

$$\mathcal{F}_t^{Y'} = \sigma(N_s^{Y'}(k,l), s \le t,\ k,l = 1,\dots,m') \subset \sigma(N_s^Y(k,l), s \le t,\ k,l = 1,\dots,m) = \mathcal{F}_t^Y.$$

□

2.2 Admissible Controls

In order to control the above constructed process (Z_t, X_t, Y_t) we introduce a control parameter $u \in U \subset \mathbb{R}$ and let the intensities $q^Z_{\mu\nu}(u)$ and $q^X_{ij,\mu}(u)$ depend on this parameter. To guarantee for a fixed control process $u = (u_t)$ the well-definedness and the existence of a process (Z_t, X_t, Y_t) satisfying the stochastic state differential equations (which are the controlled analogons of (2.2), (2.5) and (2.8))

$$\begin{aligned} dZ_t &= Q^Z(u_t)Z_t dt + dM^Z_t \\ dX_t &= Q^X(u_t, Z_t)X_t dt + dM^X_t \\ dY_t &= Q^Y(u_t, Z_t, X_t)Y_t dt + dM^Y_t \\ (Z_0, X_0, Y_0) &= (z_0, x_0, y_0) \end{aligned} \tag{2.12}$$

we have to make some assumptions on our control parameter and our control process.

Definition 2.8 *Let $U \subset \mathbb{R}$ be a compact set such that for all $u \in U$ holds $q^Z_{\mu\nu}(u) \geq 0$ for all $\mu \neq \nu$ and $\tilde{q}^X_{ij,\mu}(u) \geq 0$ for all $i \neq j$ and for all μ. Additionally let $u \mapsto q^Z_{\mu\nu}(u)$ and $u \mapsto \tilde{q}^X_{ij,\mu}(u)$ be continuous on U. A control (process) $u = (u_t) : [0, \infty) \to U$ satisfies assumption (A), if*

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

We define the set of admissible controls by

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

and call an element of $u \in \mathcal{U}$ accordingly admissible.

While the first assumption of (A) is rather technical, the second one means, that the control at time t is allowed to depend only on the information coming from the observations via the process (Y_s) up to time t. Especially the control is not allowed to depend on the unobserved processes (Z_s) and (X_s) up to time t and not on the future.

By the construction in section 2.1 and (2.12) we note the dependence of the process (Z_t, X_t, Y_t) and the martingales (M^Z_t, M^X_t, M^Y_t) on the control process $u = (u_t)$, but once more we neglect this fact in our notation. Keep in mind, that the randomness (which enters the system via the counting processes $N_t(\cdot, \cdot)$) in the system is influenced by the control process as it is often the case in technical applications. This is in contrast to many applications in finance where the randomness is given by an uncontrolled Lévy-process.

It is not difficult to extend the control set to the class of impulse controls (see for example Dempster and Ye (1995)), where we are able to move the process X_t directly from one state to another. That means by applying an impulse control $\Gamma \in S_X$ at time $\tau-$ the process (Z_t, X_t, Y_t) is moved from $(Z_{\tau-}, X_{\tau-}, Y_{\tau-})$ to $(Z_{\tau-}, \Gamma, \Upsilon)$, where $\Upsilon := \sum_{k=1}^{m} f_k \mathbb{1}\big(\Gamma_i \in I(k)\big)$

denotes the new state of the observation process Y_t according to Γ, assuming that if an impulse control is applied a jump of the processes at the same time is impossible (Γ_i denotes the projection of $\Gamma = e_i$ to its index i). The following computations remain true with slight modifications as long as the impulse control remains $\mathcal{F}_t^Y$-predictable.

Theorem 2.9 *For each $u \in \mathcal{U}$ and given $(z_0, x_0, y_0) \in S_Z \times S_X \times S_Y$ exists a probability measure $\mathbb{P}_u := \mathbb{P}_{u,(z_0,x_0,y_0)}$ on the given measurable space $(\Omega, \mathcal{F})$ such that there exists a process (Z_t, X_t, Y_t) satisfying* (2.12).

Proof: The assertion follows from Kolmogorov's theorem. □

Remark 2.10

a) *We can now state more precisely the notion martingale: $(M_t^Z), (M_t^X)$ and (M_t^Y) are martingales with respect to $\mathcal{F}_t^{Z,X,Y}$ on the given probability space $(\Omega, \mathcal{F}, \mathbb{P}_u)$.*

b) *$(Z_t), (X_t)$ and (Y_t) are Markovian jump processes with generator (or intensity matrix) $Q^Z(u), Q^X(u, Z)$ and $Q^Y(u, Z, X)$. But they are not any longer Markov processes (in contrast to the uncontrolled processes in section 2.1), since the controls need not to be Markovian. Since the intensity at time t for the next jump depends only on the current state of the system and the current control, the notion Markovian is justified as in Rishel (1978). If the controls are Markovian then the processes are Markovian in the usual sense.*

The next theorem states that the more information the larger the set of admissible controls. This is reasonable, since if one has more information one has more opportunities to decide.

Theorem 2.11 *Let $(I(k), k = 1, \ldots, m)$ be a finer information structure than $(I'(k), k = 1, \ldots, m')$ with associated observation process Y_t and Y_t' respectively. Then*

$$\mathcal{U}' \subset \mathcal{U},$$

where $\mathcal{U}$ and $\mathcal{U}'$ are the set of admissible controls corresponding to Y_t and Y_t'.

Proof: By theorem 2.7 follows that $\mathcal{F}_t^{Y'} \subset \mathcal{F}_t^Y$ and the assertion is an immediate consequence by the definition of the set of admissible controls in definition 2.8. □

2.3 The Optimization Problem

Denote by $\mathbb{E}_u$ the expectation with respect to $\mathbb{P}_u = \mathbb{P}_{u,(z_0,x_0,y_0)}$. Let $\beta > 0$ be a given discount factor and assume that for each $u \in \mathcal{U}$ the integrability condition

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

for a given Borel-measurable cost function $g : S_Z \times S_X \times S_Y \times U \to \mathbb{R}_+$ holds. g is assumed to be continuous in u (this assumption is of more technical nature in section 4.1 and 4.2). Then our optimization problem (P) over an infinite horizon is given by

$$(P) \begin{cases} \mathbb{E}_u \left[\int_0^\infty e^{-\beta t} g(Z_t, X_t, Y_t, u_t) dt\right] \to \min \\ dZ_t = Q^Z(u_t) Z_t dt + dM_t^Z \\ dX_t = Q^X(u_t, Z_t) X_t dt + dM_t^X \\ dY_t = Q^Y(u_t, Z_t, X_t) Y_t dt + dM_t^Y \\ (Z_0, X_0, Y_0) = (z_0, x_0, y_0) \\ u \in \mathcal{U} \end{cases}$$

We will denote for a fixed control process u the corresponding expected discounted cost by

$$J(z_0, x_0, y_0; u) := \mathbb{E}_u \left[\int_0^\infty e^{-\beta t} g(Z_t, X_t, Y_t, u_t) dt \mid Z_0 = z_0, X_0 = x_0, Y_0 = y_0\right]$$

and the optimal value of (P) for fixed (z_0, x_0, y_0) by

$$v(P) := J(z_0, x_0, y_0) := \inf_{u \in \mathcal{U}} J(z_0, x_0, y_0; u).$$

A control $u^* = (u_t^*)$ is called optimal if and only if $J(z_0, x_0, y_0; u^*) = J(z_0, x_0, y_0)$.

Note once more that Z_t and X_t are not observable directly. Thus (P) is a problem with partial information. Therefore it is not solvable directly, since the control process is allowed to depend only on the present observation $\mathcal{F}_t^Y$. We will show how a solution of (P), if there exists one, can be derived with the help of conditional probabilities and a reduced problem in the next chapter. Before we state the dependence between information and costs.

Theorem 2.12 *Assume g independent of Y and let $(I(k), k = 1, \ldots, m)$ be a finer information structure than $(I'(k), k = 1, \ldots, m')$ with associated problems (P) and (P'). Then*

$$v(P) \leq v(P').$$

Proof: The assertion is a direct consequence of theorem 2.11. □

Comparing (P) with the complete information model (P_{com}), which means, all processes are directly observable and our control is allowed to be $\mathcal{F}_t^{Z,X,Y}$-predictable, results in the following corollary.

Corollary 2.13 *It holds:*

a) $v(P_{com}) \leq v(P)$

b) If the optimal control of (P_{com}) is $\mathcal{F}_t^Y$-predictable, then it is optimal for (P) and $v(P_{com}) = v(P)$.

Proof: It is clear that the set of admissible controls for (P_{com}) is larger than the one for (P) and thus part a) follows as in theorem 2.12. If the optimal control for (P_{com}) is an element of the admissible control for (P) the equality in part b) is an obvious consequence of a). □

Part b) of the previous corollary holds for example for optimal control limit rules, if the control limit and all parameters are completely observable. It can be weakened in the following sense: if the optimal control in the complete information model is the same for all states e_i in an observation group $I(k)$, then this control is also optimal for the incomplete information model in state f_k. Sometimes this property is called structure maintaining. Thus in this case it is sufficient to know that the unobservable part of the process is in some group and the exact state does not matter to apply the optimal (complete information) control.

Corollary 2.14 *Let u^* be an optimal Markovian control of (P_{com}), which is independent of (Z_t) and the same for all $i \in I(k)$, that means $u^*(e_i) \equiv u$ for all $i \in I(k)$. Then an optimal control $\nu^*(f_k)$ of (P) in f_k is given by $\nu^*(f_k) = u$.*

Proof: We will omit the proof here and will present it in section 4.1 on page 36. □

3 The Reduction to a Model with Complete Information

We introduced problem (P) in the last chapter, an optimization problem under partial information, which is not directly solvable. Since an observation of Z_t and X_t is not completely possible, we will use the conditional probability $\mathbb{P}(Z_t = g_\mu, X_t = e_i \mid \mathcal{F}_t^Y)$ as an estimator for these both processes under the information available at time t modelled by $\mathcal{F}_t^Y$. In section 3.1 we derive an explicit martingale representation for this conditional probability (theorem 3.5). Additionally we discuss the behaviour and properties of it. Then, in section 3.2, we introduce the reduced optimization problem (P_{red}), which is strongly connected to (P), as we point out in the reduction theorem 3.13: optimal values and optimal controls are the same. This reduced optimization problem is under complete information, where the unobservable processes Z_t and X_t are replaced by their common estimator. But it is not a pure jump model anymore, whereas the behaviour between two jumps is deterministic.

Before stating the results for our setting we explain the idea for the derivation of the filter equation (which is adopted to Brémaud (1981)). Assume that (R_t) is a Markovian jump process with finite state space $S_R = \{e_1, \dots, e_n\}$ on a filtered probability space $(\Omega, (\mathcal{F}_t), \mathbb{P})$ and with the martingale representation

$$dR_t = Q^R R_t dt + dM_t^R$$

where Q^R is the generator and M_t^R the corresponding $\mathcal{F}_t$-martingale. As in section 2.1 R_t is defined via the counting process $N_t^R(i,j)$ having intensity $q_{ij}^R R_{t-}^i$. Let now $\mathcal{G}_t \subset \mathcal{F}_t$ for all $t \geq 0$ with $\mathcal{G}_0 = \{\emptyset, \Omega\}$ then

$$d\widehat{R}_t := d\mathbb{E}[R_t \mid \mathcal{G}_t] = Q^R \widehat{R}_t dt + d\widehat{M}_t,$$

where $\widehat{M}_t$ is a $\mathcal{G}_t$-martingale.

Since this introduction should be illustrative we consider a one-dimensional observation process and define for $\alpha_{ij} \in \{0, 1\}$

$$N_t := \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_{ij} N_t^R(i,j).$$

Hence N_t, counting some transitions of (R_s) in $[0,t]$, is a Poisson process with respect to $\mathcal{F}_t$ having intensity

$$\lambda(R_t) := \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_{ij} q_{ij}^R R_{t-}^i$$

and $M_t^N := N_t - \int_0^t \lambda_s(R_s) ds$ is the corresponding martingale. The quadratic covariation

$[R,N]_t$ of R_t and N_t is consequently given by

$$\begin{aligned} d[R,N]_t &= \sum_{i=1}^{n}\sum_{j=1}^{n}(e_j-e_i)d[N^R(i,j),N]_t \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}(e_j-e_i)\sum_{k=1}^{n}\sum_{l=1}^{n}\alpha_{kl}d[N^R(i,j),N^R(k,l)]_t \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}(e_j-e_i)\alpha_{ij}dN_t^R(i,j). \end{aligned}$$

We define $\mathcal{G}_t = \mathcal{F}_t^N$ then

$$\widehat{M}_t^N = N_t - \int_0^t \lambda(\widehat{R}_s)ds,$$

where $\lambda(\widehat{R_t}) := \mathbb{E}[\lambda(R_t) \mid \mathcal{F}_t^N] = \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_{ij}q_{ij}^R\widehat{R}_{t-}^i$, is a $\mathcal{F}_t^N$-martingale. Also one knows from the martingale representation theorem that there exists a unique $\mathcal{F}_t^N$-predictable process ϕ_t such that

$$d\widehat{M}_t = \phi_t d\widehat{M}_t^N.$$

Summarizing

$$d\widehat{R}_t = Q^R\widehat{R}_t dt + \phi_t d\widehat{M}_t^N.$$

To compute ϕ_t we consider with Itô's formula

$$\begin{aligned} R_tN_t &= \int_0^t R_{s-}dN_s + \int_0^t N_{s-}dR_s + [R,N]_t \\ &= \int_0^t R_{s-}dM_s^N + \int_0^t R_{s-}\lambda(R_s)ds + \int_0^t N_{s-}(Q^RR_sds + dM_s^R) \\ &\quad + \int_0^t\sum_{i=1}^{n}\sum_{j=1}^{n}(e_j-e_i)\alpha_{ij}dN_s^R(i,j) \\ &= \int_0^t R_{s-}dM_s^N + \int_0^t R_{s-}\lambda(R_s)ds + \int_0^t N_{s-}(Q^RR_sds + dM_s^R) \\ &\quad + \int_0^t\sum_{i=1}^{n}\sum_{j=1}^{n}(e_j-e_i)\alpha_{ij}q_{ij}^RR_{s-}^ids + \int_0^t\sum_{i=1}^{n}\sum_{j=1}^{n}(e_j-e_i)\alpha_{ij}dM_s^R(i,j) \\ \Rightarrow \mathbb{E}[R_tN_t] &= \mathbb{E}\left[\int_0^t N_{s-}Q^RR_sds\right] \\ &\quad + \mathbb{E}\left[\int_0^t R_{s-}\lambda(R_s)ds + \int_0^t\sum_{i=1}^{n}\sum_{j=1}^{n}(e_j-e_i)\alpha_{ij}q_{ij}^RR_{s-}^ids\right] \quad (3.1) \end{aligned}$$

and similar

$$\begin{aligned}
\widehat{R}_t N_t &= \int_0^t \widehat{R}_{s-} dN_s + \int_0^t N_{s-} d\widehat{R}_s + [\widehat{R}, N]_t \\
&= \int_0^t \widehat{R}_{s-} dM_s^N + \int_0^t \widehat{R}_{s-} \lambda(\widehat{R_s}) ds + \int_0^t N_{s-} (Q^R \widehat{R}_s ds + \phi_s dM_s^N) + \int_0^t \phi_s dN_s \\
&= \int_0^t \widehat{R}_{s-} dM_s^N + \int_0^t \widehat{R}_{s-} \lambda(\widehat{R_s}) ds + \int_0^t N_{s-} (Q^R \widehat{R}_s ds + \phi_s dM_s^N) \\
&\quad + \int_0^t \phi_s \lambda(\widehat{R_s}) ds + \int_0^t \phi_s dM_s^N \\
\Rightarrow \mathbb{E}[\widehat{R}_t N_t] &= \mathbb{E}\left[\int_0^t N_{s-} Q^R \widehat{R}_s ds\right] + \mathbb{E}\left[\int_0^t (\widehat{R}_{s-} + \phi_s) \lambda(\widehat{R_s}) ds\right]. \qquad (3.2)
\end{aligned}$$

Since (3.1) and (3.2) has to be equal we see that

$$\phi_s := \phi(\widehat{R}_{s-}) = \frac{1}{\lambda(\widehat{R_{s-}})} \sum_{i=1}^n \sum_{j=1}^n \alpha_{ij} e_j q_{ij}^R \widehat{R}_{s-}^i - \widehat{R}_{s-}$$

fulfils this condition. Consequently the filter equation is given by

$$d\widehat{R}_t = Q^R \widehat{R}_t dt + \phi(\widehat{R}_{t-}) dM_t^N. \qquad (3.3)$$

If only the jump from e_{i^*} to e_{j^*} is observable, set $\alpha_{i^* j^*} = 1$ and all other $\alpha_{ij} = 0$, then after a jump at τ the new estimate is

$$\widehat{R}_\tau = \widehat{R}_{\tau-} + \phi(\widehat{R}_{\tau-}) = e_j,$$

thus we know with probability one, that R_t is in state e_j. If on the other hand no jump is observable, thus $\alpha_{ij} \equiv 0$, then

$$d\widehat{R}_t = Q^R R_t dt,$$

which is the Kolmogorov's backward differential equation. If $\alpha_{ij} \equiv 1$, then all jumps are counted and the result coincides with the one in Brémaud (1981).

3.1 Filter Equation for the Unobservable Process

To keep the notation manageable we will drop in this section the dependence of the control process and compute all following formulas for the uncontrolled case, but they remain true for admissible controls u, since controls have to be $\mathcal{F}_t^Y$-predictable, $u \mapsto q_{kl}^Y(u, Z, X)$ and $u \mapsto q_{ij}^X(u, Z)$ are continuous and U is compact. First of all define the conditional probability of (Z_t, X_t) given the present information $\mathcal{F}_t^Y$

$$p_t(i, \mu) := \mathbb{P}\left(X_t = e_i, Z_t = g_\mu \mid \mathcal{F}_t^Y\right) \qquad (3.4)$$

and $p_t := (p_t(1,1), \ldots, p_t(1,d), p_t(2,1), \ldots, p_t(n,1), \ldots, p_t(n,d)) \in \triangle^{nd}$, where

$$\triangle^{\kappa} := \{x \in [0,1]^{\kappa} \mid \sum_{i=1}^{\kappa} x_i = 1\}$$

denotes the κ-dimensional probability simplex.

The following relation holds for the marginal distributions.

Lemma 3.1

a) $\mathbb{P}(X_t = e_i \mid \mathcal{F}_t^Y) = p_t(i,\cdot) = \sum_{r=1}^{d} p_t(i,r)$

b) $\mathbb{P}(Z_t = g_\mu \mid \mathcal{F}_t^Y) = p_t(\cdot,\mu) = \sum_{i=1}^{n} p_t(i,\mu)$

Proof: The claim is an immediate consequence of the properties of marginal distributions. □

The following shows the connection between the conditional probabilities and the filter technique. We make the following convention: if we write $X_t Z_t$, we mean $X_t Z_t^\top$ and similar $e_i g_\mu$ should be understood as $e_i g_\mu^\top$.

Lemma 3.2 *It holds:*

a) $p_t(i,\mu) = \mathbb{E}[(X_t Z_t)_{i\mu} \mid \mathcal{F}_t^Y]$

b) $p_t(i,\cdot) = \mathbb{E}[X_t^i \mid \mathcal{F}_t^Y]$

c) $p_t(\cdot,\mu) = \mathbb{E}[Z_t^\mu \mid \mathcal{F}_t^Y]$

Proof: We only prove part a), since the others are immediate consequences of lemma 3.1, by:

$$\mathbb{E}[(X_t Z_t)_{i\mu} \mid \mathcal{F}_t^Y] = \mathbb{E}[\mathbb{1}\big(X_t = e_i, Z_t = g_\mu\big) \mid \mathcal{F}_t^Y] = \mathbb{P}(X_t = e_i, Z_t = g_\mu \mid \mathcal{F}_t^Y) = p_t(i,\mu).$$

□

We introduce the operator $\mathcal{S} : \mathbb{R}^{n\times d} \to \mathbb{R}^{nd}$, who writes the rows of a matrix $A = (a_{ij}) \in \mathbb{R}^{n\times d}$ in one row of a vector, by

$$\mathcal{S}A := (a_{11}, \ldots, a_{1d}, a_{21}, \ldots, a_{2d}, \ldots, a_{n1}, \ldots, a_{nd})^\top \in \mathbb{R}^{nd}.$$

As a consequence of the last lemma, this definition and $\widehat{X_t Z_t} := \mathbb{E}[X_t Z_t \mid \mathcal{F}_t]$ it is immediately true that

$$\mathcal{S}\left(\widehat{XZ}\right) = p.$$

The next lemma is an easy algebraic transformation, but will be useful in the presentation of the filter equation for p_t.

Lemma 3.3 *Let $X = (x_{ij}) \in [0,1]^{n\times d}$ be a matrix, whose entries sum up to 1, that means $\sum_{i=1}^{n}\sum_{j=1}^{d} x_{ij} = 1$. Then for the matrix*

$$A(X) := \begin{pmatrix} a_{11}x_{11}+c_{11} & a_{12}x_{12}+c_{12} & \cdots & a_{1d}x_{1d}+c_{1d} \\ a_{21}x_{21}+c_{21} & \cdots & \ddots & \vdots \\ \vdots & \ddots & & \vdots \\ a_{n1}x_{n1}+c_{n1} & a_{n2}x_{n2}+c_{n2} & \cdots & a_{nd}x_{nd}+c_{nd} \end{pmatrix}$$

exists a matrix $\mathcal{A} \in \mathbb{R}^{nd\times nd}$ such that for $\lambda \in \mathbb{R}$

$$\mathcal{S}(\lambda A(X)) = \lambda \mathcal{A}\mathcal{S}X.$$

Especially $\mathcal{A}$ is given by

$$\mathcal{A} = \begin{pmatrix} a_{11}+c_{11} & c_{11} & c_{11} & \cdots & c_{11} \\ c_{12} & a_{12}+c_{12} & c_{12} & \cdots & c_{12} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ c_{nd} & c_{nd} & \cdots & \cdots & a_{nd}+c_{nd} \end{pmatrix}.$$

Additionally $\mathcal{S}(A_1(X)+A_2(X)) = \mathcal{A}_1\mathcal{S}X + \mathcal{A}_2\mathcal{S}X$.

We have to derive a representation in form of a stochastic differential equation for $p_t \in \triangle^{nd}$. If we can do this, we also have a representation for the marginal distributions by lemma 3.1. Because of (2.11) the filter technique can be applied to each $N_t^Y(k,l)$ instead of Y_t and the conditional probability p_t can be expressed in terms of the Poisson processes $N_t^Y(k,l)$. The derivation well be done as described in the opening of this section on page 18, whereas now

- X_tZ_t is the unknown process R_t
- $N_t^Y(k,l)$ acts as N_t
- $\alpha_{ij}^{kl}(I) = \mathbb{1}\big(i \in I(k), j \in I(l)\big)$ exchanges α_{ij} and depends on the information structure $(I(k), k = 1,\ldots,m)$

It is well known by Brémaud (1981) and Last and Brandt (1995), that the (predictable) $\mathcal{F}_t^Y$-intensity of $N_t^Y(k,l)$ is given by

$$\begin{aligned} q_{kl}^Y(p_{t-})Y_{t-}^k &:= \mathbb{E}[q_{kl}^Y(Z_t,X_t)Y_{t-}^k \mid \mathcal{F}_{t-}^Y] \\ &= \left(\sum_{i\in I(k)}\sum_{j\in I(l)}\sum_{r=1}^{d}\left\{\tilde{q}_{ij,r}^X + \sum_{\nu=1}^{d}\delta_{ij}^{r\nu}q_{r\nu}^Z\right\}p_{t-}(i,r)\right)Y_{t-}^k. \end{aligned}$$

Take care, despite the similar notation, about the difference of $q^Y_{kl}(Z,X)Y$, which is the $\mathcal{F}^{Z,X,Y}_t$-intensity, and the $\mathcal{F}^Y_t$-intensity $q^Y_{kl}(p)$. Note that $p \mapsto q^Y_{kl}(p)$ is linear.

With (2.8) in mind we get the $\mathcal{F}^Y_t$-representation of Y_t as (see Brémaud (1981), Last and Brandt (1995))

$$dY_t = Q^Y(p_t)Y_t dt + d\widehat{M}^Y_t, \tag{3.5}$$

where $\widehat{M}^Y_t$ is a $\mathcal{F}^Y_t$-martingale defined by

$$\widehat{M}^Y_t(k) = \sum_{l=1}^m \Big(N^Y_t(l,k) - N^Y_t(k,l) - \int_0^t q^Y_{lk}(p_s)Y^l_s ds \Big), \quad k = 1,\dots,m,$$

similar to (2.9) and $Q^Y(p) = (q^Y_{kl}(p))$ with $q^Y_k(p) := -q^Y_{kk}(p) := \sum\limits_{\substack{l=1\\ l\neq k}}^m q^Y_{kl}(p)$.

As in the opening of this section we derive the filter equation for $\widehat{X_tZ_t}$ in the following lemma. Note that $q^Y_{kl}(p) = q^Y_{kl}(\widehat{XZ})$.

Lemma 3.4 *Define* $\phi_{(k,l)}(t) := \phi_{(k,l)}(\widehat{X_{t-}Z_{t-}})$ *by*

$$\phi_{(k,l)}(t) := \frac{1}{q^Y_{kl}(\widehat{X_{t-}Z_{t-}})} \sum_{i\in I(k)} \sum_{j\in I(l)} e_j \sum_{\mu=1}^d \big(g_\mu \tilde{q}^X_{ij,\mu} + \sum_{\nu=1}^d \delta^{\mu\nu}_{ij} g_\nu q^Z_{\mu\nu} \big) (\widehat{X_{t-}Z_{t-}})_{i\mu} - \widehat{X_{t-}Z_{t-}}$$

then the filter equation is given by

$$\begin{aligned}
& d\widehat{X_tZ_t} \\
= \; & \Big\{ \sum_{i=1}^n \sum_{\mu=1}^d (e_i Q^Z g_\mu + (\tilde{Q}^X_\mu + \tilde{Q}^Z_\mu) e_i g_\mu) (\widehat{X_tZ_t})_{i\mu} \\
& \qquad + \sum_{i=1}^n \sum_{j=1}^n \sum_{\mu=1}^d \sum_{\nu=1}^d \delta^{\mu\nu}_{ij} (e_j - e_i)(g_\nu - g_\mu) (\widehat{X_tZ_t})_{i\mu} q^Z_{\mu\nu} \Big\} dt \\
& + \sum_{k=1}^m \sum_{l=1}^m \phi_{(k,l)}(t) \big(dN^Y_t(k,l) - q^Y_{kl}(p_t) Y^k_t dt \big) \\
=: \; & A(\widehat{X_tZ_t}) dt + \sum_{k=1}^m \sum_{l=1}^m \phi_{(k,l)}(t) \big(dN^Y_t(k,l) - q^Y_{kl}(p_t) Y^k_t dt \big)
\end{aligned} \tag{3.6}$$

Considering (3.6) we see it is of the same structure as (3.3):

- The dt-term is the $\mathcal{F}^Y_t$-generator of $\widehat{X_tZ_t}$ and is independent of the information structure. In particular if there are no common jumps of X_t and Z_t then all $\delta^{\mu\nu}_{ij} = 0$.

- $\phi_{(k,l)}(t)$ is of the form

$$\frac{1}{q^Y_{kl}(\widehat{XZ})}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha^{kl}_{ij}(I)e_j\sum_{\mu=1}^{d}\Big(g_\mu\tilde{q}^X_{ij,\mu}+\sum_{\nu=1}^{d}\delta^{\mu\nu}_{ij}g_\nu q^Z_{\mu\nu}\Big)(\widehat{XZ})_{i\mu}-\widehat{XZ}.$$

The proof of the lemma is done in three steps, where the proof of these individual steps are excluded in the appendix. First we compute a martingale representation for $X_tZ_tN^Y_t(k,l)$, then one for $\widehat{X_tZ_t}N^Y_t(k,l)$. Since the expectation of these two expressions has to be the same we determine the innovation gain $\phi_{(k,l)}(t)$ function by comparison of the coefficients.

Proof: Compute $X_tZ_tN^Y_t(k,l)$ (see lemma A.5):

$$\begin{aligned}
&X_tZ_tN^Y_t(k,l) \qquad (3.7)\\
=\;&\int_0^t X_{s-}Z_{s-}dN^Y_s(k,l)+\int_0^t N^Y_{s-}(k,l)\Big\{\sum_{i=1}^{n}\sum_{\mu=1}^{d}(e_iQ^Zg_\mu+(\tilde{Q}^X_\mu+\tilde{Q}^Z_\mu)e_ig_\mu)X^i_sZ^\mu_sds\\
&\qquad+\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\mu=1}^{d}\sum_{\nu=1}^{d}\delta^{\mu\nu}_{ij}(e_j-e_i)(g_\nu-g_\mu)X^i_{s-}dN^Z_s(\mu,\nu)\Big\}\\
&+\int_0^t N^Y_{s-}(k,l)X_{s-}dM^Z_s+\int_0^t N^Y_{s-}(k,l)Z_{s-}dM^X_s\\
&+\int_0^t\sum_{i\in I(k)}\sum_{j\in I(l)}(e_j-e_i)X^i_{s-}Z_sY^k_{s-}d\tilde{N}^X_s(i,j)\\
&+\int_0^t\sum_{i\in I(k)}\sum_{j\in I(l)}\sum_{\mu=1}^{d}\sum_{\nu=1}^{d}\delta^{\mu\nu}_{ij}(e_jg_\nu-e_ig_\mu)X^i_{s-}Z^\mu_{s-}Y^k_{s-}dN^Z_s(\mu,\nu).
\end{aligned}$$

Compute $\widehat{X_tZ_t}N^Y_t(k,l)$ (see lemma A.7):

$$\begin{aligned}
&\widehat{X_tZ_t}N^Y_t(k,l) \qquad (3.8)\\
=\;&\int_0^t N^Y_{s-}(k,l)\Big\{\sum_{i=1}^{n}\sum_{\mu=1}^{d}(e_iQ^Zg_\mu+(\tilde{Q}^X_\mu+\tilde{Q}^Z_\mu)e_ig_\mu)(\widehat{X_sZ_s})_{i\mu}\\
&\qquad+\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\mu=1}^{d}\sum_{\nu=1}^{d}\delta^{\mu\nu}_{ij}(e_j-e_i)(g_\nu-g_\mu)(\widehat{X_sZ_s})_{i\mu}q^Z_{\mu\nu}ds+d\widehat{M}_s\Big\}\\
&+\int_0^t\widehat{X_{s-}Z_{s-}}dN^Y_s(k,l)+\int_0^t\phi_{(k,l)}(s)dN^Y_s(k,l).
\end{aligned}$$

Since the expectations of (3.7) and (3.8) have to be the same, we are finally able to determine the innovation gain function $\phi_{(k,l)}(t):=\phi_{(k,l)}(\widehat{X_{t-}Z_{t-}})$. By the arguments of

Fubini we see that the $\int_0^t N_{s-}^Y(k,l)\{\dots\}$ from (3.7) and (3.8) are under the expectation the same, thus we have to concentrate only on the other summands. Hence choose

$$\phi_{(k,l)}(s) = \frac{1}{q_{kl}^Y(\widehat{X_{s-}Z_{s-}})} \sum_{i\in I(k)} \sum_{j\in I(l)} e_j \sum_{\mu=1}^{d} \big(g_\mu \tilde{q}_{ij,\mu}^X + \sum_{\nu=1}^{d} \delta_{ij}^{\mu\nu} g_\nu q_{\mu\nu}^Z\big)(\widehat{X_{s-}Z_{s-}})_{i\mu} - \widehat{X_{s-}Z_{s-}},$$

thus equation of the expectations is attained. Summarizing we have found a $\mathcal{F}_t^Y$-representation of $\widehat{X_tZ_t}$, given by (see lemma A.6):

$$\begin{aligned} & \widehat{X_tZ_t} \qquad (3.9)\\ = \;& \widehat{X_0Z_0} + \int_0^t \sum_{i=1}^{n}\sum_{\mu=1}^{d} (e_iQ^Zg_\mu + (\tilde{Q}_\mu^X + \tilde{Q}_\mu^Z)e_ig_\mu)(\widehat{X_sZ_s})_{i\mu} ds \\ & + \int_0^t \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\mu=1}^{d}\sum_{\nu=1}^{d} \delta_{ij}^{\mu\nu}(e_j-e_i)(g_\nu-g_\mu)(\widehat{X_sZ_s})_{i\mu} q_{\mu\nu}^Z ds \\ & + \sum_{k=1}^{m}\sum_{l=1}^{m} \phi_{(k,l)}(s)\big(dN_s^Y(k,l) - q_{kl}^Y(\widehat{X_sZ_s})Y_s^k ds\big) \end{aligned}$$

which is exactly (3.6). □

Based on lemma 3.2 and lemma 3.4 we transform the filter equation (3.6) with the help of the operator $\mathcal{S}$ to the representation of p_t.

Theorem 3.5 *Define*

$$\begin{aligned} Bp &:= \mathcal{S}A(\widehat{XZ}) = \mathcal{A}\mathcal{S}\widehat{XZ} \\ \Phi_{kl}(p) &:= \mathcal{S}\phi_{(k,l)}(\widehat{XZ}) =: \left(\frac{1}{q_{kl}^Y(p)}\varphi_{(k,l)} - I\right)p \end{aligned}$$

then p_t is the unique solution of

$$\begin{aligned} dp_t &= Bp_t dt + \sum_{k=1}^{m}\sum_{l=1}^{m} \Phi_{kl}(p_{t-})\big(dN_t^Y(k,l) - q_{kl}^Y(p_t)Y_t^k dt\big) \\ &= \left(B - \sum_{k=1}^{m}\sum_{l=1}^{m}\Big(\varphi_{(k,l)} - q_{kl}^Y(p_t)I\Big)Y_t^k\right)p_t dt + \sum_{k=1}^{m}\sum_{l=1}^{m}\Phi_{kl}(p_{t-})dN_t^Y(k,l) \qquad (3.10) \end{aligned}$$

where I denotes the $nd \times nd$-unit matrix.

We note that B is independent of Y_t and consequently independent of the information structure. This is clear, since here only the X and Z terms are mixed. At the end of this section we discuss the behaviour of p_t and for this we introduce the following abbreviation

$$b(y,p) := \left(B - \sum_{k=1}^{m}\sum_{l=1}^{m}\Big(\varphi_{(k,l)} - q_{kl}^Y(p)I\Big)y_k\right)p. \qquad (3.11)$$

$b(y,p)$ describes the deterministic flow between two jumps of p_t and is bilinear-quadratic in p.

p_t as (unique) solution of (3.10) is a piecewise-deterministic process taking values in the nd-dimensional probability simplex $\triangle^{nd}$. If a jump of $N_t^Y(k,l)$ occurs at time τ (resulting from a change in the observation state Y_t from f_k to f_l), then p_t jumps with probability one to the new state

$$p_\tau = p_{\tau-} + \Phi_{kl}(p_{\tau-}). \tag{3.12}$$

If Y_t jumps at time τ from f_k to f_l (with $k \neq l$ under the assumption $m \geq 2$) we have $p_{\tau-}(i,\cdot) = 0$ for all $i \in I(l)$ and $p_\tau(i,\cdot) \geq 0$ (where at least one is > 0) for all $i \in I(l)$ and $p_\tau(i,\cdot) = 0$ for all $i \notin I(l)$. In particular each jump of Y_t leads to changes in p_t. Thus we have $\Phi_{kl}(p) \neq 0$ for $k \neq l$ and a one-to-one relation between the observation Y_t and the estimator p_t. As an immediate consequence we have

$$\mathcal{F}_t^Y = \mathcal{F}_t^p,$$

without any further assumptions as needed in Miller et al. (2005).

Let us specialize the representation formula (3.10) of p_t to the five special case introduced on page 11 and discuss the behaviour of it. Some results are presented in a previous work of the author (Winter (2007)).

Complete-Observation-Model
In the case of complete information, remember that we skip the environment process, we have $Y_t \equiv X_t$ and $p_t(i) := \mathbb{P}(X_t = e_i \mid \mathcal{F}_t^X)$ only takes values in $\{0,1\}$, this means p_t is always in a corner of $\triangle^n$. Thus we get if X_t jumps from e_i to e_j

$$\Phi_{ij}(p) = e_j - e_i \quad \text{and} \quad b(y,p) \equiv 0.$$

Group-Observation-Model
Again without environment process (Z_t) we observe the jump size of p_t by

$$\Phi_{kl}(p) = \frac{1}{q_{kl}^Y(p)} \begin{pmatrix} \sum_{i\in I(k)} \sum_{1\in I(l)} q_{i1}^X \, p_i \\ \vdots \\ \sum_{i\in I(k)} \sum_{n\in I(l)} q_{in}^X \, p_i \end{pmatrix} - p\,.$$

Hence the a-posteriori-probability after a jump is distributed in relation to the intensities for a jump from f_k to f_l. In particular for the 0-1-observation model, we get for jumps to f_1

$$\Phi_{21}(p) = e_1 - p.$$

Thus after a jump the probability of being in state e_1 is equal to 1, which has to be the case, since we have complete observation. Additionally $b(f_1, p) \equiv 0$ (hence the probability for e_1 remains 1 until we leave this observation state f_1) and for jumps from f_1 to f_2

$$\Phi_{12}(p) = \frac{1}{q_{12}^X + q_{13}^X + \ldots + q_{1n}^X} \begin{pmatrix} 0 \\ q_{12}^X \\ \vdots \\ q_{1n}^X \end{pmatrix} - e_1 \,.$$

No-Observation-Model
In the case of no information about the state process (X_t) no jumps of (Y_t) occur, since $q_{kl}^Y(p) \equiv 0$, and therefore no jumps of (p_t) occur. Accordingly we get

$$b(y, p) = Q^X p,$$

thus the filter equation is equal to Kolmogorov's backward differential equation.

Hidden-Markov-Model
In the (classical) Hidden-Markov-Model the jump size is determined by

$$\Phi_{ij}(p) = \frac{1}{\sum_{r=1}^{d} \tilde{q}_{ij,r}^X \, p(r)} \begin{pmatrix} \tilde{q}_{ij,1}^X \, p(1) \\ \vdots \\ \tilde{q}_{ij,d}^X \, p(d) \end{pmatrix} - p.$$

since the jumps of (Z_t) has no influence to (X_t). If the jumps of (Z_t) interact with the jumps of (X_t) then the form of $\Phi_{ij}(p)$ is similar since

$$\Phi_{ij}(p) = \frac{1}{\sum_{r=1}^{d} q_{ij,r}^X \, p(r)} \begin{pmatrix} q_{ij,1}^X \, p(1) \\ \vdots \\ q_{ij,d}^X \, p(d) \end{pmatrix} - p.$$

but now $\tilde{q}_{ij,\mu}^X$ has been replaced by $q_{ij,\mu}^X := \tilde{q}_{ij,\mu}^X + \sum_{\nu=1}^{d} \delta_{ij}^{\mu\nu} q_{\mu\nu}^Z$ which pays attention to the jumps of (Z_t). In both cases the post-jump state is relative to the estimated jump intensities.

Bayesian-Model
With $Q^Z \equiv 0$ and from Z_0 only $P(Z_0 = z) = p_0$ is known (where $z \in S_Z$) and with $I(k) = \{k\}$ (complete observation about the process X_t), the jump behaviour is described by

$$\Phi_{ij}(p) = \frac{1}{\sum_{r=1}^{d} q_{ij,r}^X p(r)} \begin{pmatrix} q_{ij,1}^X \, p(1) \\ \vdots \\ q_{ij,n}^X \, p(n) \end{pmatrix} - p.$$

thus the post-jump state in each component is relative to the estimated intensities.

The finer the information structure, the "better" the estimator process as the following result demonstrates.

Theorem 3.6 *Let $(I(k), k = 1, \ldots, m)$ be a finer information structure than $(I'(k), k' = 1, \ldots, m')$ and denote by (Y_t) and (Y'_t) the corresponding observation processes, then $\mathbb{P}(X_t = e_i, Z_t = g_\mu \mid \mathcal{F}_t^{Y'})$ is measurable with respect to $\mathcal{F}_t^Y$.*

Proof: This statement is a direct consequence of $\mathcal{F}_t^{Y'} \subset \mathcal{F}_t^Y$ as stated in theorem 2.7. □

Remark 3.7 *The extension to the case of countable state spaces S_Z and S_X is straightforward under the assumption, that the intensity matrices are all conservative and the diagonal elements are finite (remember remark 2.3).*

Considering the stochastic differential equation (3.10) for the estimator p_t we see that it is nonlinear in p. But the existence of a solution is always guaranteed by the strong connection between the estimator p and the conditional expectation $\mathbb{E}[X_t Z_t \mid \mathcal{F}_t^Y]$ (which exists) given in lemma 3.2. Additionally, p_t is strongly connected to the non-normalized estimator process q_t as defined in Elliott et al. (1997) in the following way. There they define an equivalent probability measure Q such that $N_t^Y(k,l)$ are standard Poisson processes under Q. For this purpose define the Girsanov-density

$$L_t := \left(\prod_{0<s\leq t} \sum_{l=1}^{m} q_{kl}^Y(Z_s, X_s) \triangle N_s^Y(Y_{s-}, l) \right) \exp\left\{ \int_0^t \left(1 - \sum_{l=1}^{m} q_{Y_{s-}l}^Y(Z_s, X_s) \right) ds \right\}$$

under the assumption that L_t is a $\mathbb{P}$-martingale. Notice that L_t is the stochastic exponential of $\int_0^t (1 - \sum_{l=1}^{m} q_{Y_{s-}l}^Y(Z_s, X_s)) dM_s^Y$. Then the relation between the measures $\mathbb{P}$ and Q is given by

$$\mathbb{E}_Q \left[\frac{d\mathbb{P}}{dQ} \mid \mathcal{F}_t \right] = L_t.$$

Define then

$$q_t(i, \mu) := \mathbb{E}_Q \left[L_t (X_t Z_t)_{i\mu} \mid \mathcal{F}_t^Y \right]$$

and we state the analogon to theorem 3.5 for q_t.

Theorem 3.8

a) *q_t is (the unique) solution of the following linear stochastic differential equation under Q (so-called Zakai-equation), where the jump processes $N_t^Y(k,l)$ are standard Poisson processes*

$$dq_t = B q_t dt + \sum_{k=1}^{m} \sum_{l=1}^{m} \left(\varphi_{(k,l)} - I \right) q_t \left(dN_t^Y(k,l) - dt \right).$$

b) It holds:

$$p_t(i,\mu) = \frac{q_t(i,\mu)}{\sum_{j=1}^{n}\sum_{\nu=1}^{d} q_t(j,\nu)}$$

and due to this relation $q_t(i,\mu)$ is called unnormalized estimator process.

Denote by τ_n the jump times of the observation process (Y_t). These are the same jump times as of (p_t), see (3.10). Assume $Y_{\tau_n} = y, p_{\tau_n} = p$, then up to the next jump τ_{n+1} of Y_t the process p_t evolves according to

$$\begin{cases} \dot{p} &= b(y,p) \\ p_0 &= p \end{cases} \tag{3.13}$$

remember (3.11). Denote by $\phi_t(p)$ the solution of this deterministic partial differential equation. Notice that every y results in another $b(y,p)$ and hence in another $\phi_t(p)$, but we neglect this in the notation. Then after a jump of Y_t at time τ the estimator process p_t is equal to $\phi_{t-\tau}(p_\tau)$ up to the next jump as stated next.

Theorem 3.9

a) For $t \in [\tau_n, \tau_{n+1})$ it holds under $Y_t = y$ that $p_t = \phi_{t-\tau_n}(p_{\tau_n})$.

b) $p \mapsto b(y,p)$ is Lipschitz continuous.

c) $t \mapsto \phi_t(p)$ is Lipschitz continuous.

Proof:

a) The statement follows directly from the representation theorem 3.5.

b) $\triangle^{nd}$ is a convex and compact subset of $\mathbb{R}^{nd}$ and $p \mapsto b(y,p)$ is a bilinear-quadratic function and in particular continuously differentiable. As a consequence $p \mapsto b(y,p)$ is Lipschitz continuous on $\triangle^{nd}$.

c) Since $\frac{\partial}{\partial t}\phi_t(p) = b(y,p)$ and $b(y,p)$ is continuous by b) on the compact set $\triangle^{nd}$ the Lipschitz continuity is an immediate consequence.

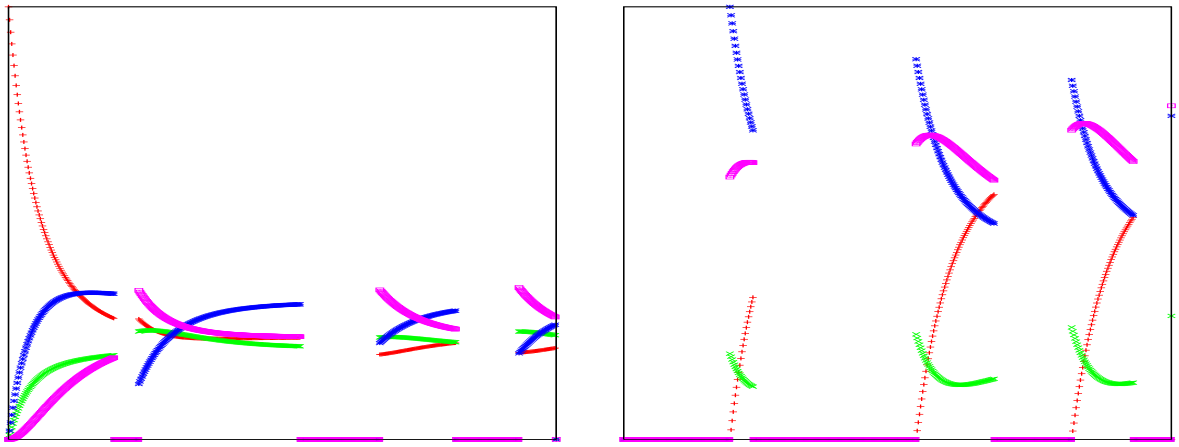
□

By the definition of the information structure and the jump behaviour of p_t described in (3.12) it holds for the conditional probabilities between two jumps, this means for $t \in [\tau_n, \tau_{n+1})$, if the observation process Y_t is in f_k that

$$p_t(i,\mu) = \phi_{t-\tau_n}(p_{\tau_n})(i,\mu) \begin{cases} > 0 & i \in I(k) \\ = 0 & i \notin I(k) \end{cases}$$

Thus we see that between two jumps the probability measure of the marginal probability $p_t(i,\cdot)$ is concentrated on states $i \in I(k)$. This is reasonable, since the observations tell, that X_t has to be in a state $i \in I(k)$.

Between two jumps the conditional probability $p_t = \phi_{t-\tau_n}(p_{\tau_n})$ moves into an equilibrium solution of $\dot{p} = b(y,p)$. Under some mild conditions this is the stationary distribution of (X_t, Z_t) restricted on $I(k)$. The following figure 2 demonstrates this behaviour. On the x-axis the time is marked, on the y-axis the interval $[0,1]$. The different lines are the conditional probabilities $p_t(i,\mu)$. We do not want to go in more detail, since in the following sections the estimate process p_t (in particular the intensities) depends on u. Hence it is quite hard to talk about a stationary distribution. But more details can be found in Davis (1993) and Braun (1993).



p on $I(1)$ p on $I(2)$

Figure 2: $d = 2$, $n = 4$ and $I(1) = \{1,2\}, I(2) = \{3,4\}$.

We have seen above that the random behaviour of p_t is completely described by the Poisson process $N_t^Y(k,l)$. The time between two jumps of $N_t^Y(k,l)$ is $\exp(q_{kl}^Y(p)Y^k)$-distributed. If there occurs a jump the jump-size of $N_t^Y(k,l)$ is equal to 1 and by construction of p_t its (deterministic) jump size is $\Phi_{kl}(p_{t-})$. Our next goal is to find an equivalent process construction, where the distribution function of the sojourn times is independent of the states of Y_t. This technique is called uniformization technique and will be helpful in section 4.2 and 5.

Define for fixed (k,l) with $\tau_0^{kl} := 0$ a sequence $(\tau_{n+1}^{kl} - \tau_n^{kl})$ of iid random variables which are $\exp(\alpha)$-distributed, where the uniformization parameter α is greater or equal to all diagonal elements of $Q^Y(Z,X)$, hence set

$$\alpha := \sup_{k,\mu,i} \left\{ q_k^Y(g_\mu, e_i) \right\}.$$

Furthermore define a sequence of Bernoulli($\frac{1}{\alpha} q^Y_{kl}(\phi_{t-\tau^{kl}_n}(p_{\tau^{kl}_n}))Y^k_{\tau^{kl}_n}$)-distributed random variables (W^{kl}_n). Then we are able to express $N^Y_t(k,l)$ in dependence of (W^{kl}_n) and (τ^{kl}_n).

Lemma 3.10 *It holds:*

$$N^Y_t(k,l) \stackrel{d}{=} \sum_{\substack{n=0 \\ \tau^{kl}_n \le t}}^{\infty} W^{kl}_n = \sum_{n=0}^{\tau^{kl}(t)} W^{kl}_n,$$

where $\tau^{kl}(t) := \sup\{n \ge 0 \mid \tau^{kl}_n \le t\}$.

Proof: This result is well-known from the Markov-chain theory, see Massey and Whitt (1998). □

From the last theorem we can rewrite our stochastic differential equation (3.10) for p_t as

$$dp_t = b(Y_t, p_t)dt + \sum_{k=1}^{m}\sum_{l=1}^{m} \Phi_{kl}(p_{t-})d\left(\sum_{n=0}^{\tau^{kl}(t)} W^{kl}_n\right) \tag{3.14}$$

The interpretation is the following: the distribution of the sojourn times is now independent of the state process Y_t and the estimate p_t. But there are more jumps of the underlying stochastic system with $\exp(\alpha)$-distribution compared to the former one. To compensate for these additional jumps, we have to skip some jumps of W, in the sense that these jumps do not lead to changes in p_t. This is done with the help of the Bernoulli-distribution. Characterization (3.14) is advantageously for simulations.

3.2 The Reduced Problem

After calculating a representation for our estimator (p_t) we express our cost criterion as a function of p_t, such that it is $\mathcal{F}^Y_t$-measurable, in particular it depends only on the available information. Note that (p_t) depends now on the control process u, since all intensities (and therefore $N^Y_t(k,l)$) may be controllable, see the construction in section 2.2. Hence the estimator process is influenced by the control process. This feature raises the level of difficulty significantly. In many other applications as for example in finance models the randomness in the system, given by a Brownian motion for example, is independent of the control. Consequently the estimators (for example for unknown parameters) are also independent of the control. Similar transformations for a time-discrete setting are pointed out in Bhulai (2002), Koole (1998) and Hernandez-Lerma (1989). The next theorem states that the expected discounted costs as a function of (Z_t, X_t, Y_t) are the same as of (Y_t, p_t).

Theorem 3.11 *It holds for all* $u \in \mathcal{U}$*:*

$$\mathbb{E}_u\Big[\int_0^\infty e^{-\beta s} g(Z_s, X_s, Y_s, u_s)ds\Big] = \mathbb{E}_u\Big[\int_0^\infty e^{-\beta s} g(p_s, Y_s, u_s)ds\Big], \tag{3.15}$$

where

$$g(p_t, y, u) := \mathbb{E}_u[g(z, x, y, u) \mid \mathcal{F}_t^Y] = \sum_{i=1}^{n} \sum_{\mu=1}^{d} g(g_\mu, e_i, y, u) p_t(i, \mu).$$

Proof: Since $g \geq 0$ we have by Fubini

$$\mathbb{E}_u\Big[\int_0^\infty e^{-\beta s} g(Z_s, X_s, Y_s, u_s) ds \Big] = \lim_{T \to \infty} \mathbb{E}_u\Big[\int_0^T e^{-\beta s} g(Z_s, X_s, Y_s, u_s) ds \Big].$$

With Wong and Hajek (1985) we conclude

$$\begin{aligned} \mathbb{E}_u\Big[\int_0^T e^{-\beta s} g(Z_s, X_s, Y_s, u_s) ds \Big] &= \mathbb{E}_u\Big[\mathbb{E}_u\Big\{ \int_0^T e^{-\beta s} g(Z_s, X_s, Y_s, u_s) ds \mid \mathcal{F}_T^Y \Big\} \Big] \\ &= \mathbb{E}_u\Big[\int_0^T \mathbb{E}_u\Big\{ e^{-\beta s} g(Z_s, X_s, Y_s, u_s) \mid \mathcal{F}_s^Y \Big\} ds \Big] \\ &= \mathbb{E}_u\Big[\int_0^T e^{-\beta s} g(p_s, Y_s, u_s) ds \Big] \end{aligned}$$

and the assertion follows. □

After transforming the objective function and the unobservable state process into estimated counterparts, we define the separated (transformed/reduced) problem based on (P). As mentioned earlier the separated control problem is with complete information, since all functions therein are $\mathcal{F}_t^Y$-adapted. That means they depend only on information available through $(Y_s)_{0 \leq s \leq t}$ up to the current time t. Define

$$(P_{\text{red}}) \begin{cases} \mathbb{E}_u\big[\int_0^\infty e^{-\beta s} g(p_s, Y_s, u_s) ds \big] \to \min \\ dp_t = b(u_t, Y_t, p_t) dt + \sum_{k=1}^{m} \sum_{l=1}^{m} \Phi_{kl}(u_t, p_{t-}) dN_t^Y(k, l) \\ dY_t = Q^Y(u_t, p_t) Y_t dt + d\widehat{M}_t^Y \\ (p_0, Y_0) = (p, y_0) \\ u \in \mathcal{U}, \end{cases}$$

where $b(u, y, p)$ is defined as in (3.13) in the controlled sense. Furthermore define $p := \mathbb{P}(Z_0 = z_0, X_0 = x_0 \mid \mathcal{F}_0^Y)$. We note that the set of admissible controls $\mathcal{U}$ is the same as for problem (P), since controls are only allowed to depend on the observation $\mathcal{F}_t^Y$. We introduce the following abbreviation:

$$\begin{aligned} J(y, p; u) &:= \mathbb{E}_u \left[\int_0^\infty e^{-\beta s} g(p_s, Y_s, u_s) ds \mid Y_0 = y, p_0 = p \right] \\ J(y, p) &:= \inf_{u \in \mathcal{U}} J(y, p; u) \end{aligned}$$

and we call a control process $u^* \in \mathcal{U}$ optimal if and only if $J(y, p, u^*) = J(y, p)$, where $J(y, p)$ is the minimal expected discounted cost over an infinite horizon starting in (y, p) (called the optimal value of (P_{red})).

In order to complete the reduction we have to prove that the objective function in the original (incomplete information) model and in the transformed (complete information) model are the same. This step is very often omitted in the literature, although it is not hard to prove.

Theorem 3.12 *Understand $\mathcal{U}$ as $\mathcal{U}[t,\infty)$, then it holds for all $t \geq 0$:*

a) $\mathbb{E}_u\Big[\int_t^\infty e^{-\beta s} g(Z_s, X_s, Y_s, u_s)ds \mid \mathcal{F}_t^Y\Big] = \mathbb{E}_u\Big[\int_t^\infty e^{-\beta s} g(p_s, Y_s, u_s)ds \mid \mathcal{F}_t^Y\Big] \quad \forall u \in \mathcal{U}$

b) $\inf\limits_{u\in\mathcal{U}} \mathbb{E}_u\Big[\int_t^\infty e^{-\beta s} g(Z_s, X_s, Y_s, u_s)ds \mid \mathcal{F}_t^Y\Big] = \inf\limits_{u\in\mathcal{U}} \mathbb{E}_u\Big[\int_t^\infty e^{-\beta s} g(p_s, Y_s, u_s)ds \mid \mathcal{F}_t^Y\Big].$

Proof: Recall that the set of admissible controls $\mathcal{U}$ for (P) and (P_{red}) are the same, then part a) follows as in theorem 3.11. Part b) is a direct consequence of part a). □

After deriving the connection between the objective functions we state the connection between the optimal controls, which is given by: the optimal control of the reduced model is an optimal control for the incomplete information model (and vice versa).

Theorem 3.13 *The following assertions are immediate consequences of theorem 3.12:*

a) *$u = (u_t)$ is optimal for $(P_{\text{red}}) \Longleftrightarrow u = (u_t)$ is optimal for (P)*

b) *The optimal values of (P_{red}) and (P) are the same.*

Theorem 3.12 simplifies in the case of Markovian controls to the following:

Corollary 3.14 *If the (optimal) control (u_t^*) is Markovian and (p_t^*, Y_t^*) is the corresponding state process, then:*

$$\mathbb{E}_u\left[\int_t^\infty e^{-\beta s} g(p_s^*, Y_s^*, u_s^*) \mid \mathcal{F}_t^{Y^*}\right] = \mathbb{E}_u\left[\int_t^\infty e^{-\beta s} g(p_s^*, Y_s^*, u_s^*) \mid p_t^*, Y_t^*\right].$$

We have seen that the two problems (P) and (P_{red}) are strongly connected to each other. If we can solve the reduced problem we have a solution for the original problem (P). In particular we have proven that $J(z,x,y;u) = J(y,p;u)$ and $J(z,x,y) = J(y,p)$. In general the existence of an optimal solution for one of these two problems is not guaranteed, but we will come back to this question in section 4.2. Before we state some properties of the value function $J(y,p;u)$ and $J(y,p)$.

Theorem 3.15

a) *For all $u \in \mathcal{U}$ holds*

$$J(y,p;u) = \sum_{i=1}^{n}\sum_{r=1}^{d} J(g_r, e_i, y; u)p(i,r)$$

b) For the optimal value function holds

$$J(y,p) \geq \sum_{i=1}^{n}\sum_{r=1}^{d} J(g_r, e_i, y)p(i,r)$$

c) $p \mapsto J(y,p)$ is concave.

Proof:

a) The equation is obvious by making use of the definition of $J(z,x,y;u)$ and $J(y,p;u)$ with the help of the conditional expectation.

b) With a) we conclude:

$$\begin{aligned} J(y,p) &= \inf_{u\in\mathcal{U}} J(y,p;u) = \inf_{u\in\mathcal{U}} \sum_{i=1}^{n}\sum_{r=1}^{d} J(g_r,e_i,y;u)p(i,r) \\ &\geq \sum_{i=1}^{n}\sum_{r=1}^{d} \inf_{u\in\mathcal{U}} \{J(g_r,e_i,y;u)\}\, p(i,r) = \sum_{i=1}^{n}\sum_{r=1}^{d} J(g_r,e_i,y)p(i,r). \end{aligned}$$

c) Again from a) we get for $p \in \triangle^{nd}, q \in \triangle^{nd}$ and $\rho \in [0,1]$:

$$\begin{aligned} & J(y,\rho p + (1-\rho)q) = \inf_{u\in\mathcal{U}} J(y,\rho p + (1-\rho)q;u) \\ = & \inf_{u\in\mathcal{U}} \left\{ \sum_{i=1}^{n}\sum_{r=1}^{d} J(g_r,e_i,y;u)(\rho p(i,r) + (1-\rho)q(i,r)) \right\} \\ \geq & \inf_{u\in\mathcal{U}} \left\{ \sum_{i=1}^{n}\sum_{r=1}^{d} J(g_r,e_i,y;u)\rho p(i,r) \right\} + \inf_{u\in\mathcal{U}} \left\{ \sum_{i=1}^{n}\sum_{r=1}^{d} J(g_r,e_i,y;u)(1-\rho)q(i,r) \right\} \\ = & \rho J(y,p) + (1-\rho)J(y,q) \end{aligned}$$

□

Corollary 3.16 *From part c) of the last theorem it follows that $p \mapsto J(y,p)$ is locally Lipschitz continuous.*

Proof: Since $J(y,p)$ is concave in p the assertion is an immediate consequence from analysis as stated for example in Rockafellar (1996). □

4 Solving the Reduced Model

In this chapter we introduce two different solution techniques for the reduced model, which was introduced in section 3.2 as

$$(P_{\text{red}}) \begin{cases} \mathbb{E}_u\big[\int_0^\infty e^{-\beta s} g(p_s, Y_s, u_s) ds\big] \to \min \\ dp_t = b(u_t, Y_t, p_t)dt + \sum_{k=1}^{m}\sum_{l=1}^{m} \Phi_{kl}(u_t, p_{t-}) dN_t^Y(k,l) \\ dY_t = Q^Y(u_t, p_t) Y_t dt + d\widehat{M}_t^Y \\ (p_0, Y_0) = (p, y_0) \\ u \in \mathcal{U} \end{cases}$$

We show that these procedures at the end are connected but have completely different fundamentals. The first solution technique is a generalization of the classical verification technique using the Hamilton-Jacobi-Bellman equation (HJB). It is well known that if the value function is sufficient differentiable it satisfies the HJB-equation and an optimal control can be computed with the help of this HJB-equation. But the differentiability condition is a very strong one and various authors tried to overcome this difficulty for example by viscosity solutions (see Fleming and Soner (1993)) or numerical approaches (see Kushner and Dupuis (2001)). We use a weaker form of differentiability, introduced by Clarke (1983), and extend the HJB-equation and the verification technique with the help of the Clarke derivative in section 4.1. We give necessary and sufficient conditions for an optimal control.

In section 4.2 we use the piecewise-deterministic behaviour of our transformed state process (Y_t, p_t) and utilize an idea of Davis (1993). We define a time-discrete Markovian-Decision-Process (MDP), where every action is a function of the state after the last jump and of the time elapsed since the last jump. Here we state an existence theorem for an optimal control. We generalize present results to discounted and uniformized models. These extensions are in our opinion much more practicable for the computation of optimal strategies and for proving properties of optimal strategies. This procedure together with the generalized HJB-approach will help us in section 5 to characterize optimal controls.

Notice that there are more than these two solution techniques. One is the maximum principle, which is connected to the verification technique. It is the extension of Pontryagin's maximum principle to the field of stochastic optimization problems. It does not use the special structure of the piecewise-deterministic process. Therefore it can be applied to various optimization problems. Since this technique makes use of the Lagrangian function and one has to solve the so-called adjoint equation, which is a deterministic partial differential equations, very often computational problems arise. For details we refer to Øksendal and Sulem (2005), Framstad et al. (2004), Rishel (1978) and Haussmann (1986).

Another approach, which is more technical, is the martingale optimality. It claims, that the value function for a fixed control is always a submartingale. But it is even a martingale if and only if the control is optimal. This result is very general and is applied often in the context of financial mathematics (see e.g. Karatzas and Shreve (2001)).

4.1 The Generalized HJB-Equation and Verification Technique

The classical verification technique with the help of the Hamilton-Jacobi-Bellman-equation goes back to Bellman (see e.g. Bellman (1977)) and is well understood, see for example Fleming and Rishel (1975), Yong and Zhou (1999) or Øksendal and Sulem (2005). The HJB-equation can be defined for every control problem and gives under some assumptions a characterization of the optimal value function (remember section 3.2)

$$J(y,p) = \inf_{u\in\mathcal{U}} \mathbb{E}_u \left[\int_0^\infty e^{-\beta s} g(p_s, Y_s, u_s) ds \mid Y_0 = y, p_0 = p \right].$$

In other words: $J(y,p)$ is the minimal expected discounted cost over an infinite horizon, when the process (Y_t, p_t) starts at time $t = 0$ in state $(y,p) \in S_Y \times \triangle^{nd}$.

The following theorem is well-known as Bellman's principle. The proof is standard and can be found for example in Gihman and Skorohod (1979) or Hanson (2007).

Theorem 4.1 *For all $\mathcal{F}_t^Y$-stopping times $\tau \geq t$ it holds*

$$e^{-\beta t} J(y,p) \quad = \quad \inf_{u\in\mathcal{U}[t,\tau)} \mathbb{E}_u \left[\int_t^\tau e^{-\beta s} g(p_s, Y_s, u_s) ds + e^{-\beta\tau} J(Y_\tau, p_\tau) \mid Y_t = y, p_t = p \right],$$

where $\mathcal{U}[t,\tau)$ denotes the set of admissible controls in the interval $[t,\tau)$.

This result will be used in the proof of theorem 4.3 where we claim that the value function is a solution of the HJB-equation. Additionally, it is useful for the proof of corollary 2.14 we give next.

Proof of corollary 2.14: Assume $Y_t = f_k$ and define $\tau := \inf\{s > t \mid Y_s \neq f_k\}$ as the first jump time point after time t. Consider $(u_s^*)_{s\in[t,\tau)}$ under the condition $Y_t = f_k$. Then we know that u_s^* is $\mathcal{F}_s^Y$-predictable for $s \in [t,\tau)$, since u_s^* is the same for all $i \in I(k)$. Furthermore we know that if v^* is an optimal control of

$$\left\{ \begin{array}{rcl} \int_0^T g(x, v_s) ds & \rightarrow & \min \\ x & \in & A \end{array} \right.$$

which is independent of x then v^* is also optimal for

$$\left\{ \begin{array}{rcl} \int_0^T \int_A g(x, v_s) \mu(dx) ds & \rightarrow & \min \\ \mu & \in & P(A) \end{array} \right.$$

where $P(A)$ is set of all probability measures over A. We conclude with Bellman's principle

of theorem 4.1 with $y = f_k$, that

$$\begin{aligned}
& e^{-\beta t} J(y,p) \\
= & \inf_{u \in \mathcal{U}[t,\tau)} \mathbb{E}_u \left[\int_t^\tau e^{-\beta s} g(p_s, u_s) ds + e^{-\beta \tau} J(Y_\tau, p_\tau) \mid Y_t = y, p_t = p \right] \\
= & \inf_{u \in \mathcal{U}[t,\tau)} \mathbb{E}_u \left[\int_t^\tau e^{-\beta s} \sum_{i=1}^n g(\cdot, e_i, u_s) p_s(i,\cdot) ds + e^{-\beta \tau} J(Y_\tau, p_\tau) \mid Y_t = y, p_t = p \right] \\
= & \inf_{u \in \mathcal{U}[t,\tau)} \mathbb{E}_u \left[\int_t^\tau e^{-\beta s} \sum_{i \in I(k)} g(\cdot, e_i, u_s) p_s(i,\cdot) ds + e^{-\beta \tau} J(Y_\tau, p_\tau) \mid Y_t = y, p_t = p \right] \\
= & \mathbb{E}_{u^*} \left[\int_t^\tau e^{-\beta s} \sum_{i \in I(k)} g(\cdot, e_i, u_s^*) p_s(i,\cdot) ds + e^{-\beta \tau} J(Y_\tau, p_\tau) \mid Y_t = y, p_t = p \right] \\
= & \mathbb{E}_{u^*} \left[\int_t^\tau e^{-\beta s} g(p_s, u_s^*) ds + e^{-\beta \tau} J(Y_\tau, p_\tau) \mid Y_t = y, p_t = p \right]
\end{aligned}$$

and therefore $(u_s^*)_{s \in [t,\tau)}$ is optimal under $Y_t = f_k$. □

Instead of requiring strict differentiability of the value function as in the classical HJB-theory it can be shown, that it is sufficient to assume that the value function is locally Lipschitz continuity. This generalized technique was first introduced by Clarke (1983) and then extended by Davis (1993). We first define the Clarke derivative, which is an extension of the classical theory of differentiability.

Let $f : \mathbb{R}^b \to \mathbb{R}$ be a locally Lipschitz continuous function. Then for $x, y \in \mathbb{R}^b$ the upper generalized directional derivative of f at x in direction y is defined by

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

The lower generalized directional derivative of f at x in direction y is defined analogously by

$$f_0(x;y) := \liminf_{\substack{z \to x \\ \varepsilon \to 0}} \frac{f(z+\varepsilon y) - f(z)}{\varepsilon}.$$

If $f_0(x;y) = f^0(x;y)$ then $\lim_{z \to x, \varepsilon \to 0} \frac{f(z+\varepsilon y) - f(z)}{\varepsilon}$ exists and f is differentiable in x in direction y and everything breaks down to the well-known directional derivative.

The Clarke generalized gradient of f at x is defined by

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

which is a nonempty, convex and compact subset of $\mathbb{R}^b$. We want to understand ξ as row vector. If $f(x)$ is differentiable in x with derivative $f'(x)$ then $\partial f(x) = \{f'(x)\}$. Due to this the classical HJB-approach is included here if the value function is (piecewise) differentiable. It holds further

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

By the local Lipschitz continuity we conclude that f is differentiable almost everywhere and we can find for every $x \in \mathbb{R}^b$ a sequence (x_n) with $x_n \in \mathbb{R}^b$ 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 $\nabla f(x_n)$, that means

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

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 $f^0(x;y) = f'(x;y)$. By Clarke (1983) every concave function f (which is even locally Lipschitz by Rockafellar (1996)) is regular.

Finally we mention one important case of the chain rule, the computation procedure for combined functions. General formulas can be found in Clarke (1983).

Lemma 4.2 *Let $g : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^m \to \mathbb{R}^n$ are locally Lipschitz continuous. Assume g is regular and h is strictly differentiable, then for $f(x) := g(h(x))$ it holds*

$$\partial f(x) = \partial g(h(x)) \circ h'(x) \tag{4.1}$$

where we want to understand $\partial g(h(x)) = \partial g(z) \mid_{z=h(x)}$. The meaning of (4.1) is that every element $\xi \in \partial f(x)$ can be represented as a composition of a map $\psi \in \partial g(h(x))$ and $h'(x)$.

In the one-dimensional case $b = 1$ only two directions are possible: to the right and to the left. Hence we will sometimes speak of the upper generalized right hand side derivative in direction y (if $y > 0$) when we consider $f^0(x;y)$ (similar for the other cases). If $f^0(x;1) = f_0(x;1)$ then f is differentiable on the right hand side in the usual way. If additionally $f^0(x;-1) = f_0(x;-1)$ then f is differentiable in x.

Thus we state the generalized HJB-equation for a locally Lipschitz continuous function W as

$$\begin{aligned} & \beta W(y,p) \qquad\qquad (4.2) \\ = & \inf_{\substack{\xi \in \partial_p W(y,p) \\ u \in U}} \left\{ g(p,y,u) + \xi b(u,y,p) + \sum_{l=1}^{m} (W(f_l, p + \Phi_{yl}(u,p)) - W(y,p)) q_{yl}^Y(u,p) \right\} \end{aligned}$$

where $\partial_p W(y,p)$ is the Clarke generalized gradient with respect to p. The next theorem is the justification of the HJB-equation, since it states that the value function $J(y,p)$ is solution of it. Additionally, it states a necessary condition for an optimal control.

Theorem 4.3 *It holds:*

a) *The value function $J(y,p)$ satisfies the generalized HJB-equation for all $(y,p) \in S_Y \times \triangle^{nd}$.*

b) *If there exists an optimal control (u_t^*) with corresponding state process (Y_t^*, p_t^*) then*

$$\beta J(Y_t^*, p_t^*) = \inf_{\xi \in \partial_p J(Y_t^*, p_t^*)} \Big\{ g(p_t^*, Y_t^*, u_t^*) + \xi b(u_t^*, Y_t^*, p_t^*) + \sum_{l=1}^{m} (J(f_l, p_t^* + \Phi_{Y_t^* l}(u_t^*, p_t^*)) - J(Y_t^*, p_t^*)) q^Y_{Y_t^* l}(u_t^*, p_t^*) \Big\}$$

for almost all $t \geq 0$.

Proof:

a) Denote by τ_n the jump times of Y_t, especially this are the jump times of p_t too. Since $t \mapsto e^{-\beta t} J(Y_t, \phi_t^u(p))$ is locally Lipschitz continuous (since $p \mapsto J(y,p)$ is locally Lipschitz by corollary 3.16, $t \mapsto \phi_t^u(p)$ by theorem 3.9 and $t \mapsto e^{-\beta t}$ is Lipschitz continuous for $t \in [0,\infty)$) there exists for all $0 =: \tau_0 < \tau_1 < \tau_2 < \dots$ a function $D\big(e^{-\beta s} J(Y_s, p_s)\big)$ such that

$$e^{-\beta \tau_i -} J(Y_{\tau_i -}, p_{\tau_i -}) - e^{-\beta \tau_{i-1}} J(Y_{\tau_{i-1}}, p_{\tau_{i-1}}) = \int_{\tau_{i-1}}^{\tau_i -} D\big(e^{-\beta s} J(Y_s, p_s)\big) ds.$$

Due to the local Lipschitz continuity of $e^{-\beta s} J(Y_s, p_s)$ the function $D\big(e^{-\beta s} J(Y_s, p_s)\big)$ may be chosen as its derivative with respect to s, which exists almost everywhere on $[0,\infty)$. Hence with theorem 3.9

$$\begin{aligned} D\big(e^{-\beta s} J(Y_s, p_s)\big) &= -\beta e^{-\beta s} J(Y_s, p_s) + e^{-\beta s} J_p(Y_s, p_s) \dot{\phi}_s^u(p_s) \\ &= e^{-\beta s} \Big(-\beta J(Y_s, p_s) + J_p(Y_s, p_s) b(u_s, Y_s, p_s) \Big). \end{aligned}$$

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

$$\begin{aligned} & e^{-\beta t} J(Y_t, p_t) \\ = \; & J(Y_0, p_0) + \int_0^t D\big(e^{-\beta s} J(Y_s, p_s)\big) ds + \sum_{0 < s \leq t} \big(e^{-\beta s} J(Y_s, p_s) - e^{-\beta s -} J(Y_{s-}, p_{s-}) \big) \\ = \; & J(Y_0, p_0) + \int_0^t D\big(e^{-\beta s} J(Y_s, p_s)\big) ds \\ & + \int_0^t e^{-\beta s} \sum_{l=1}^{m} \big(J(f_l, p_{s-} + \Phi_{Y_{s-} l}(u_s, p_{s-})) - J(Y_{s-}, p_{s-}) \big) dN_s^Y(Y_{s-}, f_l). \end{aligned}$$

Denote by τ the first jump time point after time t then we obtain from Bellman's principle in theorem 4.1 and by the considerations above with a time-shift from $[0,t]$ to $[t, t' \wedge \tau]$ that for all $u \in \mathcal{U}[t,\tau)$ and $t' > t$

$$\begin{aligned}
& e^{-\beta t} J(y,p) \\
\leq\ & \mathbb{E}_u \left[\int_t^{\tau \wedge t'} e^{-\beta s} g(p_s, Y_s, u_s) ds + e^{-\beta(\tau \wedge t')} J(Y_{\tau \wedge t'}, p_{\tau \wedge t'}) \mid Y_t = y, p_t = p \right] \\
=\ & \mathbb{E}_u \Bigg[\int_t^{\tau \wedge t'} e^{-\beta s} g(p_s, Y_s, u_s) ds + e^{-\beta t} J(Y_t, p_t) + \int_t^{\tau \wedge t'} D\big(e^{-\beta s} J(Y_s, p_s)\big) ds \\
& \quad + \int_t^{\tau \wedge t'} e^{-\beta s} \sum_{l=1}^{m} \big(J(f_l, p_{s-} + \Phi_{Y_{s-} l}(u_s, p_{s-})) - J(Y_{s-}, p_{s-}) \big) \cdot \\
& \qquad \cdot dN_s^Y(Y_{s-}, f_l) \mid Y_t = y, p_t = p \Bigg].
\end{aligned}$$

Since

$$D\big(e^{-\beta s} J(Y_s, p_s)\big) = e^{-\beta s} \big(-\beta J(Y_s, p_s) + DJ(Y_s, p_s) \big)$$

$Y_t = y, p_t = p$ and the intensity of $N_t^Y(Y_{s-}, Y_s)$ is given by $q^Y_{Y_{s-} Y_s}(u_s, p_{s-})$ we proceed to

$$\begin{aligned}
0 \ \leq\ & \mathbb{E}_u \Bigg[\int_t^{\tau \wedge t'} e^{-\beta s} \Big(g(p_s, Y_s, u_s) - \beta J(Y_s, p_s) + DJ(Y_s, p_s) \\
& \qquad + \sum_{l=1}^{m} \big(J(f_l, p_{s-} + \Phi_{Y_{s-} l}(u_s, p_{s-})) - J(Y_{s-}, p_{s-}) \big) \cdot \\
& \qquad \cdot q^Y_{Y_{s-} l}(u_s, p_{s-}) \Big) ds \mid Y_t = y, p_t = p \Bigg] \\
=:\ & \mathbb{E}_u \left[\int_t^{\tau \wedge t'} e^{-\beta s} HJ(Y_s, p_s, u_s) ds \mid Y_t = y, p_t = p \right].
\end{aligned}$$

Choose now a fixed strategy $\tilde{u}_s \equiv u$ for $s \in [t, t+\varepsilon)$, $\varepsilon > 0$, then

$$\begin{aligned}
0 \ \leq\ & \lim_{t' \downarrow t} \mathbb{E}_u \left[\frac{1}{t'-t} \int_t^{\tau \wedge t'} e^{-\beta s} HJ(Y_s, p_s, \tilde{u}_s) ds \mid Y_t = y, p_t = p \right] \\
=\ & \lim_{t' \downarrow t} \mathbb{E}_u \left[\frac{1}{t'-t} \int_t^{t'} e^{-\beta s} HJ(Y_s, p_s, \tilde{u}_s) ds \mid Y_t = y, p_t = p \right] \cdot \mathbb{P}_u(t' < \tau) \\
& + \lim_{t' \downarrow t} \mathbb{E}_u \left[\frac{1}{t'-t} \int_t^{\tau} e^{-\beta s} HJ(Y_s, p_s, \tilde{u}_s) ds \mid Y_t = y, p_t = p \right] \cdot \mathbb{P}_u(t' \geq \tau).
\end{aligned}$$

With $\alpha \geq \sup_{k,\mu,i} \sup_{u \in U} q_k^Y(u, g_\mu, e_i)$ we conclude

$$\mathbb{P}_u(t' \geq \tau) \leq 1 - e^{-\alpha(t'-t)} \to 0 \quad \text{for } t' \downarrow t.$$

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

$$0 \leq e^{-\beta t} HJ(Y_t, p_t, \tilde{u}_t) = e^{-\beta t} HJ(y,p,u) \quad \Rightarrow \quad 0 \leq \inf_{u \in U} HJ(y,p,u).$$

At points p where $J(y,p)$ is not differentiable in p the generalized gradient is given by

$$\partial_p J(y,p) = co\left\{ \lim_{n\to\infty} \nabla J(y, p_{t_n}) \mid p_{t_n} \to p_t = p \right\} = co\left\{ \lim_{n\to\infty} \nabla J(y, p_{t_n}) \mid t_n \to t \right\},$$

this means: every $\xi \in \partial_p J(y,p)$ is a convex combination of $\xi^m = \lim_{n\to\infty} \nabla J(y, p_{t_n^m})$ for sequences $t_n^m \to t$, along which $J(y,p)$ is differentiable. Since $p \mapsto J(y,p)$ is locally Lipschitz continuous (see corollary 3.16) we obtain with the chain rule for the Clarke derivative for all m that

$$\begin{aligned} 0 \quad \leq \quad & e^{-\beta t}\Big(g(p,y,u) - \beta J(y,p) + \xi^m b(u,y,p) \\ & \qquad + \sum_{l=1}^{m} \big(J(f_l, p + \Phi_{yl}(u,p)) - J(y,p) \big) q_{yl}^Y(u,p) \Big). \end{aligned}$$

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

$$0 \leq g(p,y,u) - \beta J(y,p) + \xi b(u,y,p) + \sum_{l=1}^{m} \big(J(f_l, p + \Phi_{yl}(u,p)) - J(y,p) \big) q_{yl}^Y(u,p).$$

Since $u \in U$ and $\xi \in \partial_p J(y,p)$ were chosen arbitrarily we conclude

$$\begin{aligned} 0 \quad \leq \quad & \inf_{\substack{\xi \in \partial_p J(y,p) \\ u \in U}} \Big\{ g(p,y,u) - \beta J(y,p) + \xi b(u,y,p) \\ & \qquad + \sum_{l=1}^{m} \big(J(f_l, p + \Phi_{yl}(u,p)) - J(y,p) \big) q_{yl}^Y(u,p) \Big\}. \end{aligned}$$

On the other hand for $\varepsilon > 0$ and $0 < t < t' < \infty$ with $t - t' > 0$ small enough there exists a strategy u^ε with corresponding state process (Y_t, p_t) such that we conclude from theorem 4.1

$$\begin{aligned} & e^{-\beta t} J(y,p) + \varepsilon(t'-t) \\ \geq \quad & \mathbb{E}_u \left[\int_t^{\tau \wedge t'} e^{-\beta s} g(p_s, Y_s, u_s^\varepsilon) ds + e^{-\beta(\tau \wedge t')} J(Y_{\tau \wedge t'}, p_{\tau \wedge t'}) \mid Y_t = y, p_t = p \right]. \end{aligned}$$

The same computations as before lead to

$$\begin{aligned}
\varepsilon &\geq \mathbb{E}_{u^\varepsilon}\left[\frac{1}{t'-t}\int_t^{\tau\wedge t'} e^{-\beta s} HJ(Y_s,p_s,u_s^\varepsilon)ds \mid Y_t=y,p_t=p\right] \\
&\geq \mathbb{E}_{u}\left[\frac{1}{t'-t}\int_t^{\tau\wedge t'} e^{-\beta s} \inf_{u_s\in U} HJ(Y_s,p_s,u_s)ds \mid Y_t=y,p_t=p\right] \\
&= \mathbb{E}_{u^*}\left[\frac{1}{t'-t}\int_t^{\tau\wedge t'} e^{-\beta s} HJ(Y_s,p_s,u_s^*)ds \mid Y_t=y,p_t=p\right],
\end{aligned}$$

where the existence of u_s^* is guaranteed since $u \mapsto HJ(y,p,u)$ is continuous and U compact. If $J(y,p)$ is differentiable in a point p we continue as before with

$$\begin{aligned}
\varepsilon &\geq \lim_{t'\downarrow t} \mathbb{E}_{u^*}\left[\frac{1}{t'-t}\int_t^{\tau\wedge t'} e^{-\beta s} HJ(Y_s,p_s,u_s^*)ds \mid Y_t=y,p_t=p\right] \\
&= e^{-\beta t} HJ(y,p,u_t^*) = e^{-\beta t}\inf_{u\in U} HJ(y,p,u)
\end{aligned}$$

from which we conclude, since $e^{-\beta t}>0$ and ε was arbitrarily that

$$0 \geq \inf_{u\in U} HJ(y,p,u).$$

If $J(y,p)$ is not differentiable in a point p we get that every $\xi \in \partial_p J(y,p)$ is a convex combination of ξ^m. With the same computations and with the help of the approximating sequence t_n^m as before we can show that

$$\begin{aligned}
0 \geq \inf_{\substack{\xi\in\partial_p J(y,p)\\ u\in U}} \Big\{ & g(p,y,u) - \beta J(y,p) + \xi b(u,y,p) \\
& + \sum_{l=1}^m \big(J(f_l, p+\Phi_{yl}(u,p)) - J(y,p)\big) q_{yl}^Y(u,p)\Big\}.
\end{aligned}$$

Altogether we obtain

$$\begin{aligned}
\beta J(y,p) = \inf_{\substack{\xi\in\partial_p J(y,p)\\ u\in U}} \Big\{ & g(p,y,u) + \xi b(u,y,p) \\
& + \sum_{l=1}^m \big(J(f_l, p+\Phi_{yl}(u,p)) - J(y,p)\big) q_{yl}^Y(u,p)\Big\},
\end{aligned}$$

which proves part a).

b) The necessary condition for an optimal control is proven by choosing the optimal control u_t^* instead of an arbitrary control in the beginning of the proof of a) and by choosing u_t^* instead of u^ε in the second part of the proof of a). Then overall equality holds.

□

Now we are in the position to formulate our generalized verification procedure for computing an optimal control.

Theorem 4.4

a) If $p \mapsto \tilde{J}(y,p)$ is locally Lipschitz continuous and satisfies the generalized HJB-equation (4.2) then $\tilde{J}(y,p) \leq J(y,p)$.

b) If $\tilde{J} : S_Y \times \triangle^{nd} \to \mathbb{R}$ is a locally Lipschitz continuous, regular (in p) solution of the generalized HJB-equation (4.2) and if there exists a $\left(u_t^* := u^*(Y_{t-}^*, p_{t-}^*)\right) \in \mathcal{U}$ with corresponding state process (Y_t^*, p_t^*) such that for almost all $t \geq 0$ a $\xi_t^* \in \partial_p \tilde{J}(Y_t^*, p_t^*)$ exists such that

$$\beta \tilde{J}(Y_t^*, p_t^*) = \Bigg\{ g(p_t^*, Y_t^*, u_t^*) + \xi_t^* b(u_t^*, Y_t^*, p_t^*) + \sum_{l=1}^{m} \left(\tilde{J}(f_l, p_t^* + \Phi_{Y_t^* l}(u_t^*, p_t^*)) - \tilde{J}(Y_t^*, p_t^*) \right) q^Y_{Y_t^* l}(u_t^*, p_t^*) \Bigg\}$$

then

(i) (u_t^*) is an optimal control

(ii) $J(y,p) = \tilde{J}(y,p)$ for all $(y,p) \in S_Y \times \triangle^{nd}$.

Proof:

a) Replace $J(y,p)$ by $\tilde{J}(y,p)$ which is locally Lipschitz continuous by assumption and conclude as in the proof of theorem 4.3 that

$$\begin{aligned}
& e^{-\beta t} \tilde{J}(Y_t, p_t) \\
= \; & \tilde{J}(y,p) + \int_0^t D\big(e^{-\beta s} \tilde{J}(Y_s, p_s)\big) ds + \sum_{0<s\leq t} \big(e^{-\beta s} \tilde{J}(Y_s, p_s) - e^{-\beta s-} \tilde{J}(Y_{s-}, p_{s-}) \big) \\
= \; & \tilde{J}(y,p) + \int_0^t D\big(e^{-\beta s} \tilde{J}(Y_s, p_s)\big) ds \\
& + \int_0^t e^{-\beta s} \sum_{l=1}^{m} \big(\tilde{J}(f_l, p_{s-} + \Phi_{Y_{s-} l}(u_s, p_{s-})) - \tilde{J}(Y_{s-}, p_{s-}) \big) dN_s^Y(Y_{s-}, f_l).
\end{aligned}$$

At those points p where $\tilde{J}(y,p)$ is differentiable (in particular $p_s = p_{s-}$) we know from the HJB-equation that

$$\begin{aligned} & D\big(e^{-\beta s}\tilde{J}(Y_s,p_s)\big) \\ \geq\ & e^{-\beta s}\Big(-g(p_s,Y_s,u_s) - \sum_{l=1}^{m}\big(\tilde{J}(f_l,p_{s-}+\Phi_{Y_{s-}l}(u_s,p_{s-})) - \tilde{J}(Y_{s-},p_{s-})\big)q^Y_{Y_{s-}l}(u_s,p_{s-})\Big) \end{aligned}$$

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

$$\begin{aligned} e^{-\beta t}\tilde{J}(Y_t,p_t) \ \geq\ & \tilde{J}(y,p) - \int_0^t e^{-\beta s} g(p_s,Y_s,u_s)ds \\ & + \int_0^t e^{-\beta s}\sum_{l=1}^{m}\big(\tilde{J}(f_l,p_{s-}+\Phi_{Y_{s-}l}(u_s,p_{s-})) - \tilde{J}(Y_{s-},p_{s-})\big)\cdot \\ & \qquad \cdot\big(dN^Y_s(Y_{s-},f_l) - q^Y_{Y_{s-}l}(u_s,p_{s-})ds\big). \end{aligned}$$

Let now t tend to infinity and note that $e^{-\beta t}\tilde{J}(Y_t,p_t)$ then tends to 0 we come to

$$\begin{aligned} 0 \ \geq\ & \tilde{J}(y,p) - \int_0^\infty e^{-\beta s} g(p_s,Y_s,u_s)ds \\ & + \int_0^\infty e^{-\beta s}\sum_{l=1}^{m}\big(\tilde{J}(f_l,p_{s-}+\Phi_{Y_{s-}l}(u_s,p_{s-})) - \tilde{J}(Y_{s-},p_{s-})\big)\cdot \\ & \qquad \cdot\big(dN^Y_s(Y_{s-},f_l) - q^Y_{Y_{s-}l}(u_s,p_{s-})ds\big). \end{aligned}$$

Taking expectation and remembering that the second integral is a martingale we finally get

$$\tilde{J}(y,p) \leq \mathbb{E}_u\left[\int_0^\infty e^{-\beta s} g(p_s,Y_s,u_s)ds\right],$$

which proves part a), since u was arbitrarily, consequently $\tilde{J}(y,p) \leq J(y,p)$.

b) Let t such that $\frac{\partial}{\partial t}\tilde{J}(Y_t^*,p_t^*)$ exists (which is the case almost everywhere due to the local Lipschitz continuity), then we get:

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

where the second equality is true due to the local Lipschitz continuity, the third due to the regularity and the inequality due to the properties of Clarke derivative. In particular for ξ_t^* it holds (since $p_t^* = p_{t-}^*$, since we consider t where $\frac{\partial}{\partial t}\tilde{J}(Y_t^*, p_t^*)$ exists)

$$\begin{aligned}\frac{\partial}{\partial t}\tilde{J}(Y_t^*, p_t^*) &\leq \xi_t^* b(u_t^*, Y_t^*, p_t^*) \\ &= \beta\tilde{J}(Y_t^*, p_t^*) - g(p_t^*, Y_t^*, u_t^*) \\ &\quad - \sum_{l=1}^{m}\left(\tilde{J}(f_l, p_{t-}^* + \Phi_{Y_{t-}^* l}(u_t^*, p_{t-}^*)) - \tilde{J}(Y_{t-}^*, p_{t-}^*)\right) q^Y_{Y_{t-}^* l}(u_t^*, p_{t-}^*)\end{aligned}$$

from which we conclude with

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

and by integration from 0 to t that

$$\begin{aligned}e^{-\beta t}\tilde{J}(Y_t^*, p_t^*) &\leq \tilde{J}(y,p) - \int_0^t e^{-\beta s} g(p_s^*, Y_s^*, u_s^*) ds \\ &\quad + \int_0^t e^{-\beta s}\sum_{l=1}^{m}\left(\tilde{J}(f_l, p_{s-}^* + \Phi_{Y_{s-}^* l}(u_s^*, p_{s-}^*)) - \tilde{J}(Y_{s-}^*, p_{s-}^*)\right)\cdot \\ &\qquad \cdot\left(dN_s^Y(Y_{s-}^*, f_l) - q^Y_{Y_{s-}^* l}(u_s^*, p_{s-}^*) ds\right).\end{aligned}$$

As in part a) taking expectation and letting $t \to \infty$ we obtain

$$\tilde{J}(y,p) \geq \mathbb{E}_{u^*}\left[\int_0^\infty e^{-\beta t} g(p_t^*, Y_t^*, u_t^*) dt\right].$$

Hence we conclude with a)

$$J(y,p) \geq \tilde{J}(y,p) \geq \mathbb{E}_{u^*}\left[\int_0^\infty e^{-\beta t} g(p_t^*, Y_t^*, u_t^*) dt\right] \geq J(y,p)$$

which proves the results of b).

□

Since the value function $J(y,p)$ is concave in p (see theorem 3.15) it is regular in the meaning of the Clarke derivative and it is locally Lipschitz continuous, see corollary 3.16. Thus $J(y,p)$ is the unique solution of the generalized HJB-equation. This property will hold true usually in partial observable models as we consider here. For general piecewise-deterministic optimization problems this property will not be true usually. We will apply this verification theorem in section 5.2, where we prove sufficient conditions and the existence of optimal controls.

Remark 4.5 *If the value function $J(y,p)$ is differentiable in p, then theorem 4.3 and 4.4 reduce to the classical ones. In this case the Clarke generalized gradient is the usual gradient, consequently $\partial_p J(y,p) = \{J_p(y,p)\}$.*

4.2 Solution via a Transformed MDP

Whereas the verification technique makes no use of the piecewise-deterministic behaviour of the state process (Y_t, p_t) in our reduced problem (P_{red}) the following solution procedure does. It goes back to the reduction technique for Semi-Markov-Decision-Processes (SMDP), where a time-continuous Markov chain is reduced to its embedded time-discrete Markov chain. Thus the decision time points can be reduced to the jump points of the embedded Markov chain. Various authors (Davis (1993), Dempster (1989), Forwick (1997)) extend this idea to the case of piecewise-deterministic decision processes. But they do not include discounting and did not make use of the uniformization technique (remember lemma 3.10). We will define a time-discrete Markovian-Decision-Process (MDP), which is strongly connected to the reduced control problem. We are then able to use all the tools of the MDP-theory to solve our optimization problem (P_{red}) and hence (P).

The state process (Y_t, p_t) of our reduced problem (P_{red}) is essentially described by the state after a jump-time point and the elapsed time since the last jump. Between two jumps the behaviour is deterministic as seen in section 3.1. The MDP defined below will be extended later on to a uniformized model as mentioned earlier. If the intensities or the cost rate g depend on time, one has to extend the state process by the time component in a similar way as pointed out later in remark 4.11. Remember first that the distribution function of the holding times $\tau_{n+1} - \tau_n$ is given by:

$$\begin{aligned}\mathbb{P}_u(\tau_{n+1} - \tau_n \le t \mid Y_0, p_0, \tau_1, Y_{\tau_1}, p_{\tau_1}, \ldots, Y_{\tau_n}, p_{\tau_n}) &= \mathbb{P}_u(\tau_{n+1} - \tau_n \le t \mid \tau_n, Y_{\tau_n}, p_{\tau_n}) \\ &= 1 - \exp\left\{-\int_0^t q^Y_{Y_{\tau_n}}(u_s, \phi^u_s(p_{\tau_n}))ds\right\} \\ &=: F(Y_{\tau_n}, p_{\tau_n}, u; t)\end{aligned}$$

independent of the post jump state. Then define a time-discrete MDP by

- state space: $S = S_Y \times \triangle^{nd} \ni (y, p)$
- action space: $A = \{\gamma \mid \gamma : \mathbb{R}_+ \to U \text{ measurable}\}$
- set of admissible controls: $D(y, p) = A \quad \forall (y, p) \in S$
- transition probability: for $y \neq \omega, B \subset \triangle^{nd}$

$$\begin{aligned}& q\big((y,p),\gamma,(\omega,B)\big) \\ =\ & \int_0^\infty \frac{q^Y_{y\omega}(\gamma_t, \phi^\gamma_t(p))}{q^Y_y(\gamma_t, \phi^\gamma_t(p))} \mathbb{1}\big(\phi^\gamma_t(p) + \Phi_{y\omega}(\gamma_t, \phi^\gamma_t(p)) \in B\big) F(y, p, \gamma; dt) \\ =\ & \int_0^\infty q^Y_{y\omega}(\gamma_t, \phi^\gamma_t(p)) \mathbb{1}\big(\phi^\gamma_t(p) + \Phi_{y\omega}(\gamma_t, \phi^\gamma_t(p)) \in B\big) \exp\left\{-\int_0^t q^Y_y(\gamma_s, \phi^\gamma_s(p))ds\right\} dt\end{aligned}$$

and

$$q\big((y,p),\gamma,(y,B)\big) = 0$$

- cost function:

$$\begin{aligned} & r((y,p),\gamma) \\ = & \int_0^\infty \left(\int_0^t e^{-\beta s} g(\phi_s^\gamma(p), y, \gamma_s) ds \right) F(y,p,\gamma;dt) \\ = & \int_0^\infty \left(\int_0^t e^{-\beta s} g(\phi_s^\gamma(p), y, \gamma_s) ds \right) q_y^Y(\gamma_t, \phi_t^\gamma(p)) \exp\left\{ - \int_0^t q_y^Y(\gamma_s, \phi_s^\gamma(p)) ds \right\} dt \end{aligned}$$

- discount factor:

$$\begin{aligned} \delta((y,p),\gamma) &= \int_0^\infty e^{-\beta t} F(y,p,\gamma;dt) \\ &= \int_0^\infty e^{-\beta t} q_y^Y(\gamma_t, \phi_t^\gamma(p)) \exp\left\{ - \int_0^t q_y^Y(\gamma_s, \phi_s^\gamma(p)) ds \right\} dt. \end{aligned}$$

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

$$V_{\infty,\pi}(y,p) := \mathbb{E}_\pi \left[\sum_{n=0}^\infty \left(\prod_{k=0}^{n-1} \delta\big(Y_{\tau_k}, p_{\tau_k}, f_k(Y_{\tau_k}, p_{\tau_k})\big) \right) r\big(Y_{\tau_n}, p_{\tau_n}, f_n(Y_{\tau_n}, p_{\tau_n})\big) \,\Big|\, Y_0 = y, p_0 = p \right],$$

where the expectation is taken with respect to $\mathbb{P}_\pi$ which is defined by the transition probabilities q (see e.g. Hernandez-Lerma and Lasserre (1996)). Since r is positive the expectation is well-defined and we define the set of positive functions on S by

$$\mathbb{B} := \{v : S \to \mathbb{R} \mid v \geq 0\}.$$

The value function is defined by

$$V_\infty(y,p) := \inf_{\pi \in F^\infty} V_{\infty,\pi}(y,p),$$

the minimal expected discounted costs over an infinite horizon. It holds $V_{\infty,\pi}$ and $V_\infty \in \mathbb{B}$.

Define for $v \in \mathbb{B}$ the operators $L : \mathbb{B} \to \mathbb{B}$, $\mathcal{T}_f : \mathbb{B} \to \mathbb{B}$ and $\mathcal{T} : \mathbb{B} \to \mathbb{B}$ by

$$\begin{aligned} (Lv)(y,p,\gamma) &:= r((y,p),\gamma) + \delta((y,p),\gamma) \int_{\triangle^{nd}} \sum_{\omega \in S_Y} v(\omega,\rho) q((y,p),\gamma,(d\rho,\omega)) \text{ for } \gamma \in A \\ (\mathcal{T}_f v)(y,p) &:= (Lv)(y,p,f(y,p)) \text{ for } f \in F \\ (\mathcal{T} v)(y,p) &:= \inf_{\gamma \in A} \{(Lv)(y,p,\gamma)\}. \end{aligned}$$

All operators are obviously isotone. We call $f \in F$ a minimizer of v if $\mathcal{T}_f v = \mathcal{T} v$.

Remark 4.6 *If the jump distribution of the holding times $\tau_{n+1} - \tau_n$ depends additionally on the post-jump state, that means is of the form $F((y,p),u,(\omega,B);t)$, the MDP can be generalized to*

$$q\big((y,p),\gamma,(\omega,B)\big) = \int_0^\infty \frac{q^Y_{y\omega}(\gamma_t,\phi^\gamma_t(p))}{q^Y_y(\gamma_t,\phi^\gamma_t(p))} \mathbb{1}\big(\phi^\gamma_t(p)+\Phi_{y\omega}(\gamma_t,\phi^\gamma_t(p)) \in B\big) F((y,p),\gamma,(\omega,B);dt)$$

$$r((y,p),\gamma) = \int_0^\infty \left(\int_0^t e^{-\beta s} g(\phi^\gamma_s(p),y,\gamma_s)ds\right) \int_{\triangle^{nd}} \sum_{\omega\neq y} \frac{q^Y_{y\omega}(\gamma_t,\phi^\gamma_t(p))}{q^Y_y(\gamma_t,\phi^\gamma_t(p))} \cdot$$
$$\cdot \mathbb{1}\big(\phi^\gamma_t(p)+\Phi_{y\omega}(\gamma_t,\phi^\gamma_t(p)) \in \{d\rho\}\big) F((y,p),\gamma,(\omega,d\rho);dt)$$

$$\delta((y,p),\gamma,(\omega,B)) = \int_0^\infty e^{-\beta t} F((y,p),\gamma,(\omega,B);dt)$$

The next theorem is the justification for the transformation of (P_{red}) into the just defined MDP. It shows that the above MDP is indeed strongly connected to the control problem (P_{red}), since the value function $V_\infty(y,p)$ and $J(y,p)$ are equal and if we find an optimal strategy of the MDP we have an optimal control of (P_{red}).

Theorem 4.7 *It holds:*

a) $J(y,p) = V_\infty(y,p)$

b) *If $\pi^* = (f_n) \in F^\infty$ is optimal for the MDP, then $u^* = (u^*_t) \in \mathcal{U}$ with*

$$u^*_t := f_n(Y^*_{\tau_n}, p^*_{\tau_n})(t-\tau_n) \quad \text{for } t \in [\tau_n, \tau_{n+1})$$

is optimal for (P_{red}).

Proof:

a) Denote by τ_k the last jump before t and introduce the state space of the observed past at time t as $H_k := \big(\mathbb{R}_+ \times S_Y \times \triangle^{nd}\big)^k$. Since $\tau_0 := 0$ we define $h_k \in H_k$ by

$$h_k := \{Y_0, p_0, \tau_1, Y_{\tau_1}, p_{\tau_1}, \ldots, \tau_n, Y_{\tau_k}, p_{\tau_k}\}.$$

$u = (u_t) \in \mathcal{U}$ is allowed to depend on the whole (observable) past $\mathcal{F}^Y_t$. Hence the control at time t can be written as (see e.g. Elliott et al. (1997))

$$u_t = u(Y_0, p_0, \tau_1, Y_{\tau_1}, p_{\tau_1}, \ldots, \tau_k, Y_{\tau_k}, p_{\tau_k})(t-\tau_k) \text{ for } t \in [\tau_k, \tau_{k+1}).$$

Thus we have to introduce a generalized policy $\tilde\pi = (\tilde f_n)$, which is allowed to depend on the whole past. That means $\tilde f_n \in \tilde F_n$ with $\tilde F_n := \{f : H_n \to A \text{ measurable}\}$. Then it is easy to see that for every $u = (u_t) \in \mathcal{U}$ a corresponding $\tilde\pi = (\tilde f_n) \in \times_{k=0}^\infty \tilde F_k$ can be found with

$$u_t = \tilde f_n(h_n)(t-\tau_n) \text{ for } t \in [\tau_n, \tau_{n+1}). \tag{4.3}$$

Due to this we will not differ between $\mathbb{E}_u$ and $\mathbb{E}_{\tilde{\pi}}$ in notation and we continue:

$$
\begin{aligned}
& J(y,p;u) = \mathbb{E}_u\left[\int_0^\infty e^{-\beta t} g(p_t, Y_t, u_t)dt\right] \\
= \;& \mathbb{E}_u\left[\sum_{n=0}^\infty \int_{\tau_n}^{\tau_{n+1}} e^{-\beta t} g(p_t, Y_t, u_t)dt\right] = \sum_{n=0}^\infty \mathbb{E}_u\left[\int_{\tau_n}^{\tau_{n+1}} e^{-\beta t} g(p_t, Y_t, u_t)dt\right] \\
= \;& \sum_{n=0}^\infty \mathbb{E}_u\left[e^{-\beta\tau_n}\int_0^{\tau_{n+1}-\tau_n} e^{-\beta t} g(p_{t+\tau_n}, Y_{t+\tau_n}, u_{t+\tau_n})dt\right] \\
= \;& \sum_{n=0}^\infty \mathbb{E}_u\left[e^{-\beta\tau_n}\mathbb{E}_u\left\{\int_0^{\tau_{n+1}-\tau_n} e^{-\beta t} g(p_{t+\tau_n}, Y_{t+\tau_n}, u_{t+\tau_n})dt \mid \mathcal{F}^Y_{\tau_n}\right\}\right] \\
= \;& \sum_{n=0}^\infty \mathbb{E}_u\Bigg[e^{-\beta\tau_n}\int_0^\infty \left(\int_0^s e^{-\beta t} g(\phi_t^{\tilde{f}_n}(p_{\tau_n}), Y_{\tau_n}, \tilde{f}_n(h_n)(t))dt\right)\cdot \\
& \qquad \cdot d\mathbb{P}_u(\tau_{n+1}-\tau_n \geq s \mid \mathcal{F}^Y_{\tau_n})\Bigg] \\
= \;& \sum_{n=0}^\infty \mathbb{E}_u\Bigg[e^{-\beta\tau_n}\int_0^\infty \left(\int_0^s e^{-\beta t} g(\phi_t^{\tilde{f}_n}(p_{\tau_n}), Y_{\tau_n}, \tilde{f}_n(h_n)(t))dt\right)\cdot \\
& \qquad \cdot d\mathbb{P}_u(\tau_{n+1}-\tau_n \geq s \mid \tau_n, Y_{\tau_n}, p_{\tau_n})\Bigg] \\
= \;& \sum_{n=0}^\infty \mathbb{E}_u\Bigg[e^{-\beta\tau_n}\int_0^\infty \left(\int_0^s e^{-\beta t} g(\phi_t^{\tilde{f}_n}(p_{\tau_n}), Y_{\tau_n}, \tilde{f}_n(h_n)(t))dt\right)\cdot \\
& \qquad \cdot q^Y_{Y_{\tau_n}}(\tilde{f}_n(h_n)(s), \phi_s^{\tilde{f}_n}(p_{\tau_n}))\cdot \\
& \qquad \cdot \exp\left\{-\int_0^t q^Y_{Y_{\tau_n}}(\tilde{f}_n(h_n)(t), \phi_t^{\tilde{f}_n}(p_{\tau_n}))dt\right\}ds\Bigg] \\
= \;& \sum_{n=0}^\infty \mathbb{E}_u\left[e^{-\beta\tau_n} r(Y_{\tau_n}, p_{\tau_n}, \tilde{f}_n(h_n))\right] \\
= \;& \sum_{n=0}^\infty \mathbb{E}_u\left[\prod_{k=1}^n e^{-\beta(\tau_k-\tau_{k-1})} r(Y_{\tau_n}, p_{\tau_n}, \tilde{f}_n(h_n))\right]
\end{aligned}
$$

where

$$
\begin{aligned}
r(Y_{\tau_n}, p_{\tau_n}, \tilde{f}_n(h_n)) \;=\; & \int_0^\infty \left(\int_0^s e^{-\beta t} g(\phi_t^{\tilde{f}_n}(p_{\tau_n}), Y_{\tau_n}, \tilde{f}_n(h_n)(t))dt\right)\cdot \\
& \cdot q^Y_{Y_{\tau_n}}(\tilde{f}_n(Y_{\tau_n}, p_{\tau_n})(s), \phi_s^{\tilde{f}_n}(p_{\tau_n}))\cdot \\
& \cdot \exp\left\{-\int_0^t q^Y_{Y_{\tau_n}}(\tilde{f}_n(h_n)(t), \phi_t^{\tilde{f}_n}(p_{\tau_n}))dt\right\}ds.
\end{aligned}
$$

It holds further with $\mathcal{C}_k := \{Y_{\tau_l} = y_l, p_{\tau_l} = p_l, \tau_l - \tau_{l-1} > t_l, l = 1, \ldots, k-1\}$:

$$\begin{aligned} & \mathbb{P}_u(\tau_k - \tau_{k-1} \le t, Y_{\tau_k} = y_k, p_{\tau_k} \in B_k, k = 1, \ldots, n-1) \\ = & \prod_{k=1}^{n} \mathbb{P}_u(Y_{\tau_k} = y_k, p_{\tau_k} \in B_k \mid \mathcal{C}_k) \mathbb{P}_u(\tau_k - \tau_{k-1} \le t \mid Y_{\tau_k} = y_k, p_{\tau_k} = p_k, \mathcal{C}_k) \end{aligned}$$

and we conclude

$$\begin{aligned} & \mathbb{E}_u \left[\prod_{k=1}^{n} e^{-\beta(\tau_k - \tau_{k-1})} r(Y_{\tau_n}, p_{\tau_n}, \tilde{f}_n(h_n)) \right] \\ = & \sum_{y_0, \ldots, y_n \in S_Y} \int_{\triangle^{nd}} \cdots \int_{\triangle^{nd}} r(y_n, p_n, \tilde{f}_n(h_n)) \cdot \\ & \qquad \cdot \prod_{k=0}^{n-1} \int_0^\infty e^{-\beta s_k} d\mathbb{P}_u(\tau_k - \tau_{k-1} \le s_k, Y_{\tau_k} = y_k, dp_{\tau_k}) \\ = & \mathbb{E}_u \left[r(Y_{\tau_n}, p_{\tau_n}, \tilde{f}_n(h_n)) \prod_{k=0}^{n-1} \int_0^\infty e^{-\beta s_k} F(y_k, p_k, \tilde{f}_k(h_k); ds_k) \right] \\ = & \mathbb{E}_u \left[r(Y_{\tau_n}, p_{\tau_n}, \tilde{f}_n(h_n)) \prod_{k=0}^{n-1} \delta(y_k, p_k, \tilde{f}_k(h_k)) \right] \end{aligned}$$

where

$$\delta(y, p, \gamma) = \int_0^\infty e^{-\beta t} F(y, p, \gamma; dt).$$

Taking now the infimum over all $u \in \mathcal{U}$ or by (4.3) equivalent over all $\tilde{\pi} \in \times_{k=0}^{\infty} \tilde{F}_k$ we conclude

$$J(y, p) = \inf_{\tilde{\pi}} V_{\infty, \tilde{\pi}}(y, p).$$

From Bertsekas and Shreve (1978) it is well-know that it is sufficient to consider on the right hand side the infimum over all Markov strategies $\pi \in F^\infty$, thus

$$J(y, p) = \inf_{\tilde{\pi}} V_{\infty, \tilde{\pi}}(y, p) = \inf_{\pi \in F^\infty} V_{\infty, \pi}(y, p) = V_\infty(y, p).$$

b) The statement follows from the last line of the proof of a) and (4.3) for Markov strategies $\pi = (f_n)$ which reads as

$$u_t = f_n(Y_{\tau_n}, p_{\tau_n})(t - \tau_n) \text{ for } t \in [\tau_n, \tau_{n+1}).$$

□

Corollary 4.8 *It holds:*

a) $V_\infty(y,p) = \mathcal{T}V_\infty(y,p)$ *and due to theorem 4.7* $J(y,p) = \mathcal{T}J(y,p)$.

b) If $\mathcal{T}_f V_\infty(y,p) = \mathcal{T}V_\infty(y,p)$ *then* f^∞ *is optimal.*

Proof: Both statement are well-known from the MDP-theory, see e.g. Bertsekas and Shreve (1978). □

After defining this MDP we discuss the question how to compute optimal controls. One way is to use Howard's policy improvement, which works under certain conditions. Another way is to find minimizers f_n of V_{n-1} and then taking the policy consisting of accumulation points of this sequences (f_n). This is an optimal policy for the infinite horizon MDP under compactness and continuity conditions (see e.g. Hernandez-Lerma and Lasserre (1996) and corollary 4.16). The third approach was pointed out in corollary 4.8: find a minimizer of V_∞. To close the circle to section 4.1 we discuss the latter one. Consider therefore the Bellman equation:

$$\begin{aligned}
& V_\infty(y,p) = \mathcal{T}V_\infty(y,p) \\
= \; & \inf_{\gamma\in A}\Bigg\{\mathbb{E}_u\Bigg[\int_0^\tau e^{-\beta t} g(\phi_t^\gamma(p), y, \gamma_t)dt \qquad (4.4)\\
& \qquad + e^{-\beta\tau}\Big(\sum_{\omega\neq y} V_\infty(\omega, \phi_{\tau-}^\gamma(p) + \Phi_{y\omega}(\gamma_\tau, \phi_{\tau-}^\gamma(p)))\frac{q_{y\omega}^Y(\gamma_\tau, \phi_{\tau-}^\gamma(p))}{q_y^Y(\gamma_\tau, \phi_{\tau-}^\gamma(p))}\Big)\Bigg]\Bigg\} \\
= \; & \inf_{\gamma\in A}\Bigg\{\int_0^\infty e^{-\int_0^t (q_y^Y(\gamma_s, \phi_s^\gamma(p))+\beta)ds}\Big\{g(\phi_t^\gamma(p), y, \gamma_t) \\
& \qquad + \sum_{\omega\neq y} V_\infty(\omega, \phi_t^\gamma(p) + \Phi_{y\omega}(\gamma_t, \phi_t^\gamma(p)))q_{y\omega}^Y(\gamma_t, \phi_t^\gamma(p))\Big\}dt\Bigg\}
\end{aligned}$$

where we used in the second line the proof of theorem 4.7 and then we computed the expectation. The right hand side is a (deterministic) control problem which can be solved with the help of Pontryagin's maximum principle or with the help of the (generalized) verification technique for deterministic models (remember the stochastic case was described in section 4.1). Denote by $\bar{V}(y,p)$ the value function of this deterministic control problem for fixed $V_\infty(y,p)$. Then it is well-known from the verification theorem 4.4, that if $W(y,p)$ is locally Lipschitz continuous and regular in p and if it is for fixed $V_\infty(y,p)$ a solution of

the corresponding HJB-equation

$$\begin{aligned}
&\inf_{\substack{\xi\in\partial_p W(y,p)\\ u\in U}} \Big\{ g(p,y,u) + \xi b(u,y,p) \\
&\qquad\qquad + \sum_{\omega\neq y} V(\omega, p+\Phi_{y\omega}(u,p)) q^Y_{y\omega}(u,p) - (q^Y_y(u,p)+\beta) W(y,p) \Big\} \\
= &\inf_{\substack{\xi\in\partial_p W(y,p)\\ u\in U}} \Big\{ g(p,y,u) + \xi b(u,y,p) - \beta W(y,p) \\
&\qquad\qquad + \sum_{\omega\neq y} \Big(V(\omega, p+\Phi_{y\omega}(u,p)) - W(y,p) \Big) q^Y_{y\omega}(u,p) \Big\} = 0
\end{aligned}$$

that $W(y,p) = \bar{V}(y,p) = V_\infty(y,p)$. Hence the HJB-equation for the deterministic problem on the right hand side of the Bellman equation of the MDP simplifies to

$$\begin{aligned}
&\inf_{\substack{\xi\in\partial_p W(y,p)\\ u\in U}} \Big\{ g(p,y,u) + \xi b(u,y,p) \\
&\qquad\qquad + \sum_{\omega\in S_Y} \big(W(\omega, p+\Phi_{y\omega}(u,p)) - W(y,p) \big) q^Y_{y\omega}(u,p) \Big\} = \beta W(y,p),
\end{aligned}$$

which is exactly the HJB-equation from theorem 4.4 for the reduced problem (P_{red}).

Theorem 4.9

a) The HJB-equations of (P_{red}) given in (4.2) and the HJB-equation of the deterministic control problem in the operator $\mathcal{T}$ arising in the MDP-approach are the same.

b) Since $V_\infty(y,p) = J(y,p)$ by theorem 4.7 we see that (4.4) in $J(y,p) = \mathcal{T}J(y,p)$ is the analogon to theorem 4.1 for $t = 0$.

Therefore we see both approaches, the verification technique of section 4.1 and the MDP-approach, lead to the same HJB-equation in the end. That means computation of optimal controls with the verification technique or as minimizers encounter the same difficulties like differentiability or the right guess for the value function.

The advantage of the MDP-approach is the opportunity to make use of the whole MDP-theory to state existence results for an optimal control and to find a representation of the value function $V_\infty(y,p)$ with the value iteration.

The next both remarks complete the definition of the equivalent time-discrete MDP. The first remark introduces the uniformized MDP, where the distribution function F of the sojourn times is independent of the current state and control. The second extension in remark 4.11 contains the transformation for a finite horizon control problem.

Remark 4.10 *Using the uniformization technique introduced at the end of section 3.1 and in lemma 3.10, we define the following MDP which is equivalent to the last one. For this purpose set*

$$\alpha := \sup_{k,\mu,i} \sup_{u \in U} \left\{ q_k^Y(u, g_\mu, e_i) \right\}$$

as in section 3.1, but notice that we also take the supremum with respect to the control parameter. Thus define the distribution function of the holding times $\tau_{n+1} - \tau_n$ as

$$\mathbb{P}_u(\tau_{n+1} - \tau_n \le t \mid Y_0, p_0, \tau_1, Y_{\tau_1}, p_{\tau_1}, \ldots, Y_{\tau_n}, p_{\tau_n}) = \left(1 - e^{-\alpha t}\right) \mathbb{1}\left(t \ge 0\right) =: F(t)$$

independent of the state and control processes. Then the MDP has to be defined by:

- *state space:* $S = S_Y \times \triangle^{nd} \ni (y, p)$
- *action space:* $A = \{\gamma \mid \gamma : \mathbb{R}_+ \to U \text{ measurable}\}$
- *set of admissible controls:* $D(y, p) = A \quad \forall (y, p) \in S$
- *transition probability:*

$$q\big((y,p), \gamma, (\omega, B)\big) = \int_0^\infty e^{-\alpha t} q_{y\omega}^Y(\gamma_t, \phi_t^\gamma(p)) \mathbb{1}\left(\phi_t^\gamma(p) + \Phi_{y\omega}(\gamma_t, \phi_t^\gamma(p)) \in B\right) dt, \quad y \ne \omega$$
$$q\big((y,p), \gamma, (y, B)\big) = 1 - \sum_{\omega \ne y} q\big((y,p), \gamma, (\omega, B)\big)$$

  where $\omega \in S_Y, B \subset \triangle^{nd}$

- *cost function:*

$$r((y,p), \gamma) = \int_0^\infty \alpha e^{-\alpha t} \left(\int_0^t e^{-\beta s} g(\phi_s^\gamma(p), y, \gamma_s) ds \right) dt$$

- *discount factor:*

$$\delta = \int_0^\infty e^{-\beta t} \alpha e^{-\alpha t} dt = \frac{\alpha}{\alpha + \beta} < 1$$

  independent of state and control process.

If the cost function $g(p, y, u)$ does not depend on p and u, that means it is of the form $g(y)$, then the cost rate simplifies to:

$$\begin{aligned} r(y, p, \gamma) &= g(y) \int_0^\infty \left(\int_0^t e^{-\beta t} ds \right) \alpha e^{-\alpha t} dt = -\frac{g(y)}{\beta} \int_0^\infty (e^{-\beta t} - 1) \alpha e^{-\alpha t} dt \\ &= -\frac{g(y)\alpha}{\beta} \left(-\frac{1}{\alpha + \beta} + \frac{1}{\alpha} \right) = \frac{g(y)}{\alpha + \beta} =: r(y), \end{aligned}$$

which coincides with the formula in the case of a pure jump SMDP. Note that $g(p,y,u)$ depends via the controlled state process on the control. Observe also that for its independence of the estimator process, $g(z,x,y,u)$ has to be independent of z and x.

The Bellman equation for the general uniformized model is given by:

$$\begin{aligned} v(y,p) &= \mathcal{T}v(y,p) \\ &= \inf_{\gamma\in A}\Bigg\{\int_0^\infty e^{-(\alpha+\beta)t}\Big\{g(\phi_t^\gamma(p),y,\gamma_t) \\ &\qquad + \sum_{\omega\neq y} v(\omega,\phi_t^\gamma(p)+\Phi_{y\omega}(\gamma_t,\phi_t^\gamma(p)))q_{y\omega}^Y(\gamma_t,\phi_t^\gamma(p)) \\ &\qquad + v(y,\phi_t^\gamma(p))(\alpha-\sum_{\omega\neq y} q_{y\omega}^Y(\gamma_t,\phi_t^\gamma(p)))\Big\}dt\Bigg\}, \end{aligned}$$

which can be computed as in the non-uniformized case.

If $g(p,y,u)$ is bounded, then the value function $V_\infty(y,p)$ is bounded too and we conclude that $\mathcal{T}$ is a contracting operator on the set of bounded function with contraction parameter $\delta<1$, since

$$||\mathcal{T}v-\mathcal{T}w||\leq\delta||v-w||,$$

where $||v(y,p)||:=\sup_{y,p}|v(y,p)|$. Consequently one can use Banach's fixed point theorem to prove without continuity-assumptions, that $\lim_{n\to\infty} V_n(y,p)=V_\infty(y,p)$ and for the computation of optimal strategies procedures as Howard's policy improvement algorithm may be applied.

Remark 4.11 *Considering a control problem with finite horizon T and terminal cost $h(y,p)$ we have to extend the state process by the time component t. Therefore the state process (t,Y_t,p_t) evolves between two jump-time point, that means for $t\in[\tau_n,\tau_{n+1})$, as $(t,Y_{\tau_n},\phi^u_{t-\tau_n}(p_{\tau_n}))$, where $\phi_t^u(p)$ is the unique solution of*

$$\begin{cases} \dot p &= b(u,y,p) \\ p_0 &= p \end{cases}$$

Notice that $\phi_t^u(p)$ is defined as in theorem 3.9. Again the jump distribution function is uniformized, hence

$$\mathbb{P}_u(\tau_{n+1}-\tau_n\leq t\mid Y_0,p_0,\tau_1,Y_{\tau_1},p_{\tau_1},\ldots,Y_{\tau_n},p_{\tau_n})=\big(1-e^{-\alpha t}\big)\,\mathbb{1}\big(t\geq 0\big)\mathbb{1}\big(\tau_n+t\leq T\big)$$

and then the corresponding uniformized (infinite horizon) MDP is accordingly given by:

- *state space: $S=[0,T]\times S_Y\times\triangle^{nd}\ni(t,y,p)$*

- *action space:* $A = D(t,y,p) = \{\gamma \mid \gamma : \mathbb{R}_+ \to U\}$
- *transition probability:*

$$\begin{aligned} & q\big((t,y,p),\gamma,([t_1,t_2),\omega,B)\big) \\ = & \int_0^{T-t} e^{-\alpha s} q^Y_{y\omega}(\gamma_s,\phi^\gamma_s(p)) \mathbb{1}\big((s,\phi^\gamma_s(p)+\Phi_{y\omega}(\gamma_s,\phi^\gamma_s(p))) \in [t_1,t_2)\times B\big) ds \end{aligned}$$

  with $[t_1,t_2) \subset [0,T]$, $\omega \in S_Y$ *and* $B \subset \triangle^{nd}$.

- *cost function:*

$$\begin{aligned} & r(t,y,p,\gamma) \\ = & \int_0^{T-t} \alpha e^{-\alpha s} \left(\int_0^s e^{-\beta r} g(\phi^\gamma_r(p),y,\gamma_r) dr \right) ds + e^{-(\alpha+\beta)(T-t)} h(y,\phi^\gamma_{T-t}(p)) \end{aligned}$$

- *discount factor:* $\delta(t) = \int_0^{T-t} \alpha e^{-(\alpha+\beta)s} ds = \frac{\alpha}{\alpha+\beta}\left(1 - e^{-(\alpha+\beta)(T-t)}\right)$

The Bellman equation is given by

$$\begin{aligned} & v(t,y,p) = \mathcal{T}v(t,y,p) \\ = & \inf_{\gamma\in A} \left\{ r(t,y,p,\gamma) + \delta(t) \int_{\triangle^{nd}} \sum_{\omega\in S_Y} v(t,\omega,\rho) q\big((t,y,p),\gamma,(\delta_t,\omega,d\rho)\big) \right\} \\ = & \inf_{\gamma\in A} \Bigg\{ \int_0^{T-t} e^{-(\alpha+\beta)s} \Big\{ g(\phi^\gamma_s(p),y,\gamma_s) \\ & \qquad + \sum_{\omega\neq y} v(s,\omega,\phi^\gamma_s(p)+\Phi_{y\omega}(\gamma_s,\phi^\gamma_s(p))) q^Y_{y\omega}(\gamma_s,\phi^\gamma_s(p)) \\ & \qquad + v(s,y,\phi^\gamma_s(p))(\alpha - \sum_{\omega\neq y} q^Y_{y\omega}(\gamma_s,\phi^\gamma_s(p))) \Big\} ds \\ & \qquad + e^{-(\alpha+\beta)(T-t)} h(y,\phi^\gamma_{T-t}(p)) \Bigg\} \end{aligned}$$

If g *and* h *are bounded then the model is obviously substochastic and for the operator* $\mathcal{T}$ *holds*

$$||\mathcal{T}v - \mathcal{T}w|| \leq (1 - e^{-(\alpha+\beta)T})||v-w||,$$

in particular $\mathcal{T}$ *is a contracting operator.*
If the cost function g *in the origin (infinite horizon) model depends on time* t*, then we have to extend the state space in the transformed MDP by the time component.*

In a pure jump SMDP it is well-known that an optimal control exists if A is finite, since in these models it is sufficient to consider actions $\gamma \in A$ which are constant between two jumps. In general the existence of an optimal policy is guaranteed under continuity and compactness assumptions, which are satisfied for the wider class of relaxed controls in a suitable topology. Then additionally the convergence of the n-stage value functions $V_n(y,p)$ to $V_\infty(y,p)$ is true. Assume for this U is convex.

Definition 4.12 *A measurable function $r : \mathbb{R}_+ \to \triangle(U)$ is called a relaxed control, where $\triangle(U)$ is the probability simplex of U. The set of all relaxed controls will be denoted by $\mathcal{R}$. We say in this context that $\gamma \in A$ is deterministic. If a relaxed control takes only value in $\{0,1\}$, that means r is always a corner of $\triangle(U)$, then r is called pure.*

Define for a relaxed control

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

Instead of choosing at each time point t a fixed control $u \in U$ we randomize now over the set of possible values in U. The case of deterministic controls is always included by choosing the Dirac-measure δ on U. Because of the relaxation we have to consider our state process (Y_t, p_t) now with respect to relaxed controls. As in section 2.1 we define for a relaxed control $r = (r_t)$ the corresponding state process (Y_t^r, p_t^r) in the first component by its (relaxed) intensities

$$q_{kl}^Y(r, g_\mu, e_i) = \int_U q_{kl}^Y(u, g_\mu, e_i) r(du).$$

In the second component define the state process $p_t^r = \phi_{t-\tau}^r(p_\tau)$ under $Y_t = y$ between two jumps as in lemma 3.9 as the solution of

$$\dot{p}^r = b\left(\bar{r}, y, p^r\right).$$

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

$$\phi_t^r(p) = \phi_t^{\bar{r}}(p). \tag{4.5}$$

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

$$\Phi(r,p) = \int_U \Phi(u,p) r(du).$$

The process (Y_t^r, p_t^r) as a solution of the above introduced characteristics is well-defined (see Davis (1993)). Finally define the cost rate for a relaxed control r as

$$g(p,y,r) = \int_U g(p,y,u) r(du).$$

The following construction is adopted from Davis (1993). A similar consideration can be found in Presman and Sonin (1990). Endow $\mathcal{R}$ with the Young topology, which is at the end a suitable topology, which guarantees compactness and continuity. For this purpose denote first $L_1(\mathbb{R}_+; C(U))$, the space of functions $h(t,u)$, which are integrable over $(\mathbb{R}_+; C(U))$. Hence every function $h(t,u) \in L_1(\mathbb{R}_+; C(U))$ is measurable in t, continuous in u and satisfies $||h|| := \int_0^\infty \max_{u\in U} |h(t,u)|dt < \infty$. Then one can conclude that $L_1(\mathbb{R}_+; C(U))$ is a Banach space.

The dual space is accordingly given by $L_\infty(\mathbb{R}_+; C^*(U))$, where $C^*(U)$ is the dual space of $C(U)$ (consisting of the set of signed measures on U under the total variation norm), consisting of measurable functions $v : \mathbb{R}_+ \to C^*(U)$ such that $||v||_* = \text{esssup}_{t\in\mathbb{R}_+} ||v_t||_{C^*} < \infty$.

Introduce the unit ball $B_1 := \{v \in L_\infty(\mathbb{R}_+; C^*(U)) \mid ||v||_* \leq 1\}$ which is compact in the weak* topology on $L_\infty(\mathbb{R}_+; C^*(U))$. Then the Young topology on $\mathcal{R}$ is the relative weak* topology of $\mathcal{R}$ considered as a subset of B_1.

In the next lemma we prove that all continuity and compactness conditions for the existence of an optimal (relaxed) policy are fulfilled. We refer for them to Hernandez-Lerma and Lasserre (1996) and Bertsekas and Shreve (1978).

Lemma 4.13

a) $A^{\mathcal{R}} := \{\nu : \mathbb{R}_+ \to \triangle(U)$ *measurable*$\}$ *is compact.*

b) $(p,\nu) \mapsto \phi_t^\nu(p)$ *is continuous for* $\nu \in A^{\mathcal{R}}$ *and* $p \in \triangle^{nd}$.

Proof:

a) Davis (1993).

b) Let $\nu^n = (\nu_t^n) \to (\nu_t) = \nu$ and $p^n \to p$ and denote by $\phi_t^{\nu^n}$ and ϕ_t^ν the solution of

$$\dot{p} = b(\bar{\nu^n}, y, p) \quad \text{with } p_0 = p^n \qquad \text{and} \qquad \dot{p} = b(\bar{\nu}, y, p) \quad \text{with } p_0 = p$$

then:

$$\begin{aligned}
& |\phi_t^{\nu^n}(p^n) - \phi_t^\nu(p)| \\
= \; & |p^n + \int_0^t b(\bar{\nu_s^n}, y, \phi_s^{\nu^n})ds - p - \int_0^t b(\bar{\nu}_s, y, \phi_s^\nu)ds| \\
\leq \; & |p^n - p| + \int_0^t |b(\bar{\nu_s^n}, y, \phi_s^{\nu^n}) - b(\bar{\nu_s^n}, y, \phi_s^\nu)|ds + \int_0^t |b(\bar{\nu_s^n}, y, \phi_s^\nu(p)) - b(\bar{\nu}_s, y, \phi_s^\nu)|ds \\
\leq \; & |p^n - p| + \int_0^t L|\phi_s^{\nu^n} - \phi_s^\nu|ds + \int_0^t |b(\bar{\nu_s^n}, y, \phi_s^\nu(p)) - b(\bar{\nu}_s, y, \phi_s^\nu)|ds
\end{aligned}$$

where the last inequality is true due to the Lipschitz continuity of $p \mapsto b(u, y, p)$. Since $\nu^n \to \nu$ and $p^n \to p$ the statement follows if the second integral tends to 0. Remember that $b(\nu, y, p)$ is continuous in ν. Additionally $p \mapsto b(\nu, y, p) - b(\tilde{\nu}, y, p)$ is Lipschitz continuous on the compact set $\triangle^{nd}$ and bilinear-quadratic. Therefore the maximum point $p^*(\nu)$ is a continuous function of ν and we conclude that

$$|b(\nu, y, p) - b(\tilde{\nu}, y, p)| \leq |h(\nu, y) - h(\tilde{\nu}, y)|$$

for a in ν continuous function $h(\nu, y)$. Hence we are able to apply the Grönwall-inequality and attain

$$|\phi_s^{\nu^n} - \phi_s^{\nu}| \leq e^{Ls}|p^n - p| + e^{Ls} \int_0^s e^{-Lt} |h(\bar{\nu_t^n}, y) - h(\bar{\nu_t}, y)| dt.$$

Let $p^n \to p$ and $\nu^n \to \nu$, then we see that $|\phi_s^{\nu^n} - \phi_s^{\nu}| \to 0$ and finally

$$|\phi_t^{\nu^n}(p^n) - \phi_t^{\nu}(p)| \to 0.$$

□

With this lemma we are able to state the existence of optimal relaxed strategies. Under an additional assumption we conclude the existence of an optimal deterministic strategy. Sufficient conditions for this assumptions are stated in the following remark.

Theorem 4.14

a) There exists an optimal relaxed strategy $\pi^* = (f^*, f^*, \ldots)$, in particular $f^*(y, p) \in A^{\mathcal{R}}$ and $V_{\infty,\pi^*}(y, p) = V_\infty(y, p)$.

b) If $u \mapsto F(p, y, u)$ is convex, where

$$F(p, y, u) := g(p, y, u) + \sum_{\omega \neq y} V_\infty(\omega, p + \Phi_{y\omega}(u, p)) q_{y\omega}^Y(u, p) + V_\infty(y, p)(\alpha - \sum_{\omega \neq y} q_{y\omega}^Y(u, p))$$

then

$$\bar{\pi^*} = (\bar{f}^*, \bar{f}^*, \ldots)$$

is an optimal deterministic strategy, especially $\bar{f}^*(y, p) \in A$ and

$$V_{\infty,\bar{\pi^*}}(y, p) = V_\infty(y, p).$$

Proof:

a) By corollary 4.8 it is sufficient to prove that there exists f^* with

$$\mathcal{T}_{f^*}V_\infty(y,p) = \inf_{\gamma\in A^{\mathcal{R}}}(LV_\infty)(y,p,\gamma).$$

We know from lemma 4.13 that $A^{\mathcal{R}}$ is compact. Additionally that $\nu \mapsto \phi_t^\nu(p)$ is continuous. Since $u \mapsto q_{kl}^Y(u,p)$ and $u \mapsto g(p,y,u)$ are continuous and $p \mapsto V_\infty(y,p)$ is concave, we conclude that

$$\nu \mapsto (LV_\infty)(y,p,\nu)$$

is continuous. Hence there exists a minimizer $f^*(y,p)$ of $(LV_\infty)(y,p,\nu)$. Since $u \mapsto (LV_\infty)(y,p,u)$ is continuous and $(LV_\infty)(y,p,u) \geq 0$ we conclude that $f^*(y,p)$ is measurable.

b) Since U is convex $\bar{f}^* \in U$. On the one hand we have

$$\mathcal{T}V_\infty(y,p) = \inf_{\gamma\in A}(LV_\infty)(y,p,\gamma) \geq \inf_{\nu\in A^{\mathcal{R}}}(LV_\infty)(y,p,\nu).$$

On the other hand we get:

$$\begin{aligned}(LV_\infty)(y,p,\nu) &= \int_0^\infty\int_U e^{-(\alpha+\beta)t}F(\phi_t^\nu(p),y,u)\nu(du)dt\\ &\overset{(4.5)}{=} \int_0^\infty\int_U e^{-(\alpha+\beta)t}F(\phi_t^{\bar\nu}(p),y,u)\nu(du)dt\\ &\geq \int_0^\infty e^{-(\alpha+\beta)t}F(\phi_t^{\bar\nu}(p),y,\bar\nu)dt = (LV_\infty)(y,p,\bar\nu).\end{aligned}$$

Hence

$$\inf_{\nu\in A^{\mathcal{R}}}(LV_\infty)(y,p,\nu) \geq \inf_{\nu\in A^{\mathcal{R}}}(LV_\infty)(y,p,\bar\nu) = \inf_{\gamma\in A}(LV_\infty)(y,p,\gamma) = \mathcal{T}V_\infty(y,p).$$

Summarizing:

$$\mathcal{T}V_\infty(y,p) = \inf_{\nu\in A^{\mathcal{R}}}(LV_\infty)(y,p,\nu).$$

In particular we get:

$$\mathcal{T}V_\infty(y,p) = (LV_\infty)(y,p,f^*) \geq (LV_\infty)(y,p,\bar f^*) = \mathcal{T}_{\bar f^*}V_\infty(y,p),$$

thus $\bar f^*$ is a minimizer of $V_\infty(y,p)$. The measurability of $\bar f^*$ follows as in a).

□

Remark 4.15

a) *Due to theorem 4.7 we have by the existence of optimal (relaxed/deterministic) strategy of the MDP in theorem 4.14 the existence of an optimal (relaxed/deterministic) controls for the reduced problem* (P_{red}) *and hence for* (P).

b) $u \mapsto F(p, y, u)$ *is convex if*

- $g(p, y, u)$ *is convex in* u *and*
- $q_{kl}^{Y}(u, p)$ *is linear in* u *and*
- $\Phi_{kl}(u, p)$ *is independent of* u.

Denote by $V_{n,\pi}(y, p)$ the expected discounted cost over n periods under a fixed policy $\pi = (f_0, f_1, \ldots, f_{n-1}) \in F^n$. This means

$$V_{n,\pi}(y, p) := \mathbb{E}_\pi \left[\sum_{k=0}^{n-1} \delta^k r(Y_{\tau_k}, p_{\tau_k}, f_k(Y_{\tau_k}, p_{\tau_k})) \;\middle|\; Y_0 = y, p_0 = p \right].$$

Define by $V_n(y, p) := \sup_{\pi \in F^n} V_{n,\pi}(y, p)$ the n-period value function.

Corollary 4.16 *Under the assumption of theorem 4.14 b) holds:*

a) $\lim_{n\to\infty} V_n(y, p) = V_\infty(y, p)$

b) *If* f_n *is a minimizer of* V_{n-1} *then* $(\liminf_{n\to\infty} f_n)^\infty$ *is an optimal policy.*

Proof: Both statements are well-known MDP-results which are true under the continuity and compactness assumptions (Bertsekas and Shreve (1978)). □

In section 5.2.3 we demonstrate the power of the value iteration.

5 Application to Parallel Queueing with Incomplete Information

Queueing models are very popular, since they appear in various applications of the real world, for example in telecommunication, in the internet or in supply chain management. Additionally, the basic concepts of queueing models are treated in various publications and are well-understood. We skip them here and refer to the works of Asmussen (2003), Kitaev and Rykov (1995) and Brémaud (1981). Most of the queueing applications have in common, that they can be modelled as a SMDP, but that statements about the optimal control (or even properties of it) are quite hard to compute and to prove. In this section we recall our introductory example in section 1.1 and consider a parallel queueing model with two servers, which is illustrated by the following picture:

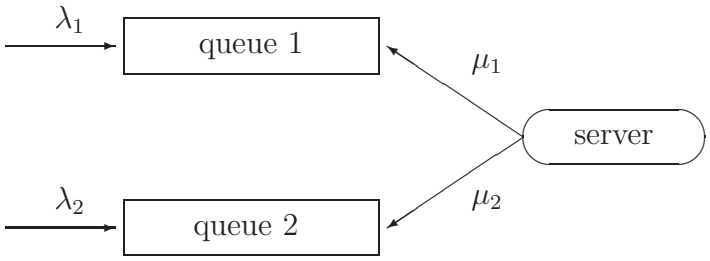


Figure 3: Parallel Queueing Model

There are two queues with infinite buffer, where customers arrive at queue i corresponding to a Poisson process with arrival rate λ_i, $i = 1, 2$. 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 i is $\exp(\mu_i)$-distributed, $i = 1, 2$. We assume that arrivals and service time are independent. For each waiting customer a cost at rate c_i, $i = 1, 2$, occurs and we want to find a service strategy for the server minimizing the expected discounted waiting costs over an infinite horizon.

In the next section we model these circumstances as a MDP. We then prove the optimality of an $c\mu$-rule in the complete information case as in Asmussen (2003) and extend his result to sufficient fine information structures. In the following sections we discuss various models under incomplete information, especially models where the information structure is not sufficiently fine anymore. First, in section 5.2 we discuss the case of unknown (Bayesian) service rates. In this part we make use of the solution technique derived in section 4.1 and 4.2. The main results in this incomplete information case are the explicit representation of the estimator process, a closed formula for the value function, the existence of an optimal (almost surely pure) control and sufficient conditions for optimal controls in several models. The same model with a reward criteria is solved completely. In section 5.3

the underlying information structure is given by a 0-1-observation about the number of waiting customers in queue 1. Here we investigate numerically the performance of several reasonable strategies.

5.1 The Model and the Complete Information Case

Denote by $\alpha := \lambda_1 + \lambda_2 + \mu_1 + \mu_2$ the uniformization parameter. Then the uniformized MDP for the above described parallel queueing model is defined under the uniformized distribution function of the holding times $\tau_{n+1} - \tau_n$

$$F(t) = \big(1 - e^{-\alpha t}\big)\mathbb{1}\big(t \geq 0\big)$$

by:

- *state space:* $S = \mathbb{N}_0^2$, $i = (i_1, i_2) \in S$ with
  $i_1 \;\hat{=}$ number of customers in queue 1
  $i_2 \;\hat{=}$ number of customers in queue 2
- *action space:* $A = \{a \mid a \in \{1, 2\}\}$ with
  $a = 1 \;\hat{=}$ serve queue 1
  $a = 2 \;\hat{=}$ serve queue 2
- *set of admissible actions:* $D(i) = A \quad \forall i \in S$
- *transition probabilities:*

$$q(i, a, j) = \begin{cases} \frac{\lambda_k}{\alpha} & j = i + e_k,\ k = 1, 2 \\ \frac{\mu_a}{\alpha}\delta(i_a) & j = i - e_a \\ 1 - \sum_{k=1}^{2} \frac{\lambda_k}{\alpha} - \frac{\mu_a}{\alpha}\delta(i_a) & j = i \\ 0 & \text{else} \end{cases}$$

  where $\delta(i) := \mathbb{1}\big(i > 0\big)$
- *cost function:* $r(i, a) = \frac{c_1 i_1 + c_2 i_2}{\alpha + \beta}$
- *discount factor:* $\delta = \frac{\alpha}{\alpha + \beta}$.

Since all processes are constant between two jumps it is sufficient to consider controls $a \in A$ which are constant between jumps (see e.g. Bertsekas and Shreve (1978)). Since A is finite all continuity and compactness conditions required for value and policy iteration are fulfilled.

The optimality equation is given by

$$\begin{aligned} v(i_1, i_2) &= \mathcal{T} v(i_1, i_2) \\ &= \frac{1}{\alpha + \beta}\Big[c_1 i_1 + c_2 i_2 + \lambda_1 v(i_1 + 1, i_2) + \lambda_2 v(i_1, i_2 + 1) \\ &\qquad + \min\big\{\mu_1 v((i_1 - 1)^+, i_2) + \mu_2 v(i_1, i_2); \\ &\qquad\qquad \mu_1 v(i_1, i_2) + \mu_2 v(i_1, (i_2 - 1)^+)\big\}\Big]. \end{aligned}$$

The next theorem uses an interchange argument for the proof of the optimal policy. It states that if the "expected costs" of queue 1 are greater than the costs of queue 2, that means if $c_1\mu_1 \geq c_2\mu_2$, the optimal policy always prefers to serve queue 1, if there is a customer waiting. Hence this strategy is called $c\mu$-policy. It arises in many queueing models with a linear cost structure and can easily be extended to the case of M parallel queues. We invest some efforts in the proof of this well-known result (see for example Asmussen (2003)), since the idea and the technique of this proof will be helpful later on.

Theorem 5.1 *Assume $c_1\mu_1 \geq c_2\mu_2$ then $\pi = f^\infty$ is optimal where*

$$f(i_1, i_2) = \begin{cases} 1 & i_1 > 0 \\ 2 & i_1 = 0 \end{cases}$$

Proof: We show, that for each horizon $N \in \mathbb{N}$ the $c\mu$-policy is optimal, this means $\pi^* = (f_N, \ldots, f_1)$ with

$$f_n(i_1, i_2) = f(i_1, i_2) = \begin{cases} 1 & i_1 > 0 \\ 2 & i_1 = 0 \end{cases}$$

is optimal. This will be done by proving that f_n is a minimizer of V_{n-1} and consequently $(f_N, \ldots, f_1)$ has to be optimal for the N-period model. Taking the limit the statement follows. Start with $N = 1$ and let $V_0 \equiv 0$, then

$$\mathcal{T}V_0(i_1, i_2) = \frac{c_1 i_1 + c_2 i_2}{\alpha + \beta},$$

thus every decision rule f is a minimizer of V_0 and hence optimal for $N = 1$. Assume f is a minimizer of $V_0, \ldots, V_{N-1}$ (in particular $(f, \ldots, f) \in F^N$ is optimal for the N-period model), then it remains to prove, that f is a minimizer of V_N (see corollary 4.16). By induction and the monotonicity of $\mathcal{T}$ we conclude that $V_n(i_1, i_2)$ is monotone increasing in i_1 and i_2 respectively, since:

$$V_n(i_1, i_2) = \mathcal{T}V_{n-1}(i_1, i_2) \leq \mathcal{T}V_{n-1}(i_1 + 1, i_2) = V_n(i_1, i_2).$$

Analogously: $V_n(i_1, i_2) \leq V_n(i_1, i_2 + 1)$.

From the optimality equation we conclude

$$\begin{aligned}
i_1 = 0, i_2 > 0: &\quad \mu_1 V_N(0, i_2) + \mu_2 V_N(0, i_2) \geq \mu_1 V_N(0, i_2) + \mu_2 V_N(0, i_2 - 1) \\
&\quad \Rightarrow \quad \text{the minimizer is } f_{N+1}(0, i_2) = 2 = f(0, i_2) \\
i_1 > 0, i_2 = 0: &\quad \mu_1 V_N(i_1 - 1, 0) + \mu_2 V_N(i_1, 0) \leq \mu_1 V_N(i_1, 0) + \mu_2 V_N(i_1, 0) \\
&\quad \Rightarrow \quad \text{the minimizer is } f_{N+1}(i_1, 0) = 1 = f(i_1, 0) \\
i_1 = 0, i_2 = 0: &\quad \text{the minimizer } f_{N+1}(0, 0) \text{ is arbitrarily.}
\end{aligned}$$

Now let $i_1, i_2 > 0$ and $g \equiv 2$. Compute then

$$\begin{aligned} & V_{N+1,(g,f,\dots,f)}(i_1,i_2) - V_{N+1,(f,g,f,\dots,f)}(i_1,i_2) \\ = & \Big(\mathcal{T}_g\big(\mathcal{T}_f V_{N-1,(f,\dots,f)}\big)\Big)(i_1,i_2) - \Big(\mathcal{T}_f\big(\mathcal{T}_g V_{N-1,(f,\dots,f)}\big)\Big)(i_1,i_2) \\ = & \frac{c_1\mu_1 - c_2\mu_2}{(\alpha+\beta)^2} \geq 0. \end{aligned} \tag{5.1}$$

Let h be an arbitrary decision rule with $h(i_1,i_2) = 2$. Using the just proven inequality we conclude

$$\begin{aligned} \mathcal{T}_h V_N(i_1,i_2) &= \mathcal{T}_h V_{N,(f,\dots,f)}(i_1,i_2) = V_{N+1,(h,f,\dots,f)}(i_1,i_2) = V_{N+1,(g,f,\dots,f)}(i_1,i_2) \\ &\geq V_{N+1,(f,g,f,\dots,f)}(i_1,i_2) = \mathcal{T}_f V_{N,(g,f,\dots,f)}(i_1,i_2) \geq \mathcal{T}_f V_N(i_1,i_2), \end{aligned}$$

where the last inequality is true due to $V_{N,(g,f,\dots,f)}(i_1,i_2) \geq V_N(i_1,i_2)$ and the monotonicity of $\mathcal{T}_f$. Hence $f(i_1,i_2)$ is a minimizer of $V_N(i_1,i_2)$, which finishes the proof. □

The next theorem offers an explicit representation of the n-period value function $V_n(i_1,i_2)$ and hence of $V_\infty(i_1,i_2) = \lim_{n\to\infty} V_n(i_1,i_2)$. We omit the proof, since it is a special case of a similar theorem in the incomplete information model (see theorem 5.13). It is done with the help of the value iteration.

Theorem 5.2 *For $i_1 > 0$ and $i_2 > 0$ the n-period value function $V_n(i_1,i_2)$ is given for $n \geq 1$ by*

$$V_n(i_1,i_2) = (c_1 i_1 + c_2 i_2)K_n + (c_1\lambda_1 + c_2\lambda_2)L_{n-1} - c_1\mu_1 L_{n-1} \tag{5.2}$$

where

$$\begin{aligned} K_n &= \frac{1}{\alpha+\beta}\sum_{k=0}^{n-1}\left(\frac{\alpha}{\alpha+\beta}\right)^k = \frac{1}{\alpha+\beta}\sum_{k=0}^{n-1}\delta^k \\ L_n &= \frac{1}{\alpha+\beta}\big(K_n + \alpha L_{n-1}\big) \text{ with } L_0 := 0. \end{aligned}$$

Thus the value function for the infinite-horizon problem is given by

$$V_\infty(i_1,i_2) = \lim_{n\to\infty} V_n(i_1,i_2) = (c_1 i_1 + c_2 i_2)K + (c_1\lambda_1 + c_2\lambda_2)L - c_1\mu_1 L$$

where $K := \lim_{n\to\infty} K_n = \frac{1}{\beta}$ and $L := \lim_{n\to\infty} L_n = \frac{1}{\beta^2} = \frac{K}{\beta}$ are well-defined (see theorem 5.14).

We see that the value function consists of three terms, which have a nice interpretation: the first one $(c_1 i_1 + c_2 i_2)K$ are waiting costs for customers in the queues, the second $(c_1\lambda_1 + c_2\lambda_2)L$ are "expected" waiting costs due to arrivals and the third $c_1\mu_1 L$ reduces the costs due to "expected" served customers.

Theorem 5.1 can be formulated more precisely in view of corollary 2.14. If one of the queues is empty, then it is always optimal to serve the other queue (independent of the

service rates). If in both queues at least one customer is waiting (the exact number of waiting customers does not matter) and if one has complete information about the service rates (we assume here without loss of generality that μ_i can only take two values, that means $\mu_i \in \{\mu_i^A, \mu_i^B\}$), then it is optimal to serve queue 1 if the product of cost rate c_1 times the current service rate μ_1 is greater than the product of cost rate c_2 and service rate at queue 2 given by μ_2, that means if $c_1\mu_1 \geq c_2\mu_2$. If the inequality holds in the other direction then it is optimal to serve queue 2. Let us state this in a corollary.

Corollary 5.3 *Let $(I(k), k = 1, \ldots, m)$ be an information structure with corresponding observation process Y_t such that*

$$p_t^{i_1}(0) := \mathbb{P}_u(\text{queue 1 empty} \mid \mathcal{F}_t^Y) \in \{0,1\} \qquad p_t^{i_2}(0) := \mathbb{P}_u(\text{queue 2 empty} \mid \mathcal{F}_t^Y) \in \{0,1\}$$
$$p_t(\mu_1^A) := \mathbb{P}_u(\mu_1 = \mu_1^A \mid \mathcal{F}_t^Y) \in \{0,1\} \qquad p_t(\mu_2^A) := \mathbb{P}_u(\mu_2 = \mu_2^A \mid \mathcal{F}_t^Y) \in \{0,1\}.$$

Define $\widehat{\mu_j} := \mu_j(p^{\mu_j}) := \mu_j^A p(\mu_j^A) + \mu_j^B(1 - p(\mu_j^A))$ for $j = 1, 2$. Then the $c\widehat{\mu}$-rule $\pi = f^\infty$ is optimal, where

$$f(p^{i_1}, p^{i_2}, p(\mu_1^A), p(\mu_2^A)) := \begin{cases} 1 & p^{i_1}(0) = 0, c_1\mu_1(p^{\mu_1}) > c_2\mu_2(p^{\mu_2}) \\ 2 & p^{i_2}(0) = 0, c_2\mu_2(p^{\mu_2}) > c_1\mu_1(p^{\mu_1}) \\ 1 & p^{i_2}(0) = 1 \\ 2 & p^{i_1}(0) = 1 \end{cases}$$

If $c_1\mu_1(p^{\mu_1}) = c_2\mu_2(p^{\mu_2})$ then the optimal service allocation can be chosen arbitrarily.

Remark 5.4

a) *Extend the state space to $S_X \times S_Z$, remember remark 2.2. Then a sufficient fine information structure has to have complete information about the service rates and a 0-1-group observation about the queues. In mathematical words it has to be at least as*

$$I(1) = \left\{(0, i_2, \mu_1, \mu_2), i_2 \in \mathbb{N}_0, \mu_1 \in \{\mu_1^A, \mu_1^B\}, \mu_2 \in \{\mu_2^A, \mu_2^B\}\right\},$$
$$I(2) = \left\{(i_1, 0, \mu_1, \mu_2), i_1 \in \mathbb{N}, \mu_1 \in \{\mu_1^A, \mu_1^B\}, \mu_2 \in \{\mu_2^A, \mu_2^B\}\right\},$$
$$I(3) = \left\{(i_1, i_2, \mu_1^A, \mu_2^A), i_1, i_2 \in \mathbb{N}\right\}, I(4) = \left\{(i_1, i_2, \mu_1^B, \mu_2^A), i_1, i_2 \in \mathbb{N}\right\},$$
$$I(5) = \left\{(i_1, i_2, \mu_1^A, \mu_2^B), i_1, i_2 \in \mathbb{N}\right\}, I(6) = \left\{(i_1, i_2, \mu_1^B, \mu_2^B), i_1, i_2 \in \mathbb{N}\right\}.$$

b) *Notice that if for example $c_1 \min\{\mu_1^A, \mu_1^B\} \geq c_2 \max\{\mu_2^A, \mu_2^B\}$, then it is not even necessary to observe the service rate. In this cases queue 1 is always better than queue 2. Therefore it is sufficient to observe only if at queue 1 a customer is waiting or not (means 0-1-observation).*

But what happens if the service parameters μ_1 and μ_2 are unknown? This may be due to the fact that two different kind of customers can arrive at queue j requiring service rates μ_j^A and μ_j^B respectively and the server is not able to observe, which kind of customer arrives. We will investigate this setting in the next section.

5.2 Unknown Service Rates: the Bayesian Case

In this section we consider the case when the service rates are Bayesian, this means $\mu_j \in \{\mu_j^A, \mu_j^B\}$ unknown, but constant. 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 complete information model in the last section. There we have seen that it is never optimal to split service, thus it was no real restriction. We interpret $u \in U$ as the service capacity spent to queue 1. Hence $1 - u$ is spent to queue 2. Sometimes we will write $u(1)$ and $u(2)$ for the service rates given to queue 1 and 2, respectively. With a slight abuse of notation we will also write $u = 2$ for serving queue 2 exclusively.

We will see in the following, that the conditions of remark 4.15 are satisfied. Hence it is not necessary to distinguish relaxed and deterministic controls as pointed out in theorem 4.14 and the existence of an optimal deterministic control is guaranteed. Sometimes we will even state existence and properties of an optimal pure control, which serves one queue exclusively.

First we derive in section 5.2.1 an explicit representation of the estimator processes for the unknown service rates. This will be done with the help of the general representation theorem 3.5. In this context we are able to solve the partial differential equation, which makes it easier to prove monotonicity and continuity properties of the estimator process between two jumps. After introducing the associated MDP in the sense of section 4.2, according to the one with complete information in section 5.1 we develop in section 5.2.3 with the value iteration a characterization of the value function in the model with unknown service rates, see theorem 5.13. We will see that it is quite similar to the one under complete information given in theorem 5.2. Furthermore we prove in theorem 5.18 that the optimal control serves one queue exclusively most of the time with the help of the generalized HJB-equation introduced in section 4.1.

In the consecutive two sections we specify our model to the symmetric case and the case, in which one service rate is known. In the first one the optimal control is a control limit rule with control limit $p^* = \frac{1}{2}$ (in particular here the certainty equivalence principle is true). This will be shown in theorem 5.20 with the help of the verification technique of section 4.1. In the latter setup we characterize the optimal strategy as a control limit rule with implicit defined control limit and state sufficient conditions for the optimality of a control. Here we demonstrate a interchange argument as in the proof of theorem 5.1. In both settings the stay-on-a-winner property is obtained. In section 5.2.6 we change the objective function, where we consider not anymore a model with waiting costs, but a model with rewards for each served customer. In this case the model is completely solved, see theorem 5.29.

Assume now that $\mu_1 \in \{\mu_1^A, \mu_1^B\}$ and $\mu_2 \in \{\mu_2^A, \mu_2^B\}$ are unknown, but constant over time. Only known is the a-priori-distribution of μ_1 and μ_2, that means $p_0 = \mathbb{P}(\mu_1 = \mu_1^A)$ and $q_0 = \mathbb{P}(\mu_2 = \mu_2^A)$. All other state and parameter processes are observable, hence we are in the Bayesian case. In the context of section 2.1 we use the following notations:

- $S_Z = \{g_1, g_2\} \times \{g_1, g_2\}$, $Q^Z \equiv 0$
- $S_X = \{e_1^1, e_2^1, \ldots\} \times \{e_1^2, e_2^2, \ldots\}$ (in particular $S_X = \mathbb{N}_0^2$), $\delta_{ij}^{\mu\nu} \equiv 0$ for all i, j, μ, ν
- $I(k) = \{k\}$, hence $Y_t \equiv X_t$ and $\mathcal{F}_t^Y \equiv \mathcal{F}_t^X$.

Due to the optimality of the $c\mu$-rule in corollary 5.3 only the following three case are relevant:

$$\begin{array}{lllll} \text{A)} & c_1\mu_1^A < c_2\mu_2^A & c_1\mu_1^B > c_2\mu_2^B & c_1(\mu_1^A - \mu_1^B) < c_2(\mu_2^A - \mu_2^B) < 0 \\ \text{B)} & c_1\mu_1^A < c_2\mu_2^A & c_1\mu_1^B > c_2\mu_2^B & c_1(\mu_1^A - \mu_1^B) < c_2(\mu_2^A - \mu_2^B) = 0 \\ \text{C)} & c_1\mu_1^A < c_2\mu_2^A & c_1\mu_1^B > c_2\mu_2^B & 0 < c_2(\mu_2^A - \mu_2^B) \leq -c_1(\mu_1^A - \mu_1^B) \end{array} \tag{5.3}$$

Case A) can be simplified to $c_1\mu_1^B > c_2\mu_2^B > c_2\mu_2^A > c_1\mu_1^A$. Case B) implies that $\mu_2^A = \mu_2^B$, in particular the service parameter μ_2 is known. We will treat this case later on in 5.2.5 in more detail. Case C) with equality in the last conditions will be topic of section 5.2.4 and referred to as the symmetric case. All other cases are completely covered by the $c\mu$-rule, for example

$$c_1\mu_1^B > c_1\mu_1^A > c_2\mu_2^B > c_2\mu_2^A$$

implies that the optimal controller always prefers queue 1.

5.2.1 The Estimator Process

In this section we analyze the estimator processes $p_t = \mathbb{P}_u(\mu_1 = \mu_1^A \mid \mathcal{F}_t^X)$ and $q_t := \mathbb{P}_u(\mu_2 = \mu_2^A \mid \mathcal{F}_t^X)$.

The $\mathcal{F}_t^X$-generator of (X_t) is given as in section 3.1 corresponding to section 5.1 by $Q^X(u,p,q) = \big(q_{ij}^X(u,p,q)\big)$ with

$$q_{ij}^X(u,p,q) = \begin{cases} \mu_1(p)u & j = i - e_1 \\ \mu_2(q)(1-u) & j = i - e_2 \\ \lambda_1 & j = i + e_1 \\ \lambda_2 & j = i + e_2 \\ -\lambda_1 - \lambda_2 - \mu_1(p)u - \mu_2(q)(1-u) & j = i \end{cases}$$

where $\mu_1(p) := \mu_1^A p + \mu_1^B(1-p)$ is the estimated service rate (the conditional mean) at queue 1 at time t, similar $\mu_2(q) := \mu_2^A q + \mu_2^B(1-q)$.

Theorem 5.5 *The estimator process p_t is the unique solution of*

$$dp_t \;=\; u_t(1)\big(\mu_1^B - \mu_1^A\big)p_t(1-p_t)\mathbb{1}\big(X_t^1 > 0\big)dt + \Phi_1(p_{t-})dN_t^{X_1}(X_{t-}^1, X_{t-}^1 - 1) \tag{5.4}$$

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

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

Similar q_t is described as the unique solution of

$$dq_t = u_t(2)\big(\mu_2^B - \mu_2^A\big)q_t(1-q_t)\mathbb{1}\big(X_t^2 > 0\big)dt + \Phi_2(q_{t-})dN_t^{X_2}(X_{t-}^2, X_{t-}^2 - 1)$$

where the control independent jump-size $\Phi_2(q)$ is given by

$$\Phi_2(q) = \frac{1}{\mu_2(q)}\mu_2^A q - q.$$

Proof: The assertions are a direct consequence of theorem 3.5. □

The following results hold in a similar fashion for q_t but we will omit them here. 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^A and μ_1^B with respect to the pre-jump-estimate p_{t-}, that means

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

We omit in the following $\mathbb{1}\big(X_t^1 > 0\big)$ since it is obvious that new information is only available through service, which is reasonable only if customers are waiting in queue 1. Between two jumps p_t is described by the deterministic part of (5.4)

$$\dot{p} = u(1)(\mu_1^B - \mu_1^A)p(1-p) \tag{5.5}$$

and can be calculated explicitly.

Theorem 5.6 *Denote by τ_n the n-th jump time of (p_t, q_t). For $t \in [\tau_n, \tau_{n+1})$ holds $p_t = \phi^u_{t-\tau_n}(p_{\tau_n})$ where*

$$\phi_t^u(p) = \frac{\exp\{(\mu_1^B - \mu_1^A)\int_0^t u_s(1)ds\}}{\exp\{(\mu_1^B - \mu_1^A)\int_0^t u_s(1)ds\} - \frac{p-1}{p}}. \tag{5.6}$$

In particular: $\phi_0^u(p) = p$.

Proof: The first statement is clear by theorem 3.9. The second one follows by differentiating (5.6) with $M(t) := (\mu_1^B - \mu_1^A)\int_0^t u_s(1)ds$:

$$\begin{aligned}
\frac{\partial}{\partial t}\phi_t^u(p) &= \frac{1}{\Big(\exp(M(t)) - \frac{p}{p-1}\Big)^2}\Bigg[\left(\exp(M(t)) - \frac{p-1}{p}\right)\exp(M(t))\left(\mu_1^B - \mu_1^A\right)u_t(1) \\
&\qquad - \exp(M(t))\exp(M(t))\left(\mu_1^B - \mu_1^A\right)u_t(1)\Bigg] \\
&= \frac{1}{\Big(\exp(M(t)) - \frac{p}{p-1}\Big)^2}\left[-\exp(M(t))\left(\mu_1^B - \mu_1^A\right)u_t(1)\frac{p-1}{p}\right] \\
&= u_t(1)\big(\mu_1^B - \mu_1^A\big)\phi_t^u(p)(1-\phi_t^u(p))
\end{aligned}$$

which is exactly (5.5) and obviously $\phi_0^u(p) = p$. □

Remark 5.7

1) *If $p = 1$ then $\phi_t^u(p) \equiv 1$. On the other hand if $p = 0$ then $\phi_t^u(p) \equiv 0$. Summarizing if we have complete information about the real value of μ_1 then the estimator does not change between two jumps. It will also remain constant due to jumps shown later on. Thus it will be constant over time and the complete information is not destroyed.*

2) *If $u_s(2) \equiv 1$ (means only queue 2 is served) for $s \in [t, t+\varepsilon]$ then $\phi_s^u(p) \equiv \phi_t^u(p)$ for all $s \in [t, t+\varepsilon]$. Thus the estimator p_t is updated only if queue 1 is served.*

3) *p_t is independent of the length of the queues i_1, i_2.*

4) *$T_t(1) := \int_0^t u_s(1)ds$ denotes the time spent for serving queue 1 in $[0, t]$.*

We assume now without loss of generality

$$\mu_1^B > \mu_1^A \tag{5.7}$$

and investigate the behaviour of $\phi_t^u(p)$ in dependence of t.

Lemma 5.8

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 $p \in [0, 1]$.*

Proof:

a) The derivative $\frac{\partial}{\partial t}\phi_t^u(p)$ is by theorem 5.6 given by $u_t(1)(\mu_1^B - \mu_1^A)\phi_t^u(p)(1 - \phi_t^u(p))$ which is greater or equal 0 due to assumption (5.7).

b) The statement is a direct consequence of theorem 3.9.

□

By part a) if $u_s \equiv 1$ for $s \in [t, t+\varepsilon]$ we have strong monotonicity for $p \in (0, 1)$. But if $u_s \equiv 2$ then $\phi_s^u(p)$ is constant in $[t, t+\varepsilon]$. In other words: if one serves queue 1 (that means $u_s = 1$) then the parameter μ_1^A becomes more likely. This is reasonable since $\mu_1^B > \mu_1^A$ which means that a service under rate $\mu_1 = \mu_1^B$ tends to result in an earlier completion than under rate $\mu_1 = \mu_1^A$.

After considering the estimator process in dependence of time, we analyze it in dependence of the a-priori-estimate p. The greater the a-priori-probability p, the greater is the estimate $\phi_t^u(p)$ up to the next jump as the following lemma claims in part a).

Lemma 5.9

a) $p \mapsto \phi_t^u(p)$ *is monotone increasing for all* $t \geq 0$.

b) $p \mapsto \phi_t^u(p)$ *is Lipschitz continuous uniformly in* u *for all* $t \geq 0$ *with Lipschitz parameter* $\exp\{(\mu_1^B - \mu_1^A)t\}$ *which is* < 1 *for* $t > 0$. *Hence it is contractive.*

c) $p \mapsto \phi_t^u(p)$ *is concave for all* $t \geq 0$.

Proof:

a) By differentiation we get:

$$\frac{\partial}{\partial p}\phi_t^u(p) = \frac{\exp\{(\mu_1^B - \mu_1^A)\int_0^t u_s(1)ds\} \cdot \frac{1}{p^2}}{\left(\exp\{(\mu_1^B - \mu_1^A)\int_0^t u_s(1)ds\} - \frac{p-1}{p}\right)^2} \geq 0.$$

b) We first note that for the denominator of $\phi_t^u(p)$ in (5.6) holds, that

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

and we conclude with $p, q \in [0,1]$:

$$\begin{aligned}
& |\phi_t^u(p) - \phi_t^u(q)| \\
= & \left| \frac{p\exp\{(\mu_1^B - \mu_1^A)\int_0^t u_s(1)ds\}}{p\exp\{(\mu_1^B - \mu_1^A)\int_0^t u_s(1)ds\} - p + 1} - \frac{q\exp\{(\mu_1^B - \mu_1^A)\int_0^t u_s(1)ds\}}{q\exp\{(\mu_1^B - \mu_1^A)\int_0^t u_s(1)ds\} - q + 1} \right| \\
= & \left| \frac{(p-q)\exp\{(\mu_1^B - \mu_1^A)\int_0^t u_s(1)ds\}}{\left(p\exp\{(\mu_1^B - \mu_1^A)\int_0^t u_s(1)ds\} - p + 1\right)\left(q\exp\{(\mu_1^B - \mu_1^A)\int_0^t u_s(1)ds\} - q + 1\right)} \right| \\
\leq & \left| (p-q)\exp\{(\mu_1^B - \mu_1^A)\int_0^t u_s(1)ds\} \right| \\
\leq & \exp\{(\mu_1^B - \mu_1^A)t\}|p-q|
\end{aligned}$$

c) Define $K := \exp\{(\mu_1^B - \mu_1^A)\int_0^t u_s(1)ds\} > 0$ then:

$$\frac{\partial^2}{\partial^2 p}\phi_t^u(p) = \frac{-2\left((K - \frac{p-1}{p}\right)^2 K\frac{1}{p^3} - 2K\frac{1}{p^2}\left(K - \frac{p-1}{p}\right)\frac{1}{p^2}}{\left(K - \frac{p-1}{p}\right)^4} \leq 0,$$

since $p \in [0,1]$ and $\frac{p-1}{p} \leq 0$.

□

As an immediate consequence of the last statements we get that under $\mu_1^B > \mu_1^A$ (see (5.7)):

- $t \mapsto \mu_1(\phi_t^u(p))$ is monotone decreasing
- $p \mapsto \mu_1(\phi_t^u(p))$ is monotone decreasing.

After we discussed the behaviour of p_t between two jumps in the last lemma we analyze now the jump behaviour of p_t, where the jump-size is independent of the current control. It can be proven, that a jump reduces the probability that μ_1^A is the true parameter, since $\mu_1^B > \mu_1^A$ by assumption (5.7), thus jumps are more likely under the hypothesis $\mu_1 = \mu_1^B$.

Lemma 5.10

a) $p + \Phi_1(p) = \frac{1}{\mu_1(p)}\mu_1^A p < p$ *if* $p \in (0,1)$, *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) If $p \in (0,1)$ (the case $p \in \{0,1\}$ is discussed after this proof) then due to (5.7)

$$\mu_1(p) = \mu_1^A p + \mu_1^B(1-p) > \mu_1^A$$

and hence $\frac{\mu_1^A}{\mu_1(p)} < 1$.

b) First note that for $p = 1$ and $p = 0$ the jump-size is equal to 0 and p_t is constant. $p + \Phi_1(p)$ is the new state of the conditional probability after a jump and it holds:

$$\begin{aligned}\frac{\partial}{\partial p}\left(p + \Phi_1(p)\right) &= \frac{\left((\mu_1^A - \mu_1^B)p + \mu_1^B\right)\mu_1^A - \mu_1^A p(\mu_1^A - \mu_1^B)}{\left((\mu_1^A - \mu_1^B)p + \mu_1^B\right)^2} \\ &= \frac{\mu_1^B \mu_1^A}{\left((\mu_1^A - \mu_1^B)p + \mu_1^B\right)^2} \geq 0.\end{aligned}$$

c) Since $\mu_1(p) \geq \mu_1^A$ and

$$\mu_1(\phi_t^u(q))\phi_t^u(p) - \mu_1(\phi_t^u(p))\phi_t^u(q) = \mu_1^B(\phi_t^u(p) - \phi_t^u(q))$$

we conclude

$$\begin{aligned}|\phi_t^u(p) + \Phi_1(\phi_t^u(p)) - \phi_t^u(q) - \Phi_1(\phi_t^u(q))| &= \left|\frac{\mu_1^A \phi_t^u(p)}{\mu_1(\phi_t^u(p))} - \frac{\mu_1^A \phi_t^u(q)}{\mu_1(\phi_t^u(q))}\right| \\ = \left|\frac{\mu_1(\phi_t^u(q))\mu_1^A\phi_t^u(p) - \mu_1(\phi_t^u(p))\mu_1^A\phi_t^u(q)}{\mu_1(\phi_t^u(p))\mu_1(\phi_t^u(q))}\right| &\leq \frac{\mu_1^B}{\mu_1^A}\left|\phi_t^u(p) - \phi_t^u(q)\right|.\end{aligned}$$

□

We obtain from part a): if $p = 0$ then $0 + \Phi_1(0) = 0$ and if $p = 1$ then $1 + \Phi_1(1) = 1$ (remember remark 5.7). 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 $\mu_1^A = 0$ we have $p_t = p_{t-} + \Phi_1(p_{t-}) = \frac{1}{\mu_1(p_{t-})}\mu_1^A p_{t-} = 0$. Hence after a jump (which is impossible under the hypothesis $\mu_1 = \mu_1^A$) the conditional probability $p_t \equiv 0$ for all upcoming time-points t.
Part b) reads as the greater the a-priori-probability p_{t-} the greater the a-posteriori-probability p_t after a jump. In particular we saw in the proof that under the hypothesis $\mu_1 = \mu_1^A = 0$ the function $p \mapsto p + \Phi_1(p)$ is constant, which is not astonishing, since if $\mu_1^A = 0$ after the first jump the conditional probability for μ_1^A has to be zero (since under the hypothesis $\mu_1^A = 0$ a jump never occurs).
Part c) is the analogon to lemma 5.9 where we proved the Lipschitz continuity in p. Similar to the proof we conclude the Lipschitz continuity in $p \mapsto p + \Phi_1(p)$.

5.2.2 The Reduced MDP

We introduce next the reduced (uniformized) MDP as in section 4.2. For this denote

$$\alpha := \lambda_1 + \lambda_2 + \mu_1^A + \mu_1^B + \mu_2^A + \mu_2^B$$

the uniformization parameter. Hence the distribution function of the sojourn times $\tau_{n+1} - \tau_n$ is given by

$$F(t) = \big(1 - e^{-\alpha t}\big)\, \mathbb{1}\big(t \geq 0\big).$$

Then define the MDP as follows (see remark 4.10):

- state space: $S = \mathbb{N}_0^2 \times [0,1]^2$, $(i_1, i_2, p, q) \in S$ with
  $i_1 \;\hat{=}$ number of waiting customers in queue 1
  $i_2 \;\hat{=}$ number of waiting customers in queue 2
  $p \;\hat{=}$ conditional probability for $\mu_1 = \mu_1^A$ discussed in section 5.2.1
  $q \;\hat{=}$ conditional probability for $\mu_2 = \mu_2^A$

- action space: $A = \{\gamma \mid \gamma : \mathbb{R}_+ \to [0,1]\}$, where $\gamma \in A$ denotes the fraction of service capacity spent to queue 1 (sometimes we will write $\gamma(1)$ and $\gamma(2) := 1 - \gamma(1)$ and with a slight abuse of notation $\gamma = 2$ for serving queue 2 exclusively)

- set of admissible actions: $D(i_1, i_2, p, q) = A \quad \forall (i_1, i_2, p, q) \in S$.

Between two jumps describe $\phi_t^\gamma(p)$ as the unique solution of (5.5)

$$\dot{p} = \gamma(1)(\mu_1^B - \mu_1^A)p(1-p)$$

with initial condition $p_0 = p$ given by

$$\phi_t^\gamma(p) = \frac{\exp\{(\mu_1^B - \mu_1^A)\int_0^t \gamma_s(1)ds\}}{\exp\{(\mu_1^B - \mu_1^A)\int_0^t \gamma_s(1)ds\} - \frac{p-1}{p}}$$

as in theorem 5.6. $\varphi_t^\gamma(q)$ corresponding to q_t is the unique solution of

$$\dot{q} = \gamma(2)(\mu_2^B - \mu_2^A)q(1-q).$$

We continue with the definition of the MDP:

- transition probabilities: for $\omega = (i_1, i_2, p, q)$ and $B \subset [0,1]^2$ set

$$\begin{aligned}
q\big(\omega, \gamma, (i_1+1, i_2, B)\big) &= \int_0^\infty e^{-\alpha t}\lambda_1 \mathbb{1}\big((\phi_t^\gamma(p), \varphi_t^\gamma(q)) \in B\big)dt \\
q\big(\omega, \gamma, (i_1, i_2+1, B)\big) &= \int_0^\infty e^{-\alpha t}\lambda_2 \mathbb{1}\big((\phi_t^\gamma(p), \varphi_t^\gamma(q)) \in B\big)dt \\
q\big(\omega, \gamma, (i_1, i_2, B)\big) &= 1 - \sum_{\substack{k \in \{(i_1+1,i_2),(i_1,i_2+1),\\ (i_1-1,i_2),(i_1,i_2-1)\}}} q\big(\omega, \gamma, (k, B)\big) \\
q\big(\omega, \gamma, (i_1-1, i_2, B)\big) &= \int_0^\infty e^{-\alpha t}\mu_1(\phi_t^\gamma(p))\mathbb{1}\big(i_1 > 0\big)\gamma_t(1) \cdot \\
&\qquad \cdot \mathbb{1}\big(\big(\phi_t^\gamma(p) + \Phi_1(\phi_t^\gamma(p)), \varphi_t^\gamma(q)\big) \in B\big)dt \\
q\big(\omega, \gamma, (i_1, i_2-1, B)\big) &= \int_0^\infty e^{-\alpha t}\mu_2(\varphi_t^\gamma(q))\mathbb{1}\big(i_2 > 0\big)\gamma_t(2) \cdot \\
&\qquad \cdot \mathbb{1}\big(\big(\phi_t^\gamma(p), \varphi_t^\gamma(q) + \Phi_2(\varphi_t^\gamma(q))\big) \in B\big)dt.
\end{aligned}$$

  All other transitions have probability 0.

- cost function: $r\big((i_1, i_2, p, q), \gamma\big) = r(i_1, i_2) = \frac{c_1 i_1 + c_2 i_2}{\alpha + \beta}$
- discount factor: $\delta = \frac{\alpha}{\alpha+\beta} \in (0,1)$.

Define as in section 4.2

$$V_{\infty,\pi}(i_1, i_2, p, q) := \mathbb{E}_\pi\left[\sum_{k=0}^\infty \delta^k r(X_{\tau_k}) \;\middle|\; X_0 = (i_1, i_2), p_0 = p, q_0 = q\right]$$

and

$$V_\infty(i_1, i_2, p, q) := \inf_{\pi \in F^\infty} V_{\infty,\pi}(i_1, i_2, p, q).$$

Remark 5.11 *It is obvious that it is never optimal to serve an empty queue while in the other queue customers are waiting. Note that by serving an empty queue there is no new information available about the service parameter of this queue. This means the (pure) optimal decision $f^* = f^*(i_1, i_2, p, q)$ in the case $i_1 = 0$ or $i_2 = 0$ is given by*

$$f^*(i_1, i_2, p, q) = \begin{cases} 2 & i_1 = 0, i_2 > 0 \\ 1 & i_2 = 0, i_1 > 0 \\ \textit{arbitrarily} & i_1 = i_2 = 0 \end{cases}$$

Hence we concentrate in the following to the case $i_1 > 0, i_2 > 0$. The optimality equation is accordingly given by:

$$\begin{aligned} v(i_1, i_2, p, q) = &\ \mathcal{T} v(i_1, i_2, p, q) \\ = &\ \inf_{\gamma \in A} \Big\{ \int_0^\infty e^{-(\alpha+\beta)t} \big\{ - (c_1 i_1 + c_2 i_2) \\ &+ v\big(i_1 + 1, i_2, \phi_t^\gamma(p), \varphi_t^\gamma(q)\big)\lambda_1 + v\big(i_1, i_2 + 1, \phi_t^\gamma(p), \varphi_t^\gamma(q)\big)\lambda_2 \\ &+ v\big(i_1, i_2, \phi_t^\gamma(p), \varphi_t^\gamma(q)\big)\big(\alpha - (\lambda_1 + \lambda_2 + \mu_1(\phi_t^\gamma(p))\gamma_t(1) + \mu_2(\varphi_t^\gamma(q)))\gamma_t(2)\big) \\ &+ v\big(i_1 - 1, i_2, \phi_t^\gamma(p) + \Phi_1(\phi_t^\gamma(p)), \varphi_t^\gamma(q)\big)\mu_1(\phi_t^\gamma(p))\gamma_t(1) \\ &+ v\big(i_1, i_2 - 1, \phi_t^\gamma(p), \varphi_t^\gamma(q) + \Phi_2(\varphi_t^\gamma(q))\big)\mu_2(\varphi_t^\gamma(q))\gamma_t(2)\big\} dt \Big\}. \end{aligned}$$

Remark 5.12 *Since $u \mapsto q_{ij}^X(u, p, q)$ is linear, the cost rate is control independent, $\Phi_1(p)$ and $\Phi_2(q)$ are independent of the control too and $U = [0, 1] \subset \mathbb{R}$ is convex there exists an optimal deterministic strategy π as proven in theorem 4.14 and remark 4.15. Furthermore due to corollary 4.16 we have $\lim_{n\to\infty} V_n(i_1, i_2, p, q) = V_\infty(i_1, i_2, p, q)$.*

5.2.3 A Characterization of the Value Function and the Optimal Control

The following theorem offers an explicit characterization of the value function. Assume here again that $i_1 > 0$ and $i_2 > 0$. Otherwise the optimal control is clear, since serving an empty queue is never advantageously, see remark 5.11.

Theorem 5.13 *The n-stage value function $V_n(i_1, i_2, p, q)$ is given for $n \geq 1$ by*

$$V_n(i_1, i_2, p, q) = (c_1 i_1 + c_2 i_2)K_n + (c_1\lambda_1 + c_2\lambda_2)L_{n-1} + G_{n-1}(p, q),$$

where

$$
\begin{aligned}
K_n &= \frac{1}{\alpha+\beta}\sum_{k=0}^{n-1}\left(\frac{\alpha}{\alpha+\beta}\right)^k = \frac{1}{\alpha+\beta}\sum_{k=0}^{n-1}\delta^k \\
L_n &= \frac{1}{\alpha+\beta}\big(K_n+\alpha L_{n-1}\big) \text{ with } L_0 := 0 \\
G_n(p,q) &= \inf_{\gamma\in A}\Bigg\{\int_0^\infty e^{-(\alpha+\beta)t}\Big\{\Big(G_{n-1}(\phi_t^\gamma(p)+\Phi_1(\phi_t^\gamma(p)),\varphi_t^\gamma(q)) - G_{n-1}(\phi_t^\gamma(p),\varphi_t^\gamma(q)) \\
&\qquad -c_1K_{n-1}\Big)\mu_1(\phi_t^\gamma(p))\gamma_t(1) \\
&\qquad +\Big(G_{n-1}(\phi_t^\gamma(p),\varphi_t^\gamma(q)+\Phi_2(\varphi_t^\gamma(q))) - G_{n-1}(\phi_t^\gamma(p),\varphi_t^\gamma(q)) \\
&\qquad -c_2K_{n-1}\Big)\mu_2(\varphi_t^\gamma(q))\gamma_t(2) \\
&\qquad +G_{n-1}(\phi_t^\gamma(p),\varphi_t^\gamma(q))\alpha\Big\}dt\Bigg\} \\
&\qquad \text{with } G_0(p) := 0.
\end{aligned}
$$

The same result holds for pure controls by replacing $A = \{\gamma : \mathbb{R}_+ \to [0,1]\}$ by $A = \{a \mid a \in \{1,2\}\}$. But the convergence of $V_n(i_1,i_2,p,q)$ to $V_\infty(i_1,i_2,p,q)$ we consider in theorem 5.14, may not be true anymore for the pure control model in general.

Proof: For $n = 1$ and $V_0 \equiv 0$ we get:

$$
V_1(i_1,i_2,p,q) = \mathcal{T}V_0(i_1,i_2,p,q) = \inf_{\gamma\in A}\left\{\int_0^\infty e^{-(\alpha+\beta)t}\big(c_1i_1+c_2i_2\big)dt\right\} = \big(c_1i_1+c_2i_2\big)K_1.
$$

Assume now $V_{n-1}(i_1,i_2,p,q) = (c_1i_1+c_2i_2)K_{n-1} + (c_1\lambda_1+c_2\lambda_2)L_{n-2} + G_{n-2}(p,q)$ for some $n-1 \in \mathbb{N}$. Then the statement follows by induction, since

$$
\begin{aligned}
& V_n(i_1,i_2,p,q) = \mathcal{T}V_{n-1}(i_1,i_2,p,q) \\
= \;& (c_1i_1+c_2i_2)\frac{1}{\alpha+\beta}(1+\alpha K_{n-1}) + (c_1\lambda_1+c_2\lambda_2)\frac{1}{\alpha+\beta}(K_{n-1}+\alpha L_{n-2}) \\
& + \inf_{\gamma\in A}\Bigg\{\int_0^\infty e^{-(\alpha+\beta)t}\Big\{\big(G_{n-2}(\phi_t^\gamma(p)+\Phi_1(\phi_t^\gamma(p)),\varphi_t^\gamma) - G_{n-2}(\phi_t^\gamma,\varphi_t^\gamma(q)) \\
&\qquad -c_1K_{n-1}\big)\mu_1(\phi_t^\gamma(p))\gamma_t(1) \\
&\qquad +\big(G_{n-2}(\phi_t^\gamma(p),\varphi_t^\gamma+\Phi_2(\varphi_t^\gamma(q))) - G_{n-2}(\phi_t^\gamma,\varphi_t^\gamma(q)) \\
&\qquad -c_2K_{n-1}\big)\mu_2(\varphi_t^\gamma(q))\gamma_t(2) + G_{n-2}(\phi_t^\gamma(p),\varphi_t^\gamma(q))\alpha\Big\}dt\Bigg\}
\end{aligned}
$$

and

$$\begin{aligned} K_n &:= \frac{1}{\alpha+\beta}\big(1+\alpha K_{n-1}\big) = \frac{1}{\alpha+\beta}\sum_{k=0}^{n-1}\left(\frac{\alpha}{\alpha+\beta}\right)^k \\ L_n &:= \frac{1}{\alpha+\beta}\big(K_n+\alpha L_{n-1}\big) = \frac{1}{\alpha+\beta}\sum_{k=0}^{n-1}\left(\frac{\alpha}{\alpha+\beta}\right)^k K_{n-k}. \end{aligned}$$

□

For the next lemma remember that the continuous and compactness assumptions under the use of deterministic controls are fulfilled, see remark 5.12.

Theorem 5.14 *It holds:*

a) K_n *is monotone increasing in* n *and* $K := \lim_{n\to\infty} K_n = \frac{1}{\beta}$

b) L_n *is bounded for all* n*, monotone increasing in* n *and* $L := \lim_{n\to\infty} L_n = \frac{1}{\beta^2}$.

c) $V_\infty(i_1,i_2,p,q) = \lim_{n\to\infty} V_n(i_1,i_2,p,q) = (c_1i_1+c_2i_2)K + (c_1\lambda_1+c_2\lambda_2)L + G(p,q)$*, where* $G(p,q) := \lim_{n\to\infty} G_n(p,q)$.

d) $V_\infty(i_1,i_2,p,q)$ *is monotone increasing in* i_1 *and* i_2.

Proof:

a) The monotonicity of K_n in n is obvious. Furthermore we get

$$K = \frac{1}{\alpha+\beta}\sum_{k=0}^{\infty}\left(\frac{\alpha}{\alpha+\beta}\right)^k = \frac{1}{\alpha+\beta}\frac{1}{1-\frac{\alpha}{\alpha+\beta}} = \frac{1}{\alpha+\beta}\frac{1}{\frac{\beta}{\alpha+\beta}} = \frac{1}{\beta}.$$

b) First, $L_n = \sum_{k=0}^{n-1}\left(\frac{\alpha}{\alpha+\beta}\right)^k K_{n-k} \le \frac{1}{\beta}\sum_{k=0}^{\infty}\left(\frac{\alpha}{\alpha+\beta}\right)^k \le \frac{1}{\beta}\frac{1}{1-\frac{\alpha}{\alpha+\beta}} = \frac{1}{\beta}\frac{\alpha+\beta}{\beta} < \infty$. Additionally $L_0 \le L_1$ and by induction $L_n = \frac{1}{\alpha+\beta}(K_n+\alpha L_{n-1}) \le \frac{1}{\alpha+\beta}(K_{n+1}+\alpha L_n) = L_{n+1}$, by the monotonicity of K_n and induction hypothesis. $L = \frac{1}{\beta^2}$ follows from the recursion formula of L_n given in the proof of theorem 5.13.

c) We know that $V_\infty(i_1,i_2,p,q) = \lim_{n\to\infty} V_n(i_1,i_2,p,q)$ holds due to corollary 4.16 for deterministic strategies which satisfy the compactness and continuity assumptions. By theorem 5.13 the there given characterization of $V_n(i_1,i_2,p,q)$ holds for each n. Taking the limit and using part a) and b) the assertion in c) follows, in particular $G(p,q) := \lim_{n\to\infty} G_n(p,q)$ exists.

d) The statement follows from c) and theorem 5.13.

□

Furthermore we know from part c) that

$$\begin{aligned} G(p,q) &= \inf_{\gamma\in A}\Bigg\{\int_0^\infty e^{-(\alpha+\beta)t}\Big\{\big(G(\phi_t^\gamma(p)+\Phi_1(\phi_t^\gamma(p)),\varphi_t^\gamma(q))-G(\phi_t^\gamma(p),\varphi_t^\gamma(q)) \\ &\qquad -c_1K\big)\mu_1(\phi_t^\gamma(p))\gamma_t(1) \\ &\qquad +\big(G(\phi_t^\gamma(p),\varphi_t^\gamma(q)+\Phi_2(\varphi_t^\gamma(q)))-G(\phi_t^\gamma(p),\varphi_t^\gamma(q)) \\ &\qquad -c_2K\big)\mu_2(\varphi_t^\gamma(q))\gamma_t(2)+G(\phi_t^\gamma(p),\varphi_t^\gamma(q))\alpha\Big\}dt\Bigg\} \\ &=: TG(p,q). \end{aligned}$$

Hence $G(p,q)$ is a fixed point of the operator T. Whereas a direct computation of $G_n(p,q)$ and $G(p,q)$ is quite hard, we can give some bounds for $G_n(p,q)$ and $G(p,q)$, respectively. Define for this purpose

$$\begin{aligned} M_{\min} &:= \max\big\{\min\{c_1\mu_1^A, c_1\mu_1^B\}, \min\{c_2\mu_2^A, c_2\mu_2^B\}\big\} \\ M^{\max} &:= \max\{c_1\mu_1^A, c_1\mu_1^B, c_2\mu_2^A, c_2\mu_2^B\} \end{aligned}$$

$M_{\min}$ stands for the second worst case and $M^{\max}$ for the best case of parameters. Then

$$G_n(p,q)\in[-M_{\min}L_{n-1}, -M^{\max}L_{n-1}] \quad \text{and} \quad G(p,q)\in[-M_{\min}L, -M^{\max}L].$$

The bounds are attained if $M_{\min}$ ($M^{\max}$) are estimated with probability 1, that means if $M^{\max}=c_1\mu_1^A$ then the a priori-probability p_0 has to be 1. In particular $G(p,q)$ is negative which is clear, since $G(p,q)$ are expected reductions of the costs due to finished services (compare the interpretation after theorem 5.2). Additionally we see that the length of the interval for $G_n(p,q)$ which is given by $(M^{\max}-M_{\min})L_{n-1}$ is increasing in n, since L_{n-1} is.

To value the information (or the cost of incomplete information) we compare in the next theorem the value functions of both models, derived in theorem 5.2, 5.13 and 5.14.

Theorem 5.15 *Denote by $V^C(i_1,i_2)$ and $V^{IC}(i_1,i_2,p,q)$ the value functions of the complete and the incomplete information models from the just mentioned theorems. Then it holds:*

a) $0\geq V_n^C(i_1,i_2)-V_n^{IC}(i_1,i_2,p,q)=-c_1\mu_1L_{n-1}-G_{n-1}(p,q)$

b) $0\geq V_\infty^C(i_1,i_2)-V_\infty^{IC}(i_1,i_2,p,q)=-c_1\mu_1L-G(p,q)$

Proof: The inequalities follows from theorem 2.12, whereas the equalities are true due to theorem 5.2, 5.13 and 5.14. □

We obtain that the value of information does not depend on the lengths of the queues i_1 and i_2. In the special case of $n=1$ we have $V_1^C(i_1,i_2)-V_1^{IC}(i_1,i_2,p,q)=0$. This is not surprising, since we have seen that for $n=1$ every service allocation is optimal and the

completion of a service of a customer has no influence on the total waiting costs since after the next jump the system terminates.

For completeness we mention here, that the proof of theorem 5.2 follows directly from theorem 5.14 c), since in the complete information case the value function does not depend on p, q and therefore the optimization problem in the definition of $G(p, q)$ is easy to solve.

Lemma 5.16

a) *$p \mapsto G_n(p, q)$ and $q \mapsto G_n(p, q)$ are concave and therefore locally Lipschitz continuous for all $n \in \mathbb{N}_0$.*

b) *$p \mapsto G(p, q)$ and $q \mapsto G(p, q)$ are concave and therefore locally Lipschitz continuous.*

Proof: Both statements are immediate consequences of theorem 3.15 and the characterization of the value function in theorem 5.13 and 5.14, where we have to adapt theorem 3.15 in a similar fashion to the finite horizon case in a). □

Remark 5.17 *We note from the separation property of the value function $V_\infty(i_1, i_2, p, q)$ that the optimal control depends for $i_1 > 0$ and $i_2 > 0$ only on p and q. Furthermore, to find an optimal control one has to solve the deterministic optimization problem $TG(p, q)$. This was already pointed out in theorem 4.9.*

So far we discussed the value function of our parallel queueing model with unknown service rates. Also important is the optimal control. For this we first remember the existence of an optimal (deterministic) policy π, see remark 5.12. Due to theorem 4.7 there exists an optimal control $u^* = (u_t^*)$. After that we look for sufficient conditions to characterize the optimal control.

From theorem 3.16 we know that our value function $J(i_1, i_2, p, q) = V_\infty(i_1, i_2, p, q)$ is a solution of the generalized HJB-equation, which is given by

$$\beta W(i_1, i_2, p, q) = \inf_{\substack{\xi_p \in \partial_p W(i_1, i_2, p, q) \\ \xi_q \in \partial_q W(i_1, i_2, p, q) \\ u \in [0,1]}} \left\{ HW(i_1, i_2, p, q, u) \right\} \tag{5.8}$$

where the generalized Hamiltonian is defined as

$$\begin{aligned} HW(i_1, i_2, p, q, u) \;:=\; & c_1 i_1 + c_2 i_2 \\ & + \xi_p(\mu_1^B - \mu_1^A) p(1-p)u + \xi_q(\mu_2^B - \mu_2^A) q(1-q)(1-u) \\ & + \left(W(i_1 + 1, i_2, p, q) - W(i_1, i_2, p, q)\right) \lambda_1 \\ & + \left(W(i_1, i_2 + 1, p, q) - W(i_1, i_2, p, q)\right) \lambda_2 \\ & + \left(W(i_1 - 1, i_2, p + \Phi_1(p), q) - W(i_1, i_2, p, q)\right) \mu_1(p) u \\ & + \left(W(i_1, i_2 - 1, p, q + \Phi_2(q)) - W(i_1, i_2, p, q)\right) \mu_2(q)(1-u). \end{aligned}$$

By theorem 4.4 it is sufficient to compute (u^*, ξ_p^*, ξ_q^*) of the generalized Hamiltonian to find an optimal control such that the generalized HJB-equation is fulfilled. We note first that $HW(i_1, i_2, p, q, u)$ is linear in u. Consequently the minimum point in u will be (if it is unique) equal to 0 or 1. Hence the optimal control will serve one queue exclusively. If the minimum point in u is not unique, that means in cases where $HW(i_1, i_2, p, q, u)$ does not depend on u or $HW(i_1, i_2, p, q, 0) = HW(i_1, i_2, p, q, 1)$ with corresponding ξ_p^* and ξ_q^* we do not choose a minimum point arbitrarily. We choose u^* in such a way, that p and q remain such, that the coefficient of u in the Hamiltonian remain 0. This will be fulfilled for example if u^* is chosen such that p and q keep constant. Let us state the existence theorem first.

Theorem 5.18 *There exists an optimal control $u^* = (u_t^*) \in \mathcal{U}$ with the above stated properties. In particular if the minimum point of the HJB-equation is unique one queue is served exclusively.*

Proof: We apply the verification theorem 4.4. If we can prove that there always exists an (u^*, ξ_p^*, ξ_q^*) such that the generalized HJB-equation is fulfilled, the statement follows. The generalized HJB-equation for the queueing model is given by

$$\beta W(i_1, i_2, p, q) = \inf_{\substack{\xi_p \in \partial_p W(i_1, i_2, p, q) \\ \xi_q \in \partial_q W(i_1, i_2, p, q) \\ u \in [0,1]}} \left\{ HW(i_1, i_2, p, q, u) \right\}$$

which is fulfilled for the value function $J(i_1, i_2, p, q) = V_\infty(i_1, i_2, p, q)$, since it is locally Lipschitz continuous and regular. Due to the linearity of $HJ(i_1, i_2, p, q, u)$ in u we conclude with

$$\begin{aligned} F(i_1, i_2, p, q) \;:=\; & c_1 i_1 + c_2 i_2 + \left(J(i_1 + 1, i_2, p, q) - J(i_1, i_2, p, q)\right) \lambda_1 \\ & + \left(J(i_1, i_2 + 1, p, q) - J(i_1, i_2, p, q)\right) \lambda_2 \end{aligned}$$

that the HJB-equation can be written as

$$\begin{aligned} & \beta J(i_1, i_2, p, q) - F(i_1, i_2, p, q) \\ = \;& \min \left\{ \inf_{\substack{\xi_p \in \partial_p J(i_1, i_2, p, q) \\ \xi_q \in \partial_q J(i_1, i_2, p, q)}} \left\{ HJ(i_1, i_2, p, q, 1) \right\}, \inf_{\substack{\xi_p \in \partial_p J(i_1, i_2, p, q) \\ \xi_q \in \partial_q J(i_1, i_2, p, q)}} \left\{ HJ(i_1, i_2, p, q, 0) \right\} \right\} \\ = \;& \min \left\{ \inf_{\xi_p \in \partial_p J(i_1, i_2, p, q)} \Big\{ \xi_p (\mu_1^B - \mu_1^A) p(1-p) \right. \\ & \qquad \left. + \left(J(i_1 - 1, i_2, p + \Phi_1(p), q) - J(i_1, i_2, p, q)\right) \mu_1(p) \Big\}, \right. \\ & \quad \inf_{\xi_q \in \partial_q J(i_1, i_2, p, q)} \Big\{ \xi_q (\mu_2^B - \mu_2^A) q(1-q) \\ & \qquad \left. \left. + \left(J(i_1, i_2 - 1, p, q + \Phi_2(q)) - J(i_1, i_2, p, q)\right) \mu_2(q) \Big\} \right\} \end{aligned}$$

$$
\begin{aligned}
= \quad \min \Big\{ & J_{0,p}(i_1,i_2,p,q;1)(\mu_1^B-\mu_1^A)p(1-p) \\
& \qquad + \left(J(i_1-1,i_2,p+\Phi_1(p),q) - J(i_1,i_2,p,q)\right)\mu_1(p), \\
& J_{0,q}(i_1,i_2,p,q;1)(\mu_2^B-\mu_2^A)q(1-q) \\
& \qquad + \left(J(i_1,i_2-1,p,q+\Phi_2(q)) - J(i_1,i_2,p,q)\right)\mu_2(q) \Big\}
\end{aligned}
$$

where we used

$$
\begin{aligned}
\inf_{\xi_p \in \partial_p J(i_1,i_2,p,q)} \left\{\xi_p(\mu_1^B-\mu_1^A)p(1-p)\right\} &= \inf_{\xi_p \in \partial_q J(i_1,i_2,p,q)} \left\{\xi_p\right\}(\mu_1^B-\mu_1^A)p(1-p) \\
&= J_{0,p}(i_1,i_2,p,q;1)(\mu_1^B-\mu_1^A)p(1-p)
\end{aligned}
$$

and the definition of the lower generalized Clarke derivative $J_{0,p}(i_1,i_2,p,q;1)$ with respect to p. Since $J(i_1,i_2,p,q)$ is regular in p we conclude that $J_{0,p}(i_1,i_2,p,q;1)$ exists and is the right hand side derivative. Similar considerations hold true for $J_{0,q}(i_1,i_2,p,q;1)$. □

5.2.4 The Symmetric Case

We consider now case C) in (5.3). This means that both possible parameters at each queue are the same, but it is not known which parameter is true at which queue. Assume furthermore $c_1 = c_2 = 1$ and for the service rates we make the convention $\mu_1^A = \mu_2^B =: \mu^A$ and $\mu_1^B = \mu_2^A =: \mu^B$. We will refer to this as the symmetric case. In particular we get $\mu^B > \mu^A$ and hence we see that if the true value $\mu_1 = \mu^A$ 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 $\mu_1 = \mu^B$ then the optimal decision is vice versa. A similar model was considered in the context of bandit problems by Donchev in his works (Donchev and Yushkevich (1996), Donchev (1998) and Donchev (1999)).

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

$$p_t := \mathbb{P}_u(\mu_1 = \mu_1^A \mid \mathcal{F}_t) = \mathbb{P}_u(\mu_2 = \mu_1^B \mid \mathcal{F}_t)$$

which is the solution of

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

where

$$
\begin{aligned}
\Phi_1(p) &= \frac{1}{\mu^A p + \mu^B(1-p)}\mu^A p - p = \frac{1}{\mu(p)}\mu^A p - p \\
\Phi_2(p) &= \frac{1}{\mu^B p + \mu^A(1-p)}\mu^B p - p = \frac{1}{\mu(1-p)}\mu^B p - p.
\end{aligned}
$$

Here we set again $\mu(p) = \mu^A p + \mu^B(1-p)$. From this stochastic differential equation we see that between two jump times τ_n and τ_{n+1}

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

The interpretation of this result is as in section 5.2.1: If we serve queue 1 majoritarian then the estimator is monotone increasing, since $\mu^A < \mu^B$ and hence jumps are more unlikely. We see that now the estimator process is not separated to one queue alone, since the hypothesis of the true values of the service rates are connected. Thus the completion of a service in both queues leads to jumps and therefore to updates of the estimator process. Determining of customers at queue 1 make μ^A being the true parameter at queue 1 more unlikely, completing a customer at queue 2 makes it more probable, since $\mu^B > \mu^A$ and hence

$$p + \Phi_1(p) = \frac{\mu^A p}{\mu(p)} \leq p \quad \text{and} \quad p + \Phi_2(p) = \frac{\mu^B p}{\mu(1-p)} \geq p \tag{5.9}$$

From theorem 5.14 we know that the value function is

$$J(i_1, i_2, p) = V_\infty(i_1, i_2, p) = (i_1 + i_2)K + (\lambda_1 + \lambda_2)L + G(p)$$

where $G(p)$ is given by

$$\begin{aligned} G(p) \;=\; \inf_{(u)} \Bigg\{ \int_0^\infty e^{-(\alpha+\beta)t} \Big\{ & \big(G(\phi^u_t(p) + \Phi_1(\phi^u_t(p))) - G(\phi^u_t(p)) - K\big)\mu(\phi^u_t(p))u_t(1) \\ & + \big(G(\phi^u_t(p) + \Phi_2(\phi^u_t(p))) - G(\phi^u_t(p)) - K\big)\mu(1-\phi^u_t(p))(1-u_t(1)) \\ & + G(\phi^u_t(p))\alpha \Big\} dt \Bigg\}. \end{aligned}$$

Since $p \mapsto G(p)$ is concave it is locally Lipschitz continuous and regular. Because we are in the symmetric case with the same cost rate at each queue it is clear that

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

The function $p \mapsto G(p)$ is symmetric and concave, hence it is 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 $p < \frac{1}{2}$ and $\xi \leq 0$ for $p > \frac{1}{2}$ and since $G(p)$ attains a maximum in $p = \frac{1}{2}$ that $\{0\} \in \partial_p G(\frac{1}{2})$. Additionally we have

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

where we understand $-F$ as the set $\{-a \mid a \in F\}$. Especially we get, $\partial_p G(\frac{1}{2})$ is a symmetric interval with respect to 0. Since we are in the symmetric case it would be desirable that the symmetry also holds for the optimal control. This is the case as the following theorem shows (as long as $i_1 > 0, i_2 > 0$).

Theorem 5.19 *Let $i_1 > 0$ and $i_2 > 0$ and denote by $u^* = (u_t^*)$ the optimal control, then $u^*(p) = 1 - u^*(1-p)$.*

Proof: From theorem 4.3 we know that the value function $J(i_1, i_2, p)$ and the optimal control (u_t^*) with corresponding state process (p_t^*) fulfils the generalized HJB-equation for almost all $t \geq 0$. That means by the separation property of $J(i_1, i_2, p)$ in theorem 5.14

$$\inf_{\xi \in \partial_p G(p_t^*)} \Big\{ \big(G(p_t^* + \Phi_1(p_t^*)) - G(p_t^*) - K\big)\mu(p_t^*)u_t^* \\ + \big(G(p_t^* + \Phi_2(p_t^*)) - G(p_t^*) - K\big)\mu(1-p_t^*)(1-u_t^*) \\ + \xi(\mu^B - \mu^A)p_t^*(1-p_t^*)(2u_t^* - 1) - \beta G(p_t^*) \Big\} = 0.$$

Consider the generalized Hamiltonian in this symmetric case given by

$$\begin{aligned} HG(p,u) &:= \big(G(p + \Phi_1(p)) - G(p) - K\big)\mu(p)u \\ &\quad + \big(G(p + \Phi_2(p)) - G(p) - K\big)\mu(1-p)(1-u) \\ &\quad + \xi(p) \cdot (\mu^B - \mu^A)p(1-p)(2u-1) - \beta G(p) \\ &=: M(p,G)u + R(p,G) \end{aligned}$$

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

$$\begin{aligned} HG(1-p,u) &= \big(G(1-p + \Phi_1(1-p)) - G(1-p) - K\big)\mu(1-p)u \\ &\quad + \big(G(1-p + \Phi_2(1-p)) - G(1-p) - K\big)\mu(p)(1-u) \\ &\quad + \xi(1-p) \cdot (\mu^B - \mu^A)(1-p)p(2u-1) - \beta G(1-p) \\ &= \big(G(p + \Phi_2(p)) - G(p) - K\big)\mu(1-p)u \\ &\quad + \big(G(p + \Phi_1(p)) - G(p) - K\big)\mu(p)(1-u) \\ &\quad - \xi(p) \cdot (\mu^B - \mu^A)p(1-p)(2u-1) - \beta G(p) \\ &=: -M(p,G)u + \tilde{R}(p,G). \end{aligned}$$

By the linearity of $HG(p,u)$ in u and the symmetry of the coefficients $M(p,G)$ of u in $HG(p,u)$ and $HG(1-p,u)$ it follows, that if $u^*(p) = 1$ is a minimum point of $u \mapsto HG(p,u)$, then $u^*(1-p) = 0$ is a minimum point of $u \mapsto HG(1-p,u)$. The same conclusion holds true if $u^*(p) = 0$. If $p = \frac{1}{2}$ we will see in the upcoming that 0 and 1 are minimum points in the generalized HJB-equation. 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. This means if $p < \frac{1}{2}$ it is more likely that the better rate μ^B is the true value at queue 1. Thus it is optimal to serve queue 1. If both hypothesis are equiprobable, the optimal control divides service fifty-fifty between both queues (although every allocation would be optimal). This was explained from a technical point of view before the existence theorem 5.18 to keep the estimator process constant. It is motivated from an intuitive point of view in the proof of the next theorem which proves the optimality of a threshold-strategy with threshold $p^* = \frac{1}{2}$. In other words the certainty equivalence principle to the $c\mu$-rule holds true, since with $c_1 = c_2 = 1$

$$\mu_1(p) = \mu(p) \geq \mu(1-p) = \mu_2(p) \quad \Longleftrightarrow \quad p \leq \frac{1}{2}.$$

Theorem 5.20 *Assume $i_1 > 0$ and $i_2 > 0$ and set*

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

then $\big(u^*(p^*_{t-})\big)$ is optimal (as long as customers are waiting in the queue which is served). In particular the optimal control u^* is a threshold control with threshold $p^* = \frac{1}{2}$.

Proof: Choose $\xi^* = G_{0,p}(p;1)$ for $p \leq \frac{1}{2}$, since we know from the proof of the existence theorem 5.18 that according to the generalized verification theorem 4.4 it is sufficient to compute minimum points in u of the right hand side of the generalized HJB-equation. Due to the separation property of the value function $J(i_1, i_2, p)$ we only have to compute the minimum point in u of the Hamiltonian, which is independent of i_1, i_2 (as long as $i_1 > 0, i_2 > 0$), which is given by

$$\begin{aligned} HG(p,u) \;=\; & \Bigg\{ \Big(G\left(\frac{\mu^A p}{\mu(p)}\right) - G(p) - K\Big)\mu(p) - \Big(G\left(\frac{\mu^B p}{\mu(1-p)}\right) - G(p) - K\Big)\mu(1-p) \\ & \qquad +2\xi^*(\mu^B - \mu^A)p(1-p)\Bigg\} u \\ & \qquad -\xi^*(\mu^B - \mu^A)p(1-p) + \Big(G\left(\frac{\mu^B p}{\mu(1-p)}\right) - G(p) - K\Big)\mu(1-p). \end{aligned}$$

It is sufficient to prove that

$$\begin{aligned} M(p) \;:=\; & \left(G\left(\frac{\mu^A p}{\mu(p)}\right) - G(p) - K\right)\mu(p) - \left(G\left(\frac{\mu^B p}{\mu(1-p)}\right) - G(p) - K\right)\mu(1-p) \\ & \qquad +2\xi^*(\mu^B - \mu^A)p(1-p) < 0 \end{aligned}$$

for $p < \frac{1}{2}$ and equal to 0 for $p = \frac{1}{2}$ by theorem 5.19. For this purpose define

$$h(p) := 2\xi^*(\mu^B - \mu^A)p(1-p) - K\mu(p) + K\mu(1-p).$$

Then

$$M(p) = \left\{ \big(G\left(\frac{\mu^A p}{\mu(p)}\right) - G(p)\big)\mu(p) - \big(G\left(\frac{\mu^B p}{\mu(1-p)}\right) - G(p)\big)\mu(1-p) + h(p) \right\} u.$$

Analyzing $h(p)$ we get, that $h(p) < 0$ for $p < \frac{1}{2}$, which can be seen as follows:

- $p \mapsto h(p)$ is continuous
- $h(0) = -K(\mu^B - \mu^A) < 0$
- $h\left(\frac{1}{2}\right) = (\mu^B - \mu^A)(\xi^*\left(\frac{1}{2}\right) - \frac{1}{2}\xi^*\left(\frac{1}{2}\right)) \leq 0$ since $\xi^*\left(\frac{1}{2}\right) = G_{0,p}\left(\frac{1}{2}; 1\right) \in \partial_p G\left(\frac{1}{2}\right)$ is ≤ 0 due the fact, that $\partial_p G\left(\frac{1}{2}\right)$ is symmetric interval around 0 and $G_{0,p}\left(\frac{1}{2}; 1\right)$ is equal to the left bound
- the maximum point of $h(p)$ is $p^* = \frac{K+\xi^*}{2\xi^*} > \frac{1}{2}$, since $K > 0$ and $\xi^* \geq 0$ for all $p \in [0, \frac{1}{2})$.

It remains to prove that

$$\big(G\left(\frac{\mu^A p}{\mu(p)}\right) - G(p)\big)\mu(p) - \big(G\left(\frac{\mu^B p}{\mu(1-p)}\right) - G(p)\big)\mu(1-p) \leq 0$$

for $p < \frac{1}{2}$ (note for $p = \frac{1}{2}$ this expression is equal 0). 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 $0 \leq p_1 \leq p \leq p_2 \leq 1$ such that $p_2 \neq p_1$. Choose now $p_1 = \frac{\mu^A p}{\mu(p)}$ and $p_2 = \frac{\mu^B p}{\mu(1-p)}$ and conclude

$$\begin{aligned} & (p_2 - p_1)G(p)\mu(p)\mu(1-p) = (\mu^B p\mu(p) - \mu^A p\mu(1-p))G(p) \\ \geq\ & (\mu(p)p - \mu^A p)\mu(1-p)G\left(\frac{\mu^B p}{\mu(1-p)}\right) + (\mu^B p - \mu(1-p)p)\mu(p)G\left(\frac{\mu^A p}{\mu(p)}\right). \end{aligned}$$

Dividing by p we continue with

$$\begin{aligned} 0 \ \geq\ & \big(G\left(\frac{\mu^A p}{\mu(p)}\right) - G(p)\big)\mu^B\mu(p) - G\left(\frac{\mu^A p}{\mu(p)}\right)\mu(p)\mu(1-p) \\ & -\big(G\left(\frac{\mu^B p}{\mu(1-p)}\right) - G(p)\big)\mu^A\mu(1-p) + G\left(\frac{\mu^B p}{\mu(1-p)}\right)\mu(p)\mu(1-p) \\ =\ & \Big\{\big(G\left(\frac{\mu^A p}{\mu(p)}\right) - G(p)\big)\mu(p) - \big(G\left(\frac{\mu^B p}{\mu(1-p)}\right) - G(p)\big)\mu(1-p)\Big\}(\mu^B - \mu^A)(1-p) \end{aligned}$$

and since $(\mu^B - \mu^A)(1-p)$ we conclude that $M(p) < 0$ for $p < \frac{1}{2}$.

If $p = \frac{1}{2}$ then $(u, \xi) = (1, G_p^0(\frac{1}{2}; 1))$ and $(u, \xi) = (0, G_{0,p}(\frac{1}{2}; 1))$ as well fulfil the HJB-equation. Hence both allocations would be optimal. We choose $u^*\left(\frac{1}{2}\right) = \frac{1}{2}$ as linear combination of 0 and 1, which is optimal with corresponding $\xi^* = \frac{1}{2}\left(G_{0,p}(\frac{1}{2}; 1) + G_p^0(\frac{1}{2}; 1)\right) = 0 \in \partial_p G(\frac{1}{2})$, since $\partial_p G(\frac{1}{2})$ is symmetric. Although the choice of ξ^* is not necessary, since its coefficient is equal to 0 for $u^* = \frac{1}{2}$. Additionally the estimator $\phi^*(p)$ remains constant $\frac{1}{2}$ for $u^* = \frac{1}{2}$ up to the next jump. Thus the service rate is split up fifty-fifty and the next served customer determines which queue is served next exclusively, since directly after a jump the estimator p_t is not equal $\frac{1}{2}$ anymore. □

Remark 5.21 *From the proof we conclude the stay-on-a-winner property: If the server finishes the service of a customer in queue i, then it will continue serving queue i (assuming queue i is not empty). This is true since the completion of the service of a customer in queue i makes queue i more probable to be the "better" queue, see (5.9).*

The following table illustrates the results above in a numerical context. We simulate a parallel queueing model with unknown service rates which are symmetric. We consider different strategies and compare them to each other.

strategy	simulated cost
$c\mu$ rule (complete information)	8.4795
prefer queue 2	8.4975
prefer queue 1	8.7538
allocate constant 50 : 50	8.6410
uniform distribution on $\{0, 1\}$	8.6061
control limit with $p = \frac{1}{2}$	8.5224

Table 1: Symmetric case with $\mu^A = 0.1, \mu^B = 0.3, \lambda_1 = \lambda_2 = 0, i_1 = 3, i_2 = 5, p_0 = 0.9$, $\beta = 0.9$ and true parameter $\mu_1 = \mu^A$.

Donchev proved in his symmetric bandit models, considered in Donchev and Yushkevich (1996), Donchev (1998) and Donchev (1999) with the help of Presman and Sonin (1990) and Presman (1990), that the control limit strategy with control limit $p = \frac{1}{2}$ is optimal. They use a logarithmic scale in their proof which works due to their special structure of their cost rate. In our model their approach seems not to work.

5.2.5 Complete Information about One Service Rate

Consider now case B) in (5.3). Hence the parameter μ_2 is known and the three inequalities can be rewritten as

$$c_1\mu_1^B > c_2\mu_2 > c_1\mu_1^A. \tag{5.10}$$

Since one parameter is completely observable the estimator process q_t is dispensable and p_t is the solution of

$$dp_t = u_t(1)\big\{(\mu_1^B - \mu_1^A)p_t(1-p_t)\big\}\mathbb{1}\big(X_t^1 > 0\big)dt + \Phi(p_{t-})dN_t^X(X_{t-}^1, X_{t-}^1 - 1) \tag{5.11}$$

where $\Phi(p) = \frac{1}{\mu_1(p)}\mu_1^A p$, see also section 5.2.1. The results of lemma 5.8 and 5.9 can be carried through to this section. Thus $\phi_t^u(p)$ is

- monotone increasing in t and p
- Lipschitz continuous in t and p.

Again the separation property of the value function $J(i_1, i_2, p) = V_\infty(i_1, i_2, p)$ holds true. Hence the generalized HJB-equation given in (5.8) simplifies to

$$\begin{aligned} &\beta G(p) \qquad\qquad (5.12)\\ = &\inf_{\substack{\xi \in \partial_p G(p) \\ u \in U}} \left\{ \xi(\mu_1^B - \mu_1^A)p(1-p)u + (G(p + \Phi(p)) - G(p) - c_1 K)\mu_1(p)u - c_2\mu_2 K(1-u) \right\} \end{aligned}$$

We prove next, that in this model there even exists a pure optimal strategy. That means one queue is always served exclusively. This result is very special to case of one unknown service rate (here μ_1). It will not hold true for the case of two unknown service rates in general as pointed out in the last sections.

Theorem 5.22 *There exists an optimal pure control $u^* = (u_t^*)$, in particular $u_t^* \in \{0, 1\}$.*

Proof: The existence of an optimal deterministic control u^* is a consequence of the existence theorem 5.18. There we have seen, that the optimal control serves one queue exclusively, except the minimum point of the Hamiltonian is not unique. In this case, it is sufficient to choose the minimum point such that the estimator process remains constant (see the comments after theorem 5.18). This is here the case, if we choose $u^* = 0$. The existence of $\xi^* = G_{0,p}(p; 1)$ is guaranteed as in theorem 5.18 by the regularity of the value function in p. Summarizing we have found a pure control with the verification theorem 4.4. □

From the proof we conclude that if the optimal server switches from a non-empty queue 1 to queue 2 it will remain there until queue 2 is empty, since the completion of a service in queue 2 has no influence to the estimator process p. From the HJB-equation (5.12) and the last proof it is clear, that the optimal control is a control limit strategy. It is defined as $u^* = \big(u^*(X_{t-}^*, p_{t-}^*)\big)$ with

$$u^*(i_1, i_2, p) = \begin{cases} 1 & i_2 = 0, i_1 > 0 \\ 2 & i_1 = 0, i_2 > 0 \\ 1 & i_1 > 0, i_2 > 0, H(p) < c_2\mu_2 K \\ 2 & i_1 > 0, i_2 > 0, H(p) \geq c_2\mu_2 K \\ \text{arbitrarily} & i_1 = i_2 = 0 \end{cases}$$

where the control limit is given by

$$H(p) := (G(p + \Phi(p)) - G(p) - c_1 K)\mu_1(p) + G_{0,p}(p; 1)(\mu_1^B - \mu_1^A)p(1 - p).$$

Unfortunately, this control limit is not suitable since it depends on the function $G(p)$ which is quite hard to compute. But on the other hand it is clear that there exists a critical value τ^* defined by the first time point where the control limit $H(p_t^*)$ is greater or equal to $c_2\mu_2 K$, such that it is optimal to serve queue 1 for $t < \tau^*$ and queue 2 for $t \geq \tau^*$ as long as customers are waiting in both queues. This is the well-known stay-on-a-winner property.

The next theorem states, that if the estimate $c_1\mu_1(p)$ is greater than $c_2\mu_2$, then it is always optimal to serve queue 1.

Theorem 5.23 *If $c_1\mu_1(p) \geq c_2\mu_2$, then it is optimal to serve queue 1.*

Proof: The statement is an immediate consequence of the properties of the optimal control in the complete information case, discussed in theorem 5.1. □

This is some kind of one-step-look-ahead-rule: a myopic strategy, as in the case of complete information. Due to the monotonicity of $p \mapsto \mu_1(p)$ we are able to characterize the optimality condition of the last theorem in dependence on p. For this define

$$p^{\leq} := \sup\{p \in [0, 1] \mid c_1\mu_1(p) \geq c_2\mu_2\}$$

and if $p \leq p^{\leq}$ serving queue 1 (means $u^*(p) = 1$) is optimal. Note that $p^{\leq}$ is well-defined, since for $p = 0$ we get $c_1\mu_1(0) = c_1\mu_1^B > c_2\mu_2$ by (5.10). $p^{\leq}$ is independent of the length of the queues i_1, i_2 and can be computed explicitly as

$$p^{\leq} = \frac{c_1\mu_1^B - c_2\mu_2}{c_1(\mu_1^B - \mu_1^A)}.$$

From this equation we see that $p^{\leq}$ is monotone increasing in μ_1^B and decreasing in μ_1^A and μ_2. This makes sense, since the higher μ_2 the lower p has to be such that $c_1\mu_1(p) \geq c_2\mu_2$ is satisfied. Similar interpretations hold for the other parameters.

Notice, that $c_2\mu_2 > c_1\mu_1(p)$ does not imply the optimality of serving queue 2 as we will see in the following. If no new customers arrive at both queues we are able to give a sufficient condition for an optimal pure control in the spirit of the $c\mu$-rule as in the complete information case, see theorem 5.1. Hence assume

$$\lambda_1 = \lambda_2 = 0.$$

By this assumption we immediately get a finite state space S_X, that means

$$S_X = \{0, \ldots, i_1(0)\} \times \{0, \ldots, i_2(0)\},$$

where $i_j(0)$ is the number of customers waiting at time 0 in queue j. We first mention, that as a consequence of the finite queue length, the cost function in the MDP given by $r(i_1, i_2) = \frac{c_1 i_1 + c_2 i_2}{\alpha+\beta}$ is bounded. Since $\delta < 1$ we can apply Banach's fixed point theorem due to the boundedness of $V_\infty(i_1, i_2, p)$ and we obtain without any continuity and compactness assumptions

$$\lim_{n\to\infty} V_n(i_1, i_2, p) = V_\infty(i_1, i_2, p),$$

V_∞ is the unique fixed point of $\mathcal{T}$ and for every bounded V_0

$$||V_\infty - \mathcal{T}^n V_0|| \leq \frac{\delta^n}{1-\delta}||\mathcal{T}V_0 - V_0||.$$

Notice that Howard's policy improvement algorithm is also applicable now.

We denote the state where both queues are empty by $\mathcal{G}_0 := (0, 0)$. $\mathcal{G}_0$ is an absorbing set, since if both queues are empty the optimization problem is terminated and the state process (X_t, p_t) will never leave $\mathcal{G}_0 \times [0, 1]$, since there are no arrivals to the queues. It is clear, that it is never optimal to serve an empty queue (see remark 5.11). Therefore we restrict the set of admissible actions for $i \notin \mathcal{G}_0$ to

$$D(i_1, i_2, p) = \begin{cases} A & i_1 > 0, i_2 > 0 \\ \{\gamma \in A \mid \gamma_t \equiv 1\} & i_1 > 0, i_2 = 0 \\ \{\gamma \in A \mid \gamma_t \equiv 2\} & i_1 = 0, i_2 > 0 \end{cases}$$

As a consequence we know, that under each admissible strategy the set $\mathcal{G}_0$ is reached almost surely, if all possible service rates are strictly positive.

Lemma 5.24 *Assume $\mu_1^A > 0$. Then for all (i_1, i_2, p) and all policies $\pi \in F^\infty$ there exists a random variable $\tau := \tau(i_1, i_2, p)$ with $\mathbb{P}_\pi(\tau < \infty) = 1$ such that*

$$\mathbb{P}_\pi\Big((X_\tau, p_\tau) \in \mathcal{G}_0 \times [0, 1] \mid (X_0, p_0) = (i_1, i_2, p)\Big) = 1$$

and hence the MDP is terminating.

Proof: Avoiding to serve an empty queue, one only looks for the time point where i_1 customers in queue 1 and i_2 in queue 2 are served. The service behaviour is described by Poisson processes and we know, that i_1 and i_2 jumps happen in finite time almost surely as long as the intensities are positive, see for example Brémaud (1981). □

Since there are no arrivals we are able to use a very special solution technique, called recursion in the state space, which proceeds as follows. Define $\mathcal{G}_1 := \{(1, 0), (0, 1)\}, \mathcal{G}_2 := \{(2, 0), (1, 1), (0, 2)\}$ and so on. Then we gain a disjoint partition of $S_X = \{0, \ldots, i_1(0)\} \times \{0, \ldots, i_2(0)\} = \sum \mathcal{G}_k$. Define $N := N(i_1, i_2, p)$ as the number of jumps of p_t until $X_\tau \in \mathcal{G}_0$ and $V(i_1, i_2, p) := V_N(i_1, i_2, p)$. By the terminating property we know

$$V(i_1, i_2, p) = 0 \quad \forall (i_1, i_2) \in \mathcal{G}_0.$$

Then compute for $(i_1, i_2) \in \mathcal{G}_k$, $k = 1, 2, \ldots$, the value function $V(i_1, i_2, p)$ with the help of the Bellman equation, which is possible since for the computation of $V(i_1, i_2, p)$ for $(i_1, i_2) \in \mathcal{G}_k$ only the knowledge of $V(j_1, j_2, p)$ for $(j_1, j_2) \in \mathcal{G}_\kappa$, $\kappa = 0, \ldots, k-1$ is necessary. We will demonstrate this in the proof of theorem 5.25.

The next theorem gives a sufficient condition for an optimal control. It is the counterpart of theorem 5.23, where we gave a condition, such that serving queue 1 is optimal. Now: if the highest estimated value of $c_1\mu_1$ is less than $c_2\mu_2$ (remember the monotonicity result in lemma 5.10), then the optimal strategy is to serve queue 2 until it is empty. Hence the stay-on-a-winner property is obtained again. The idea behind this result is, that the estimate for $c_1\mu_1$ remains less than $c_2\mu_2$ all the time. Consequently there is no expected benefit to accept higher cost now for new informations in the hope of lower future cost. Introduce the operator $\mathcal{M}$ defined by

$$\mathcal{M}p := p + \Phi(p).$$

Theorem 5.25 *If $i_1 > 0, i_2 > 0$ and*

$$c_2\mu_2 > c_1\mu_1(\mathcal{M}^{i_1-1}p) \tag{5.13}$$

then it is optimal to serve queue 2 until it is empty.

Before we prove this theorem we discuss the preconditions of the theorem in more detail and receive some additional results for the proof.

Lemma 5.26 *The set of all p which fulfils (5.13) is an interval, that means*

$$\left\{p \in [0,1] \mid c_2\mu_2 > c_1\mu_1(\mathcal{M}^{i_1-1}p)\right\} = (p^>(i_1), 1],$$

where $p^>(i_1) := \inf\left\{p \in [0,1] \mid c_2\mu_2 > c_1\mu_1(\mathcal{M}^{i_1-1}p)\right\}$. $p^>(i_1)$ is well-defined since for $p = 1$ we have $c_1\mu_1(1) = c_1\mu_1^A < c_2\mu_2$ by assumption (5.10).

Proof: By lemma 5.10 we know that $p \mapsto \mathcal{M}p$ is monotone increasing and by assumption (5.10) we have $\mu_1^A < \mu_1^B$. Let $p > p^>(i_1) =: p^>$ then

$$\mathcal{M}p \geq \mathcal{M}p^> \quad \Rightarrow \quad \mathcal{M}^{i_1-1}p \geq \mathcal{M}^{i_1-1}p^> \quad \Rightarrow \quad c_1\mu_1(\mathcal{M}^{i_1-1}p) \leq c_1\mu_1(\mathcal{M}^{i_1-1}p^>) < c_2\mu_2$$

where the last inequality holds by the definition of $p^>(i_1)$. In particular $p^>(i_1)$ is well-defined and depends only on i_1 and not on i_2. □

Observe that $p^>(i_1)$ can be computed explicitly for every i_1 as for example

$$p^>(2) = \frac{\frac{\mu_1^B(\mu_2-\mu_1^B)}{\mu_1^A(\mu_1^A-\mu_1^B)}}{1-\frac{\mu_2-\mu_1^B}{\mu_1^A}}.$$

As an immediate consequence of the definition of $p^{>}(i_1)$ one can show:

$$\begin{aligned} p^{>}(i_1) &= \inf\left\{p \in [0,1] \mid c_2\mu_2 > c_1\mu_1(\mathcal{M}^{i_1-1}p)\right\} \\ &= \inf\left\{p + \Phi(p) \in [0,1] \mid c_2\mu_2 > c_1\mu_1(p + \Phi(p) + \mathcal{M}^{i_1-2}(p + \Phi(p)))\right\} \\ &= (p + \Phi(p))^{>}(i_1 - 1). \end{aligned}$$

Additionally we get the monotonicity of $i_1 \mapsto p^{>}(i_1)$. That means the more customers are waiting in queue 1, the higher $p^{>}(i_1)$, in particular the higher the threshold for the estimate of the (bad) service rate μ_1^A for serving queue 2.

Lemma 5.27 *$i_1 \mapsto p^{>}(i_1)$ is monotone increasing.*

Proof: By lemma 5.10 we know that $\Phi(p) \leq 0$, thus $\mu_1(\Phi(p)) \geq 0$ by (5.10). Hence

$$\begin{aligned} p^{>}(i_1) &= \inf\left\{p \in [0,1] \mid c_2\mu_2 > c_1\mu_1(\mathcal{M}^{i_1-1}p)\right\} \\ &\geq \inf\left\{p \in [0,1] \mid c_2\mu_2 > c_1\mu_1(\mathcal{M}^{i_1-2}p)\right\} = p^{>}(i_1 - 1). \end{aligned}$$

□

Important for the proof of the optimality in theorem 5.25 is the following lemma, which guarantees that the recursion in the state space is possible. That means, if condition (5.13) is satisfied in state (i_1, p), then it is fulfilled if one customer is served at queue 1 and the estimator has been updated from p to $p+\Phi(p)$, in particular it is fulfilled in $(i_1-1, p+\Phi(p))$.

Lemma 5.28 *If $p > p^{>}(i_1)$ then $\mathcal{M}\phi_t^{\gamma}(p) > p^{>}(i_1 - 1)$ for all $i_1 \geq 2$ and for all $t \geq 0$.*

Proof: By lemma 5.8 we know $\phi_t^{\gamma}(p) > p$ and we conclude for $i_1 = 2$

$$c_1\mu_1(\phi_t^{\gamma}(p)) \leq c_1\mu_1(p) \leq c_1\mu_1(\mathcal{M}p) < c_2\mu_2$$

where the second inequality holds due to lemma 5.10. Assume the statement for $i_2 \geq 2$ is true, then with lemma 5.8 and 5.10

$$c_1\mu_1(\mathcal{M}\phi_t^{\gamma}(p)) \leq c_1\mu_1(\mathcal{M}p) \leq c_1\mu_1(\mathcal{M}(p + \Phi(p))) < c_2\mu_2,$$

where the last inequality follows as in the case $i_2 = 2$. □

We are now in the position to give the proof of theorem 5.25. As in the complete information case in theorem 5.1 we use an interchange argument and additionally the recursion in the state space.

Proof of theorem 5.25:
The details of the proof can be found in appendix B. We illustrate here the main steps. Since the existence of an optimal pure control is guaranteed by theorem 5.22 define a decision rule $f = (f_t) = (f(X_t^1, X_t^2, p_t))$ where

$$f(i_1, i_2, p) := \begin{cases} 2 & i_1 > 0, i_2 > 0 \\ 1 & i_2 = 0 \\ 2 & i_1 = 0 \end{cases}$$

Denote the slightly modified decision rule g by

$$g(i_1,i_2,p) := \begin{cases} 1 & i_1>0, i_2>0, t\in B \\ 2 & i_1>0, i_2>0, t\notin B \\ 1 & i_2=0 \\ 2 & i_1=0 \end{cases}$$

where $B\subset[0,\infty)$. Without loss of generality assume $B=[0,\varepsilon]$. Consider then two policy

$$\pi=(f,g,f,\ldots,f)\in F^n \quad \text{and} \quad \tilde{\pi}=(g,f,\ldots,f)\in F^n$$

and compute

$$V_{n,\tilde{\pi}}(i_1,i_2,p)-V_{n,\pi}(i_1,i_2,p)=\int_0^{\varepsilon} e^{-(\alpha+\beta)t}\left(\frac{c_2\mu_2-c_1\mu_1(\phi_t^g(p))}{\alpha+\beta}\right)dt\geq 0. \tag{5.14}$$

This inequality is true, since $c_1\mu_1(\phi_t^g(p))\leq c_1\mu_1(\mathcal{M}^{i_1-1}p)<c_2\mu_2$. For this statement we have to use $\phi_t^g(p)\geq p\geq \mathcal{M}p$ (see lemma 5.10) and $\mu_1^A<\mu_1^B$ by (5.10). Finally we have to show that $(f_t)_{t\geq 0}$ is a minimizer of $V_n(i_1,i_2,p)$ for $p>p^{>}(i_1)$ and all $n\in\mathbb{N}$. Consider

$$\begin{aligned} \mathcal{T}_g V_n(i_1,i_2,p) &= \mathcal{T}_g V_{n,(f,\ldots,f)}(i_1,i_2,p)=V_{n+1,(g,f,\ldots,f)}(i_1,i_2,p) \\ &\overset{(5.14)}{\geq} V_{n+1,(f,g,f,\ldots,f)}(i_1,i_2,p)=\mathcal{T}_f V_n(i_1,i_2,p) \end{aligned}$$

and hence $f=(f_t)$ is a minimizer of $V_n(i_1,i_2,p)$. In the detailed computations we used

$$V_n(i_1,i_2,\phi_s^g(p))=V_{n,(f,\ldots,f)}(i_1,i_2,\phi_s^g(p))$$

which is true by induction hypotheses, since $\phi_s^g(p)\geq p$ and

$$\begin{aligned} V_n(i_1-1,i_2,\phi_s^g(p)+\Phi(\phi_s^g(p))) &= V_{n,(f,\ldots,f)}(i_1-1,i_2,\phi_s^g(p)+\Phi(\phi_s^g(p))) \\ V_n(i_1,i_2-1,\phi_s^g(p)) &= V_{n,(f,\ldots,f)}(i_1,i_2-1,\phi_s^g(p)) \end{aligned}$$

by the recursion in the state space, since the preconditions of this theorem are fulfilled for n in states $(i_1-1,i_2,\phi_s^g(p)+\Phi(\phi_s^g(p)))$ and $(i_1,i_2-1,\phi_s^g(p))$ by lemma 5.28. □

In the special case of $(1,i_2)$ the condition of theorem 5.25 simplifies to

$$c_2\mu_2>c_1\mu_1(p)$$

which is the certainty equivalence principle of the $c\mu$-rule, where the unknown parameter μ_1 is replaced by its estimator $\mu_1(p)$. In general the preconditions of the theorem guarantees that the highest estimated value for μ_1 is such that $c_1\mu_1(p)$ remains less than $c_2\mu_2$, which is essential for the proof of the optimality of the $c\mu$-rule. That means looking for the optimal decision knowledge of the conditional probability p is not necessary anymore.

With theorem 5.23 in mind we have found an optimal (pure) control for the model without arrivals in the intervals $[0,p^{\leq}]$ and $(p^{>}(i_1),1]$ where $p^{\leq}=p^{>}(1)$. The interval $(p^{>}(i_1),1]$ is getting smaller for increasing i_1, see lemma 5.27. This fact is illustrated in figures 4 and 5. Remember that outside these both intervals the optimal control serves also one queue exclusively as proven in theorem 5.22.

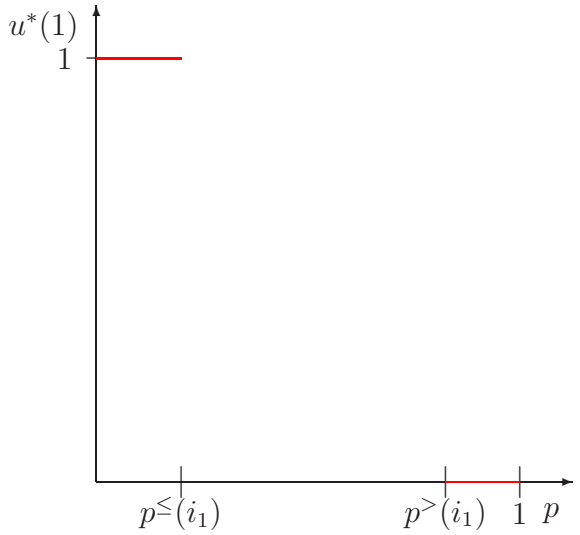


Figure 4: Optimal Control in a Waiting-Cost Model without Arrivals for fixed i_1

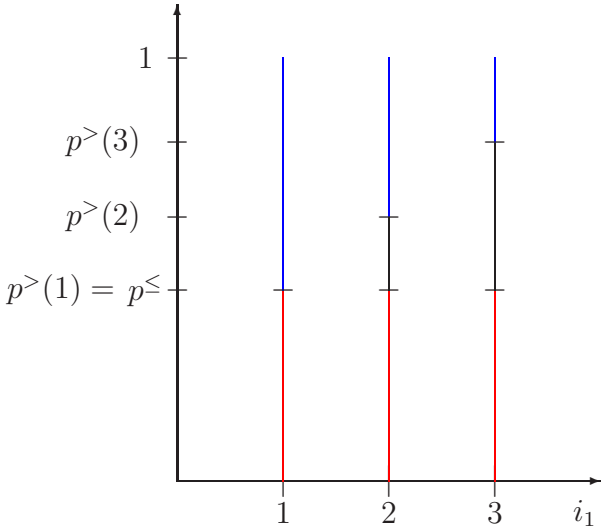


Figure 5: Optimal Control in a Waiting-Cost Model without Arrivals $u_t(1) = 1$, $u_t(1) = 0$

Knowing that the optimal strategy is a control limit strategy we propose and discuss some reasonable control limit strategies in the next table. Notice that the real control limit is not suitable, since it depends on $G(p)$ as seen on page 86, which is quite hard to compute. Unfortunately our numerical investigations do not uniquely indicate a "best" strategy to us, but the certainty equivalence control performs well mostly (see table 2).

strategy	simulated cost
$c\mu$ rule (complete information)	8.50442
split service constant 50 : 50	8.6506
allocate according to an uniform distribution on $\{0, 1\}$	8.6244
control limit with mean of $p^{>}(i_1)$ and $p^{\leq}$	8.6308
certainty equivalence	8.6139

Table 2: One Service Rate known with $c_1 = c_2 = 1$, $\mu_1^A = 0.1, \mu_1^B = 0.3, \mu_2 = 0.2$, $\lambda_1 = \lambda_2 = 0, i_1 = 3, i_2 = 5, p_0 = 0.9, \beta = 0.9$ and true Parameter $\mu_1 = \mu^A$

5.2.6 The Optimal Control in a Model with Reward-Function

We consider now a parallel queueing setup as in section 5.1, but with different cost structure. Instead of waiting costs for each customer in queue, we earn for each served customer of queue i a positive reward $r_i \geq 0$. Hence the reward criterion is

$$\mathbb{E}_u \left[\int_0^\infty e^{-\beta t} \left(r_1 dN_t^{X_1}(X_{t-}^1, X_{t-}^1 - 1) + r_2 dN_t^{X_2}(X_{t-}^2, X_{t-}^2 - 1) \right) \right] \to \max .$$

$N_t^{X_j}(i_j, i_j - 1)$ denotes the departure process of queue j, modelled by a Poisson process with intensity

$$\mu_j \mathbb{1}\big(u = j\big)\mathbb{1}\big(i_j > 0\big),$$

where i_j is the number of waiting customers in queue j. Using the definition of the intensities the objective function can be written as

$$\mathbb{E}_u \left[\int_0^\infty e^{-\beta t} \left(r_1 \mu_1 u_t(1) \mathbb{1}\big(X_t^1 > 0\big) + r_2 \mu_2 u_t(2) \mathbb{1}\big(X_t^2 > 0\big) \right) dt \right].$$

Here the reward function is bounded. Hence we do not require the continuity and compactness conditions for the convergence of V_n to V_∞ due to Banach's fixed point theorem.

As in remark 5.11 it is never advantageously to serve an empty queue. One can prove that the $r\mu$-rule is optimal in the complete information case, that means

$$r_1\mu_1 \geq r_2\mu_2 \quad \Longrightarrow \quad u_t^* := (u^*(X_{t-}^*)) \text{ with } u^*(i_1, i_2) = \begin{cases} 1 & i_1 > 0 \\ 2 & i_1 = 0 \end{cases} \text{ is optimal.}$$

The proof of this optimality statement works completely similar as in theorem 5.1, where we get in the spirit of (5.1) the following inequality

$$V_{N+1,(g,f^*,\ldots,f^*)} - V_{N+1,(f^*,g,f^*,\ldots,f^*)} = \frac{1}{\alpha+\beta} \left(1 - \frac{\alpha}{\alpha+\beta} \right) (r_2\mu_2 - r_1\mu_1) \leq 0.$$

Hence we see that this myopic control in the case of complete information is optimal.

As in the last sections we assume, that $\mu_1 \in \{\mu_1^A, \mu_1^B\}$ is not observable and similar to (5.10)

$$r_1\mu_1^B > r_2\mu_2 > r_1\mu_1^A. \tag{5.15}$$

One can derive the conditional probabilities for μ_1 as in section 5.2.1 and define the reduced MDP as in section 5.2.2 with the slight modification in the reward function given by

$$r(i_1, i_2, p, \gamma) = \int_0^\infty e^{-(\alpha+\beta)t} \left\{ r_1\mu_1(\phi_t^\gamma(p))\gamma_t(1)\mathbb{1}\big(i_1 > 0\big) + r_2\mu_2\gamma_t(2)\mathbb{1}\big(i_2 > 0\big) \right\} dt.$$

Again a optimality equation can be formulated, but for this reward criterion a more explicit solution can be given by another method: the optimal control is a pure one which is completely characterized by an index.

Remember first remark 5.7 where we introduced the notation $T_t(1)$ as service time at queue 1 in $[0, t]$, in particular

$$\begin{aligned} T_t(1) &:= \int_0^t u_t(1)dt \quad \Longrightarrow \quad dT_t(1) = u_t(1)dt \\ T_t(2) &:= \int_0^t u_t(2)dt \quad \Longrightarrow \quad dT_t(2) = u_t(2)dt. \end{aligned}$$

Due to $u_t(1) + u_t(2) = 1$ we get $T_t(1) + T_t(2) = t$. Then we define $(\tilde{p}_t, \tilde{X}_t)$ as solution of the control independent analogons to (5.11) and (2.12)

$$\begin{aligned} dp_t &= (\mu_1^B - \mu_1^A)p_t(1-p_t)dt + \Phi(p_{t-})dN_t^X(X_{t-}^1, X_{t-}^1 - 1) \\ dX_t &= Q^X(p_t)dt + d\widehat{M}_t^X \\ (p_0, X_0) &= (p, x_0), \end{aligned}$$

where $Q^X(p) = \left(q_{ij}^X(p)\right)$ with

$$q_{ij}^X(p) = \begin{cases} \mu_1(p) & j = i - e_1 \\ \mu_2 & j = i - e_2 \\ \lambda_1 & j = i + e_1 \\ \lambda_2 & j = i + e_2 \\ -\lambda_1 - \lambda_2 - \mu_1(p) - \mu_2 & j = i \end{cases}$$

These two equations are the control independent stochastic differential equations for the state process, denoting the number of waiting customers in each queue and the estimator process for the unknown service rate. Since the jump-sizes of X_t and p_t are independent of the control, it is clear that for a fixed control $u = (u_t)$ the following relations are true:

$$p_t^u = \tilde{p}_{T_t(1)} \quad \text{and} \quad \left(X_t^{1,u}, X_t^{2,u}\right) = \left(\tilde{X}_{T_t(1)}^1, \tilde{X}_{T_t(2)}^2\right).$$

Here we stress the dependence of the state process $X_t^{i,u}$ on the control process u. Thus we see there is a one-to-one relation between the process controlled through u at time t and the uncontrolled process considered at the time point $T_t(i)$, means after serving $T_t(i)$ units of time.

Additionally, we transform the objective function in a notation depending on $T_t(i)$ by

$$\begin{aligned} &\mathbb{E}_u\left[\int_0^\infty e^{-\beta t}\left(r_1\mu_1(p_t^u)\mathbb{1}\left(X_t^{1,u} > 0\right)u_t(1) + r_2\mu_2\mathbb{1}\left(X_t^{2,u} > 0\right)u_t(2)\right)dt\right] \\ = \;& \mathbb{E}\left[\int_0^\infty e^{-\beta t}\left(r_1\mu_1(\tilde{p}_{T_t(1)})\mathbb{1}\left(\tilde{X}_{T_t(1)}^1 > 0\right)dT_t(1) + r_2\mu_2\mathbb{1}\left(\tilde{X}_{T_t(2)}^2 > 0\right)dT_t(2)\right)\right], \end{aligned}$$

where we change the notation from $\mathbb{E}_u$ to $\mathbb{E}$ due to the previous comments. In the next theorem we prove the optimality of a pure index strategy.

Theorem 5.29 *Define by*

$$\Gamma(i_1, p) := esssup_{\sigma>0}\frac{\mathbb{E}\left[\int_0^\sigma e^{-\beta t}r_1\mu_1(\tilde{p}_t)\mathbb{1}\left(\tilde{X}_t^1 > 0\right)dt \mid \tilde{p}_0 = p, \tilde{X}_0^1 = i_1\right]}{\mathbb{E}\left[\int_0^\sigma e^{-\beta t}dt \mid \tilde{p}_0 = p, \tilde{X}_0^1 = i_1\right]}$$

the Gittins index for queue 1. *Then* (u_t^*) *with*

$$u_t^* := u^*(\tilde{X}^1_{T_t^*(1)}, \tilde{X}^2_{T_t^*(2)}, \tilde{p}_{T_t^*(1)}) = \begin{cases} 1 & \tilde{X}^1_{T_t^*(1)} > 0, \Gamma(\tilde{X}^1_{T_t^*(1)}, \tilde{p}_{T_t^*(1)}) \geq r_2\mu_2 \\ 2 & \tilde{X}^2_{T_t^*(2)} > 0, \Gamma(\tilde{X}^1_{T_t^*(1)}, \tilde{p}_{T_t^*(1)}) < r_2\mu_2 \\ 1 & \tilde{X}^2_{T_t^*(2)} = 0 \\ 2 & \tilde{X}^1_{T_t^*(1)} = 0 \end{cases}$$

is an optimal control, where the $T_t^*(i)$ *corresponds to* u_t^* *via* $T_t^*(i) = \int_0^t u_s^*(i)ds$.

Proof: Consider first with the product rule

$$\begin{aligned} & \int_0^\infty \left(\int_{T_t(1)}^\infty e^{-\beta s} \mu_1(\tilde{p}_s) \mathbb{1}\big(\tilde{X}_s^1 > 0\big) ds \right) de^{-\beta(t-T_t(1))} \\ = & \int_{T_t(1)}^\infty e^{-\beta s} \mu_1(\tilde{p}_s) \mathbb{1}\big(\tilde{X}_s^1 > 0\big) ds e^{-\beta(t-T_t(1))} \Big|_{t=0}^\infty \\ & \qquad - \int_0^\infty e^{-\beta(t-T_t(1))} d\left(\int_{T_t(1)}^\infty e^{-\beta s} \mu_1(\tilde{p}_s) \mathbb{1}\big(\tilde{X}_s^1 > 0\big) ds \right) \\ = & -\int_0^\infty e^{-\beta s} \mu_1(\tilde{p}_s) \mathbb{1}\big(\tilde{X}_s^1 > 0\big) ds \\ & \qquad - \int_0^\infty e^{-\beta(t-T_t(1))} d\left(\int_{T_t(1)}^\infty e^{-\beta s} \mu_1(\tilde{p}_s) \mathbb{1}\big(\tilde{X}_s^1 > 0\big) ds \right). \end{aligned}$$

Thus we get

$$\begin{aligned} & \int_0^\infty e^{-\beta t} \mu_1(\tilde{p}_{T_t(1)}) \mathbb{1}\big(\tilde{X}^1_{T_t(1)} > 0\big) dT_t(1) \\ = & -\int_0^\infty e^{-\beta(t-T_t(1))} d\left(\int_{T_t(1)}^\infty e^{-\beta s} \mu_1(\tilde{p}_s) \mathbb{1}\big(\tilde{X}_s^1 > 0\big) ds \right) \\ = & \int_0^\infty e^{-\beta s} \mu_1(\tilde{p}_s) \mathbb{1}\big(\tilde{X}_s^1 > 0\big) ds + \int_0^\infty \left(\int_{T_t(1)}^\infty e^{-\beta s} \mu_1(\tilde{p}_s) \mathbb{1}\big(\tilde{X}_s^1 > 0\big) ds \right) de^{-\beta(t-T_t(1))}. \end{aligned}$$

Taking expectation and using the Markov-property of our model we continue

$$\begin{aligned} & \mathbb{E}\left[\int_0^\infty e^{-\beta t} \mu_1(\tilde{p}_{T_t(1)}) \mathbb{1}\big(\tilde{X}^1_{T_t(1)} > 0\big) dT_t(1) \right] \\ = & \mathbb{E}\left[\int_0^\infty e^{-\beta s} \mu_1(\tilde{p}_s) \mathbb{1}\big(\tilde{X}_s^1 > 0\big) ds \right] \\ & \qquad + \mathbb{E}\left[\int_0^\infty \left(\int_{T_t(1)}^\infty e^{-\beta s} \mu_1(\tilde{p}_s) \mathbb{1}\big(\tilde{X}_s^1 > 0\big) ds \right) de^{-\beta(t-T_t(1))} \right] \\ = & \mathbb{E}\left[\int_0^\infty e^{-\beta s} \mu_1(\tilde{p}_s) \mathbb{1}\big(\tilde{X}_s^1 > 0\big) ds \right] \\ & \qquad + \mathbb{E}\left[\int_0^\infty \mathbb{E}\left\{ \int_{T_t(1)}^\infty e^{-\beta s} \mu_1(\tilde{p}_s) \mathbb{1}\big(\tilde{X}_s^1 > 0\big) ds \mid \mathcal{F}^X_{T_t(1)} \right\} de^{-\beta(t-T_t(1))} \right] \end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}\left[\int_0^\infty e^{-\beta s}\mu_1(\tilde{p}_s)\mathbb{1}\big(\tilde{X}_s^1>0\big)ds\right] \\
&\qquad +\mathbb{E}\left[\int_0^\infty \mathbb{E}\left\{\int_{T_t(1)}^\infty e^{-\beta s}\mu_1(\tilde{p}_s)\mathbb{1}\big(\tilde{X}_s^1>0\big)ds \mid \tilde{X}_{T_t(1)},\tilde{p}_{T_t(1)}\right\} de^{-\beta(t-T_t(1))}\right].
\end{aligned}
$$

Applying the representation theorem of Bank and ElKaroui (2004) (with $f(t,l) := \beta e^{-\beta t} l$ and $X_t := \int_t^\infty e^{-\beta s}\mu_1(\tilde{p}_s)\mathbb{1}\big(\tilde{X}_s^1>0\big)ds$) results in

$$
\begin{aligned}
&\mathbb{E}\left[\int_0^\infty e^{-\beta s} r_1\mu_1(\tilde{p}_s)\mathbb{1}\big(\tilde{X}_s^1>0\big)ds\right] \\
&\qquad +\mathbb{E}\left[\int_0^\infty \mathbb{E}\left\{\int_{T_t(1)}^\infty e^{-\beta s} r_1\mu_1(\tilde{p}_s)\mathbb{1}\big(\tilde{X}_s^1>0\big)ds \mid \tilde{X}_{T_t(1)},\tilde{p}_{T_t(1)}\right\} de^{-\beta(t-T_t(1))}\right] \\
&= \mathbb{E}\left[\int_0^\infty e^{-\beta s}\inf_{\nu\in[0,s]}\Gamma(\tilde{X}_\nu^1,\tilde{p}_\nu)ds\right] \\
&\qquad +\mathbb{E}\left[\int_0^\infty \mathbb{E}\left\{\int_{T_t(1)}^\infty e^{-\beta s}\inf_{\nu\in[T_t(1),s]}\Gamma(\tilde{X}_\nu^1,\tilde{p}_\nu)ds \mid \tilde{X}_{T_t(1)},\tilde{p}_{T_t(1)}\right\} de^{-\beta(t-T_t(1))}\right] \\
&\le \mathbb{E}\left[\int_0^\infty e^{-\beta s}\inf_{\nu\in[0,s]}\Gamma(\tilde{X}_\nu^1,\tilde{p}_\nu)ds\right] \\
&\qquad +\mathbb{E}\left[\int_0^\infty \left(\int_{T_t(1)}^\infty e^{-\beta s}\inf_{\nu\in[0,s]}\Gamma(\tilde{X}_\nu^1,\tilde{p}_\nu)ds\right) de^{-\beta(t-T_t(1))}\right] \qquad (5.16)\\
&= \mathbb{E}\left[\int_0^\infty e^{-\beta s}\inf_{\nu\in[0,T_s(1)]}\Gamma(\tilde{X}_\nu^1,\tilde{p}_\nu)dT_s(1)\right],
\end{aligned}
$$

where the inequality holds true due to $de^{-\beta(t-T_t(1))} \le 0$, since $e^{-\beta(t-T_t(1))}$ is monotone decreasing in t. The last equality follows analogously as in the beginning of the proof by applying the product rule.

Thus we have found the following bounds for the objective function:

$$
\begin{aligned}
&\mathbb{E}\left[\int_0^\infty e^{-\beta s}\left(r_1\mu_1(\tilde{p}_{T_s(1)})\mathbb{1}\big(\tilde{X}_{T_s(1)}^1>0\big)dT_s(1) + r_2\mu_2\mathbb{1}\big(\tilde{X}_{T_s(2)}^2>0\big)dT_s(2)\right)\right] \\
\le\ &\mathbb{E}\left[\int_0^\infty e^{-\beta s}\left(\inf_{\nu\in[0,T_s(1)]}\Gamma(\tilde{X}_\nu^1,\tilde{p}_\nu)dT_s(1) + r_2\mu_2\mathbb{1}\big(\tilde{X}_{T_s(2)}^2>0\big)dT_s(2)\right)\right] \qquad (5.17)\\
\le\ &\sup_{(\tilde{T})}\mathbb{E}\left[\int_0^\infty e^{-\beta s}\left(\inf_{\nu\in[0,\tilde{T}_s(1)]}\Gamma(\tilde{X}_\nu^1,\tilde{p}_\nu)d\tilde{T}_s(1) + r_2\mu_2\mathbb{1}\big(\tilde{X}_{\tilde{T}_s(2)}^2>0\big)d\tilde{T}_s(2)\right)\right] \qquad (5.18)
\end{aligned}
$$

We only have to prove, that for our strategy T^* equality holds in (5.17) and (5.18). Consider from (5.16) the expression

$$
\int_0^\infty \mathbb{E}\left[\int_{T_t(1)}^\infty e^{-\beta s}\left(\inf_{\nu\in[T_t(1),s]}\Gamma(\tilde{X}_\nu^1,\tilde{p}_\nu) - \inf_{\nu\in[0,s]}\Gamma(\tilde{X}_\nu^1,\tilde{p}_\nu)\right)ds \mid \tilde{X}_{T_t(1)},\tilde{p}_{T_t(1)}\right] de^{-\beta(t-T_t(1))}.
$$

If this expression is equal to 0, then we have equality in (5.16) and consequently in (5.17). Since

$$\inf_{\nu\in[T_t(1),s]} \Gamma(\tilde{X}^1_\nu, \tilde{p}_\nu) - \inf_{\nu\in[0,s]} \Gamma(\tilde{X}^1_\nu, \tilde{p}_\nu) \tag{5.19}$$

is lower-semi-right-continuous and greater or equal 0, we have equality in (5.16) if and only if

$$dT_t(1) = u_t(1)dt < 1 \quad \Longleftrightarrow \quad \inf_{\nu\in[T_t(1),s]} \Gamma(\tilde{X}^1_\nu, \tilde{p}_\nu) = \inf_{\nu\in[0,s]} \Gamma(\tilde{X}^1_\nu, \tilde{p}_\nu).$$

This is the case for T^*, since if $\Gamma(\tilde{X}_{T^*_t(1)}, \tilde{p}_{T^*_t(1)}) < r_2\mu_2$ then we have by definition of T^* that $\Gamma(\tilde{X}_{T^*_t(1)}, \tilde{p}_{T^*_t(1)}) = \inf_{\nu\in[0,T^*_t(1)]} \Gamma(\tilde{X}^1_\nu, \tilde{p}_\nu)$ and hence (5.19) is zero.

If we consider (5.18) we see that $s \mapsto \inf_{\nu\in[0,s]} \Gamma(\tilde{X}_\nu,^1 \tilde{p}_\nu)$ is monotone increasing. Hence the myopic strategy is optimal and for this control equality holds in (5.18), that means

$$u_t(1) = 1 \quad \Longleftrightarrow \quad \inf_{\nu\in[0,T_t(1)]} \Gamma(\tilde{X}^1_\nu, \tilde{p}_\nu) \geq r_2\mu_2.$$

But this condition is equivalent for $T^*_t(1)$ due to its definition to $\Gamma(\tilde{X}^1_{T^*_t(1)}, \tilde{p}_{T^*_t(1)}) \geq r_2\mu_2$. □

As an immediate consequence of theorem 5.29 we get the following characterization of the value function:

Corollary 5.30 *It holds with $(T^*_t(1), T^*_t(2))$ from theorem 5.29:*

$$J(i_1, i_2, p) = \mathbb{E}\left[\int_0^\infty e^{-\beta s}\left(\inf_{\nu\in[0,T^*_s(1)]} \Gamma(\tilde{X}^1_\nu, \tilde{p}_\nu)dT^*_s(1) + r_2\mu_2 \mathbb{1}\left(\tilde{X}^2_{T^*_s(2)} > 0\right)dT^*_s(2)\right)\right].$$

Remark 5.31

a) We have seen that if the optimal strategy has switched from queue 1 to queue 2 it will remain there unless queue 2 is empty or a new arrival occurs at queue 1. New arrivals at queue 2 have no influence to the optimal action in this moment, since $\Gamma(i_1, p)$ is independent of queue 2. On the other hand an arrival at queue 1 makes queue 1 more likely to serve, in particular the optimal server will remain at queue 1 if he was at queue 1 before the arrival since

$$\begin{aligned}
\Gamma(i_1+1, p) &= esssup_{\sigma>0} \frac{\mathbb{E}\left[\int_0^\sigma e^{-\beta t} r_1\mu_1(\tilde{p}_t)\mathbb{1}\left(\tilde{X}^1_t > 0\right)dt \mid \tilde{p}_0 = p, \tilde{X}^1_0 = i_1+1\right]}{\mathbb{E}\left[\int_0^\sigma e^{-\beta t}dt \mid \tilde{p}_0 = p, \tilde{X}^1_0 = i_1+1\right]} \\
&\geq esssup_{\sigma>0} \frac{\mathbb{E}\left[\int_0^\sigma e^{-\beta t} r_1\mu_1(\tilde{p}_t)\mathbb{1}\left(\tilde{X}^1_t > 0\right)dt \mid \tilde{p}_0 = p, \tilde{X}^1_0 = i_1\right]}{\mathbb{E}\left[\int_0^\sigma e^{-\beta t}dt \mid \tilde{p}_0 = p, \tilde{X}^1_0 = i_1\right]} \\
&= \Gamma(i_1, p),
\end{aligned}$$

since $\mathbb{1}\left(\tilde{X}^1_t(i_1+1) > 0\right) \geq \mathbb{1}\left(\tilde{X}^1_t(i_1) > 0\right)$. Hence $i_1 \mapsto \Gamma(i_1, p)$ is monotone increasing.

b) *The optimality of this index strategy can be extended from the Bayesian setting to the Hidden-Markov-Model, where μ_1 changes over time according to an unobservable environment process (Z_t). The proof works completely analogous since we did not make use of the behaviour of p_t.*

c) *If both, μ_1 and μ_2 are unknown the optimal strategy is again a index strategy, but the existence of a pure index strategy is not guaranteed anymore.*

d) *Our numerical studies indicate that if the service of a customer in queue 1 is finished it is never optimal to change to queue 2, expect it was the last waiting customer in queue 1. In particular it indicates $\Gamma(i_1, p) \geq r_2\mu_2 \Rightarrow \Gamma(i_1 - 1, p + \Phi(p)) \geq r_2\mu_2$, from which the stay-on-a-winner property follows.*

We have seen in this proof that the model with the reward criterion is a classical bandit problem. This is due to the special structure of the rewards, whereas the model with waiting costs discussed in section 5.2.2 is not a bandit problem. This is explained by the fact, that the waiting costs at both queues are not independent.

Remark 5.32 *The proof of theorem 5.29 is adopted to Bank and Küchler (2007). In this work the authors consider a bandit problem in continuous time and prove that the set of optimal allocation strategies is equal to the set of so-called index strategies. This Gittins theorem is well-known and proven in various ways, see for example ElKaroui and Karatzas (1997), Kaspi and Mandelbaum (1995) and Kaspi and Mandelbaum (1998). Our statement is slightly different, since we do not claim that every optimal control has to be of the structure of u_t^*. In particular we drop here the special assumption of the synchronisation property. But with the same ideas we are able to show that in a classical two-armed-bandit model, where one service rate is Bayesian and the other one is known, index-strategies are pure strategies. This can be obtained, by proving in the notations of Bank and Küchler (2007), that $t \in \mathcal{D}$ implies $\sigma_2(N(t)-) = 0$.*

If there are no arrivals to the system we can prove under the same conditions as in theorem 5.25 that it is optimal to serve queue 2 until it is empty. But with this reward structure the proof simplifies in various way and we do not need the recursion in the state space technique. If there are no arrivals and if $\mu_1^A > 0$ then similar to lemma 5.24 there exists a random variable $\tau(i_1, i_2, p)$ such that for all $t \geq \tau(i_1, i_2, p)$ both queues are empty and the program terminates. It is easy to prove, that the definition of the Gittins index can be extended to stopping times (see e.g. Bank and ElKaroui (2004)) as

$$\Gamma(i_1, p) = \operatorname{esssup}_{\sigma \in (0, \tau(i_1, i_2, p))} \frac{\mathbb{E}\left[\int_0^\sigma e^{-\beta t} r_1 \mu_1(\tilde{p}_t) \mathbb{1}\left(\tilde{X}_t^1 > 0\right) dt \mid \tilde{p}_0 = p, \tilde{X}_0^1 = i_1\right]}{\mathbb{E}\left[\int_0^\sigma e^{-\beta t} dt \mid \tilde{p}_0 = p, \tilde{X}_0^1 = i_1\right]}.$$

Then we are able to claim in the spirit of theorem 5.25 the following sufficient condition for an optimal control.

Theorem 5.33 *If $r_1\mu_1(\mathcal{M}^{i_1-1}p) < r_2\mu_2$ then serving queue 2 unless it is empty is optimal.*

Proof: It is sufficient due to theorem 5.29 and remark 5.31 to prove that $\Gamma(i_1, p) < r_2\mu_2$. This is true because of

$$\begin{aligned}\Gamma(i_1,p) &= \operatorname{esssup}_{\sigma\in(0,\tau(i_1,i_2,p))} \frac{\mathbb{E}\left[\int_0^\sigma e^{-\beta t} r_1\mu_1(\tilde{p}_t)\mathbb{1}\left(\tilde{X}_t^1>0\right)dt \mid \tilde{p}_0=p, \tilde{X}_0^1=i_1\right]}{\mathbb{E}\left[\int_0^\sigma e^{-\beta t}dt \mid \tilde{p}_0=p, \tilde{X}_0^1=i_1\right]} \\ &\leq r_1\mu_1(\mathcal{M}^{i_1-1}p) < r_2\mu_2.\end{aligned}$$

□

The proof can be simplified once more by making use of the objective function

$$\mathbb{E}\left[\int_0^{\tau(i_1,i_2,p)} e^{-\beta t}\left(r_1\mu_1(\tilde{p}_t)\mathbb{1}\left(\tilde{X}_t^1>0\right)\mathbb{1}\left(u_t=1\right)+r_2\mu_2\mathbb{1}\left(\tilde{X}_t^2>0\right)\mathbb{1}\left(u_t=2\right)\right)dt\right].$$

By assumption we know $r_1\mu_1(\tilde{p}_t) < r_2\mu_2$ for all $t \in [0, \tau(i_1, i_2, p)]$. Thus it is obvious, that queue 2 has a higher priority to queue 1.

This model with reward criterion is quite similar to the model completely solved in Donchev and Yushkevich (1996), Donchev (1998) and Donchev (1999). Especially the model in Donchev (1998) is very related to our model with two difference. First, we do not divide a constant flow to two servers, in contrast we have stochastic arrivals at each queue and the queues are served. As a consequence the Gittins index in our model depends on the current length of the queue. Second, Donchev (1998) assumed that both service parameters are unknown in a symmetric way as in section 5.2.4. This means $\mu_1, \mu_2 \in \{\mu^A, \mu^B\}$ with $\mu_1 = \mu^A$ if and only if $\mu_2 = \mu^B$ where the parameters change at random times. He solved the model with the help of variational inequalities, but using the theory of bandit problems yields in the same results as indicated in remark 5.31.

5.3 Unknown Length of the Queues: the 0-1-Observation

Assume now that the server is not able to observe queue 1 completely. The server can only differ if there are more than two customers waiting or not. For simplicity, assume that queue 2 and all parameters are completely known, in particular we are in the case of a 0-1-observation, see page 11. Thus the information structure for i_1 is given by

$$I(1) = \{0, 1\} \quad \text{and} \quad I(2) = \{2, 3, \ldots\}.$$

Assume furthermore that $c_1\mu_1 > c_2\mu_2$. Hence it is always optimal to apply the $c\mu$-rule and serve queue 1 if the server obtains $i_1 \in I(2)$. But what is the optimal service allocation (or at least a well-performing) if the server only knows $i_1 \in I(1)$, that means maybe there is a customer waiting or maybe not.

As in section 5.2 we derive a partial differential equation for the estimator process $p_t := \mathbb{P}_u(X_t^1 = 0 \mid \mathcal{F}_t^Y)$. Then we state the HJB-equation and suggest some reasonable strategies, which we compare numerically, since the optimal control can not be computed analytically.

If the observation changes at time τ from f_2, especially $i_1 \in I(2)$, to f_1, in particular $i_1 \in I(1)$, then the server knows that in this moment there is exactly one customer waiting in queue 1, hence $p_\tau = 0$. Starting from this a-priori probability the estimator process evolves as $p_t = \phi_{t-\tau}(0)$ where $\phi_t(0)$ is the unique solution of

$$\begin{cases} \dot{p} &= -\lambda_1 p^2 + \mu_1 u(1-p) \\ p_0 &= 0. \end{cases}$$

If queue 1 is not served, that means $u = 0$, the estimator is decreasing. This is reasonable, since no customer can leave queue 1 whereas only new customers may join queue 1. The generalized HJB-equation can be stated under the observation $Y_t = f_1$, in particular under $i_1 \in I(1)$, as

$$\begin{aligned} & \beta W(p, i_2, f_1) \\ = & \inf_{\substack{\xi \in \partial_p W(p,i_2,f_1) \\ u \in U}} \Big\{ c_1(1-p) + c_2 i_2 + \xi(\mu_1(1-p)u - \lambda_1 p^2) \\ & \quad + \big(W(0, i_2, f_2) - W(p, i_2, f_1)\big)\lambda_1(1-p) + \big(W(p, i_2+1, f_1) - W(p, i_2, f_1)\big)\lambda_2 \\ & \quad + \big(W(p, i_2-1, f_1) - W(p, i_2, f_1)\big)\mu_2(1-u) \Big\}. \end{aligned}$$

As mentioned above we discuss two reasonable strategies for this model, where the state $i_1 = 0$ is not observable completely.

5.3.1 Threshold-Strategy

The threshold-strategy will serve queue 1 if the estimated probability that no customer is waiting is under a given threshold p^*. Otherwise it serves queue 2. It is clear, that if the observation process Y_t changes from f_2 to f_1 that the server continues serving queue 1, since in the moment of the change at time τ the estimator p starts in $p_\tau = 0$. Define σ_1 as the first time when the threshold is reached after τ and σ_2 as the first time after τ where the observation changes from f_1 to f_2. Then with $\sigma := \min\{\sigma_1, \sigma_2\}$ the estimator is given for $t \in [\tau, \sigma)$ by $p_t = \phi_{t-\tau}(0)$ where

$$\phi_t(0) = \frac{-\mu_1 + \tanh\left\{\frac{1}{2}t\rho + \frac{1}{2}\ln\left(\frac{\rho+\mu_1}{\rho-\mu_1}\right)\right\}\rho}{2\lambda_1}$$

with $\rho := \sqrt{4\mu_1\lambda_1 + \mu_1^2}$.

A special threshold is the certainty equivalence threshold p^{CEP} defined by

$$p^{\mathrm{CEP}} := 1 - \frac{c_2\mu_2}{c_1\mu_1} \in (0,1).$$

Its name is justified by

$$c_1\mu_1(1-p^{\mathrm{CEP}}) = c_1\mu_1\frac{c_2\mu_2}{c_1\mu_1} = c_2\mu_2,$$

that means, p^{CEP} is the value for which the estimate

$$c_1\mu_1(1-p) = c_1\mu_1\mathbb{P}(\text{one customer is waiting in queue 1})$$

is equal to $c_2\mu_2$. If $c_2 > c_1$ under $c_1\mu_1 > c_2\mu_2$ then p^{CEP} is monotone increasing in c_2 for fixed c_1, μ_1, μ_2.

The following figure illustrates the relative costs of threshold policies with respect to the expected cost under complete information, that means on the y-axis we have

$$\gamma := \frac{V^{\text{threshold}}}{V^{\text{complete information}}},$$

for different thresholds, denoted on the x-axis. The values were chosen by $\lambda_1 = \lambda_2 = 0.1$, $\mu_1 = 0.3, \mu_2 = 0.4, c_1 = 2, c_2 = 1, \beta = 0.9, i_1(0) = 1, i_2(0) = 2$, hence $p^{\mathrm{CEP}} = \frac{1}{3}$.

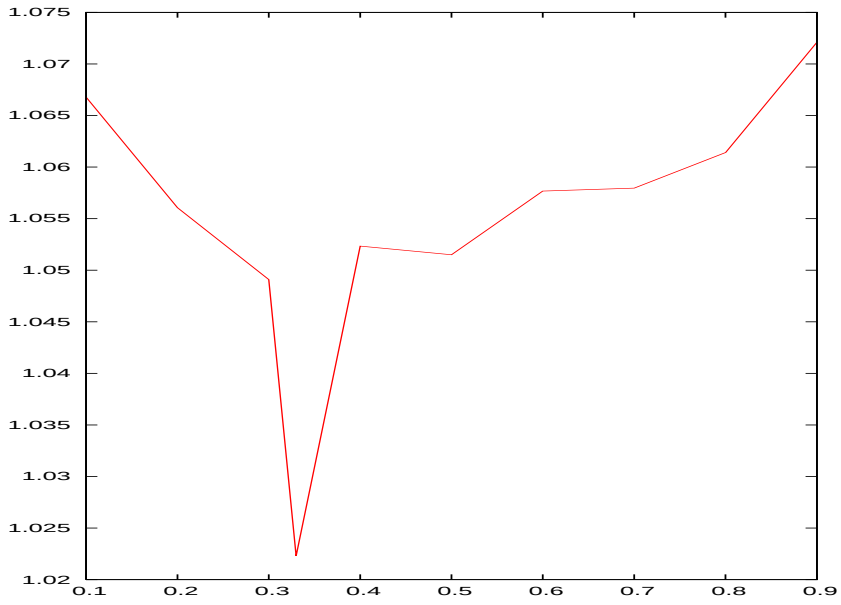


Figure 6: Relative Costs under Threshold-Strategies

The results of various numerical studies can be summarized as follows:

- If the threshold is near 1 then γ is higher, since the server undervalues the fact, that one customer is still waiting in queue 1. On the other hand, if the threshold is near 0 the server overrates this fact and probably serves an empty queue.
- The certainty equivalence principle threshold control works very well under all threshold strategies and mostly $\gamma = \frac{V^{\text{threshold}}}{V^{\text{complete information}}}$ attains its minimum for this control.
- If $c_2 > c_1$ under $c_1\mu_1 \geq c_2\mu_2$, then strategies with higher threshold fit better, which is reasonable, since the waiting costs for the last waiting customer in queue 1 are significant compared to the waiting costs in queue 2.
- γ is decreasing in i_1, which is due to the fact, that X_t^1 (denoting the length of queue 1) stays more time in the second observation group $I(2) = \{2, 3, \ldots\}$, for which the (complete information) $c\mu$-rule is optimal.

5.3.2 Double-Threshold-Strategy

We have seen that the estimator p_t is monotone decreasing if queue 1 is not served. If the threshold is reached a threshold strategy will stop the service of queue 1, the estimator decreases, is then under the threshold again and service will restart. Therefore it seems to be more reasonable to wait a certain time until the estimator p_t reaches a lower level p_*. This level is less than the first threshold p^*. If the lower level is reached the server changes service back to queue 1 until the upper threshold p^* is reached again. Through the waiting period it becomes more likely that a customer is waiting in queue 1. We will call this kind of strategy double-threshold-strategy. It is illustrated in the following figure:

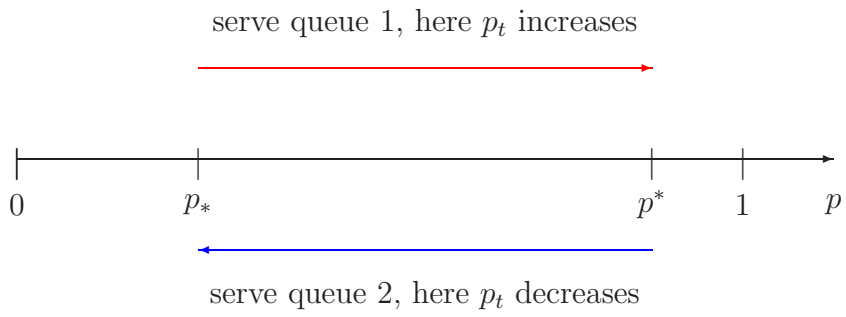


Figure 7: Double-Threshold-Strategy

For $p_* = p^*$ the double-threshold-strategy simplifies to the normal threshold-strategy. Since the double-threshold-strategy evolves like a threshold-strategy if the lower threshold was

reached at time τ we find in this case an equivalent closed formula for p_t as in in the previous section, except the process does not start in 0 (except for the case, where the observation just changed) but in p_*. Hence $p_t = \phi_{t-\tau}(p_*)$ where

$$\phi_t(p) = \frac{-\mu_1 + \tanh\left\{\frac{1}{2}t\rho + \frac{1}{2}\ln\left(\frac{\rho+\mu_1+2\lambda_1 p}{\rho-\mu_1-2\lambda_1 p}\right)\right\}\rho}{2\lambda_1}.$$

If the estimator p_t reaches the threshold p^* at time σ, then queue 2 is served and p_t is decreasing until p_* is reached again or a change in the observation occurs. If queue 1 is not served the estimator evolves as $p_t = \phi_{t-\sigma}(p^*)$ where

$$\phi_t(p) = \frac{1}{\lambda_1 t + \frac{1}{p}}.$$

In the next figure we illustrate the results of our numerical studies. Only strategies with $p_* \leq p^*$ are considered. On the left axis p_* and on the right p^* is marked. Again we consider the relative cost $\gamma = \frac{V^{\text{double-threshold}}}{V^{\text{complete information}}}$ on the vertical axis.

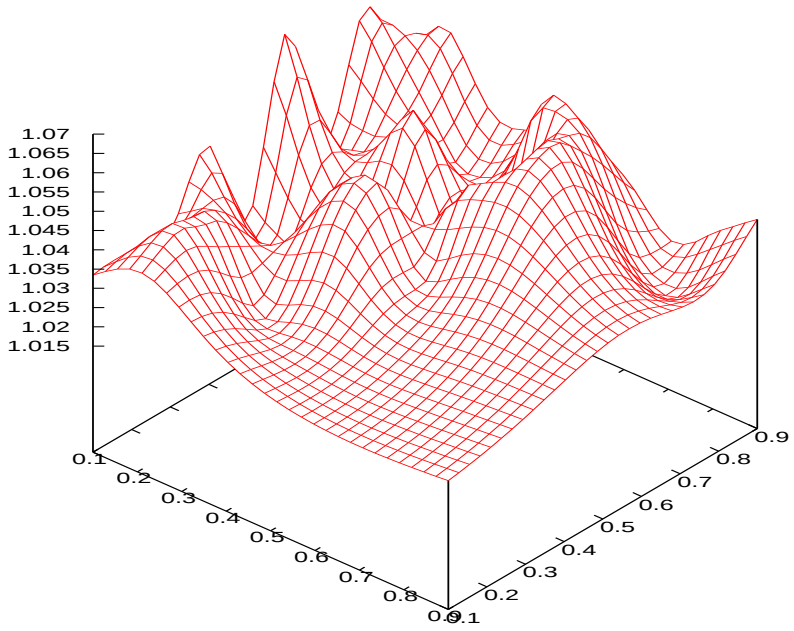


Figure 8: Relative Costs under Double-Threshold-Strategies

Our impression from the numerical investigation is that double-threshold-strategies perform better than one-threshold-strategies. But in none of our simulations a double-threshold-strategy bet the certainty equivalence principle threshold strategy.

6 Conclusion

Control models with partial information are treated in several publications over the last years. In particular, the Hidden-Markov-Model was investigated in various applications of financial portfolio optimization. Nevertheless, coarsening of the observations due to group representation of the states is considered nowhere, although it appears very often in the real world. This thesis closes this gap and introduces the notion of information structures. For the unobservable part of the state process a conditional probability is used as estimator and an explicite filter equation is derived. Additionally, we transform the optimization problem based on the unobservable process under incomplete information into one with complete information. We rigorously prove the equivalence of these two problems, often neglected by many authors in their works. Furthermore we investigate the dependence of the optimal value on the information structure.

Besides discussing properties and characteristics of the conditional probabilities and the value function we propose two methods for the solution of the transformed complete information model. We extend the Hamilton-Jacobi-Bellman equation and the corresponding verification technique by using the Clarke derivative. An advantage of this approach is that we require weaker assumptions instead of the strong ones in the classical case. These are fulfilled for our value function by its concavity. The second approach makes use of the piecewise-deterministic behaviour of the estimator process. For this purpose we define a time-discrete Markovian-Decision-Process (MDP) whose value function coincides with the value function of the original problem. Additionally, one can construct from an optimal policy of the MDP an optimal control for the original model. Hence we can use all tools of the established MDP-theory.

Combining all the developed results and applying to a parallel queueing model we analyze a setup with unknown Bayesian service rates strictly mathematically. Interesting results arise, as for example the separation property and the explicit characterization of the value function. Furthermore we show the existence of an optimal control which serves one queue exclusively almost everywhere. If one service rate is known, then the optimal control is a pure one. The last result holds true also for general bandit problems. Moreover we find sufficient conditions for the optimality of controls. The symmetric case is completely solved with the optimality of a threshold strategy. These results close a gap in the present queueing research, in particular the proofs demonstrate the power of our proposed solution procedures.

A Tools for Theorem 3.5

Here we append the lemmas and their proofs needed in the proof of theorem 3.5, where we derive the filter equation for $\widehat{X_t Z_t}$ and hence for p_t. Define for a process (H_t)

$$\triangle H_t := H_t - H_{t-}$$

and denote by $[H, \tilde{H}]_t$ its quadratic covariation with the process $(\tilde{H}_t)$.

Lemma A.1 *It holds:*

$$[X,Z]_t = \sum_{0<s\le t} \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\mu=1}^{d}\sum_{\nu=1}^{d} \delta_{ij}^{\mu\nu}(e_j - e_i)(g_\nu - g_\mu)\triangle N_s^Z(\mu,\nu) X_{s-}^i.$$

Proof:

$$\begin{aligned}
[X,Z]_t &= \sum_{0<s\le t} \triangle X_s \triangle Z_s \\
&= \sum_{0<s\le t} \Big(\sum_{i=1}^{n}\sum_{j=1}^{n} (e_j - e_i)\triangle N_s^X(i,j) \Big) \sum_{\mu=1}^{d}\sum_{\nu=1}^{d} (g_\nu - g_\mu)\triangle N_s^Z(\mu,\nu) \Big) \\
&\overset{(*)}{=} \sum_{0<s\le t} \Big(\sum_{i=1}^{n}\sum_{j=1}^{n} (e_j - e_i) \sum_{\mu=1}^{d}\sum_{\nu=1}^{d} \delta_{ij}^{\mu\nu}\triangle N_s^Z(\mu,\nu) X_{s-}^i \Big) \cdot \Big(\sum_{\mu=1}^{d}\sum_{\nu=1}^{d} (g_\nu - g_\mu)\triangle N_s^Z(\mu,\nu) \Big) \\
&\overset{(**)}{=} \sum_{0<s\le t} \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\mu=1}^{d}\sum_{\nu=1}^{d} \delta_{ij}^{\mu\nu}(e_j - e_i)(g_\nu - g_\mu)\triangle N_s^Z(\mu,\nu) X_{s-}^i,
\end{aligned}$$

where we used in $(*)$ the construction of $N_t^X(i,j)$ and the assumption that $\tilde{N}_t^X$ and N_t^Z do not jump at the same time. In $(**)$ we used $\triangle N_s^Z(\mu,\nu) \in \{0,1\}$ for all μ and ν which results in

$$\triangle N_s^Z(\mu,\nu) \cdot \triangle N_s^Z(\mu',\nu') = \begin{cases} 0 & \text{if } \mu \neq \mu' \text{ or } \nu \neq \nu' \\ \triangle N_s^Z(\mu,\nu) & \text{else.} \end{cases}$$

□

If $\delta_{ij}^{\mu\nu} \equiv 0$ for all i,j,μ,ν, (this means, there are no common jumps of X_t and Z_t), we obtain $[X,Z]_t \equiv 0$.

Remark A.2 *An analogous representation for the quadratic covariation is*

$$[X,Z]_t = \int_0^t \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\mu=1}^{d}\sum_{\nu=1}^{d} \delta_{ij}^{\mu\nu}(e_j - e_i)(g_\nu - g_\mu) dN_s^Z(\mu,\nu) X_{s-}^i.$$

Lemma A.3 *The $n \times d$-matrix $X_t Z_t$, where one entry is one and all others zero, has the following representation:*

$$\begin{aligned} X_t Z_t &= X_0 Z_0 + \int_0^t \sum_{i=1}^n \sum_{\mu=1}^d \big(e_i Q^Z g_\mu + (\tilde{Q}_\mu^X + \tilde{Q}_\mu^Z) e_i g_\mu\big) X_s^i Z_s^\mu ds \\ &\quad + \int_0^t \sum_{i=1}^n \sum_{j=1}^n \sum_{\mu=1}^d \sum_{\nu=1}^d \delta_{ij}^{\mu\nu} (e_j - e_i)(g_\nu - g_\mu) X_{s-}^i dN_s^Z(\mu,\nu) \\ &\quad + \int_0^t X_{s-} dM_s^Z + \int_0^t Z_{s-} dM_s^X. \end{aligned}$$

Proof: The proof is an easy application of the Itô-formula (convention: by $\int Z dX$ we mean $\int \big((dX)Z\big)$):

$$\begin{aligned} X_t Z_t &= X_0 Z_0 + \int_0^t X_{s-} dZ_s + \int_0^t Z_{s-} dX_s + [X,Z]_t \\ &= X_0 Z_0 + \int_0^t X_{s-}(Q^Z Z_s ds + dM_s^Z) + \int_0^t Z_{s-}(Q^X(Z_s) X_s ds + dM_s^X) + [X,Z]_t \\ &= X_0 Z_0 + \int_0^t X_s Q^Z Z_s ds + \int_0^t \big((\tilde{Q}_1^X, \ldots, \tilde{Q}_d^X) + (\tilde{Q}_1^Z, \ldots, \tilde{Q}_d^Z)\big) Z_s X_s Z_s ds \\ &\quad + [X,Z]_t + \int_0^t X_{s-} dM_s^Z + \int_0^t Z_{s-} dM_s^Z \\ &\overset{(*)}{=} X_0 Z_0 + \int_0^t \big(X_s Q^Z Z_s + \sum_{\mu=1}^d (\tilde{Q}_\mu^X + \tilde{Q}_\mu^Z) X_s Z_s^\mu g_\mu\big) ds \\ &\quad + [X,Z]_t + \int_0^t X_{s-} dM_s^Z + \int_0^t Z_{s-} dM_s^X \\ &= X_0 Z_0 + \int_0^t \sum_{i=1}^n \sum_{\mu=1}^d (e_i Q^Z g_\mu + (\tilde{Q}_\mu^X + \tilde{Q}_\mu^Z) e_i g_\mu) X_s^i Z_s^\mu ds \\ &\quad + [X,Z]_t + \int_0^t X_{s-} dM_s^Z + \int_0^t Z_{s-} dM_s^X. \end{aligned}$$

In $(*)$ we used

$$\begin{aligned} (a_1, \ldots, a_d) Z X Z &= \sum_{\mu=1}^d a_\mu Z^\mu \sum_{i=1}^n e_i X^i \sum_{r=1}^d g_r Z^r \\ = \sum_{i=1}^n \sum_{\mu=1}^d \sum_{r=1}^d X^i a_\mu Z^\mu Z^r e_i g_r &= \sum_{i=1}^n \sum_{\mu=1}^d X^i a_\mu Z^\mu e_i g_\mu = \sum_{\mu=1}^d a_\mu Z^\mu X g_\mu, \end{aligned}$$

where the second last equality holds, since exactly one entry of $Z \in S_Z$ is one and the others are zero, in particular $\sum_{r=1}^d \sum_{\nu=1}^d Z^r Z^\nu = \sum_{r=1}^d Z^r$. Lemma A.1 completes the proof. □

Since we estimate the process $X_t Z_t$ by $N_t^Y(k,l)$, which is equivalent to the estimation procedure by Y_t, we need a formula for the quadratic covariation between the unobservable process (XZ) and the observation $N_t^Y(k,l)$. The next lemma contains this formula.

Lemma A.4 *It holds for all $k,l \in \{1,\dots,m\}$:*

$$[(XZ), N^Y(k,l)]_t = \int_0^t \Big\{ \sum_{i\in I(k)} \sum_{j\in I(l)} (e_j - e_i) X^i_{s-} Z_s Y^k_{s-} d\tilde{N}^X_s(i,j) + \sum_{i\in I(k)} \sum_{j\in I(l)} \sum_{\mu=1}^{d} \sum_{\nu=1}^{d} \delta^{\mu\nu}_{ij} (e_j g_\nu - e_i g_\mu) X^i_{s-} Z^\mu_{s-} Y^k_{s-} dN^Z_s(\mu,\nu) \Big\}.$$

Proof: By construction of the processes Z_t, X_t and Y_t it is only possible that $N_t^Y(k,l)$ and $N_t^X(i,j)$ jump at the same time, if $i \in I(k)$ and $j \in I(l)$. $N_t^X(i,j)$ jumps if $\tilde{N}_t^X(i,j)$ jumps (and then Z_t does not jump) or if the jump is influenced by a jump of Z_t. Consequently we get:

$$\begin{aligned}
[(XZ), N^Y(k,l)]_t &= \sum_{0<s\le t} \triangle(X_s Z_s) \triangle N^Y_s(k,l) \\
&= \sum_{0<s\le t} \Big\{ \sum_{i=1}^{n} \sum_{j=1}^{n} (e_j - e_i) X^i_{s-} \triangle \tilde{N}^X_s(i,j) Z_s \triangle N^Y_s(k,l) \\
&\qquad + \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{\mu=1}^{d} \sum_{\nu=1}^{d} \delta^{\mu\nu}_{ij} (e_j g_\nu - e_i g_\mu) X^i_{s-} Z^\mu_{s-} \triangle N^Z_s(\mu,\nu) \triangle N^Y_s(k,l) \Big\} \\
&= \int_0^t \Big\{ \sum_{i\in I(k)} \sum_{j\in I(l)} (e_j - e_i) X^i_{s-} Z_s Y^k_{s-} d\tilde{N}^X_s(i,j) \\
&\qquad + \sum_{i\in I(k)} \sum_{j\in I(l)} \sum_{\mu=1}^{d} \sum_{\nu=1}^{d} \delta^{\mu\nu}_{ij} (e_j g_\nu - e_i g_\mu) X^i_{s-} Z^\mu_{s-} Y^k_{s-} dN^Z_s(\mu,\nu) \Big\},
\end{aligned}$$

where the last equality follows like $(**)$ in the proof of lemma A.1. □

We are now interested in the expectation of $X_t Z_t N_t^Y(k,l)$ which we need for the derivation of the filter equation. Note that we even derive in the proof of the next lemma an explicit representation for $X_t Z_t N_t^Y(k,l)$.

Lemma A.5 *It holds:*

$$\begin{aligned}
&\mathbb{E}\big[X_t Z_t N_t^Y(k,l)\big] \\
= \;& \mathbb{E}\Big[\int_0^t \sum_{i\in I(k)}\sum_{j\in I(l)} e_j\big(\sum_{\mu=1}^d g_\mu \tilde{q}^X_{ij,\mu} + \sum_{\nu=1}^d \delta_{ij}^{\mu\nu} g_\nu q^Z_{\mu\nu}\big) X^i_{s-} Z^\mu_{s-} Y^k_{s-} ds \\
&\quad + \int_0^t N^Y_{s-}(k,l)\Big\{\sum_{i=1}^n\sum_{\mu=1}^d (e_i Q^Z g_\mu + (\tilde{Q}^X_\mu + \tilde{Q}^Z_\mu) e_i g_\mu) X^i_s Z^\mu_s \\
&\qquad\qquad + \sum_{i=1}^n\sum_{j=1}^n\sum_{\mu=1}^d\sum_{\nu=1}^d \delta_{ij}^{\mu\nu}(e_j - e_i)(g_\nu - g_\mu) q^Z_{\mu\nu} Z^\mu_{s-} X^i_{s-}\Big\} ds\Big].
\end{aligned}$$

Proof: First we apply Itô (note that $N_0^Y(k,l) = 0$) and use the results from the lemmas above, hence

$$\begin{aligned}
X_t Z_t N_t^Y(k,l) &= \int_0^t X_{s-} Z_{s-} dN_s^Y(k,l) + \int_0^t N^Y_{s-}(k,l) d(X_s Z_s) + [XZ, N^Y(k,l)]_t \\
= \;& \int_0^t X_{s-} Z_{s-} dN_s^Y(k,l) + \int_0^t N^Y_{s-}(k,l)\Big\{\sum_{i=1}^n\sum_{\mu=1}^d (e_i Q^Z g_\mu + (\tilde{Q}^X_\mu + \tilde{Q}^Z_\mu) e_i g_\mu) X^i_s Z^\mu_s\Big\} ds \\
&+ \int_0^t N^Y_{s-}(k,l) \sum_{i=1}^n\sum_{j=1}^n\sum_{\mu=1}^d\sum_{\nu=1}^d \delta_{ij}^{\mu\nu}(e_j - e_i)(g_\nu - g_\mu) X^i_{s-} dN^Z_s(\mu,\nu) \\
&+ \int_0^t N^Y_{s-}(k,l) X_{s-} dM^Z_s + \int_0^t N^Y_{s-}(k,l) Z_{s-} dM^X_s \\
&+ \int_0^t \sum_{i\in I(k)}\sum_{j\in I(l)} (e_j - e_i) X^i_{s-} Z_s Y^k_{s-} d\tilde{N}^X_s(i,j) \\
&+ \int_0^t \sum_{i\in I(k)}\sum_{j\in I(l)}\sum_{\mu=1}^d\sum_{\nu=1}^d \delta_{ij}^{\mu\nu}(e_j g_\nu - e_i g_\mu) X^i_{s-} Z^\mu_{s-} Y^k_{s-} dN^Z_s(\mu,\nu).
\end{aligned}$$

Taking expectation on both sides, noticing that the expectation over the martingale-integrals is zero and using the property of the intensity of $N_t^Y(k,l)$

$$\mathbb{E}\big[dN_s^Y(k,l)\big] = \mathbb{E}\big[q^Y_{kl}(Z_s, X_s) Y^k_{s-} ds\big] = \mathbb{E}\big[\sum_{i\in I(k)}\sum_{j\in I(l)}\sum_{\mu=1}^d q^X_{ij,\mu} X^i_{s-} Z^\mu_{s-} Y^k_{s-} ds\big]$$

we get:

$$\begin{aligned}
&\mathbb{E}\big[X_t Z_t N_t^Y(k,l)\big] \\
= \;& \mathbb{E}\Big[\int_0^t X_{s-} Z_{s-} \sum_{i\in I(k)}\sum_{j\in I(l)}\sum_{\mu=1}^d q^X_{ij,\mu} X^i_{s-} Z^\mu_{s-} Y^k_{s-} ds
\end{aligned}$$

$$
\begin{aligned}
&+\int_0^t N^Y_{s-}(k,l)\Big\{\sum_{\mu=1}^d X_s Z^\mu_s Q^Z g_\mu + (\tilde{Q}^X_\mu + \tilde{Q}^Z_\mu) X_s Z^\mu_s g_\mu \\
&\qquad\qquad + \sum_{i=1}^n \sum_{j=1}^n \sum_{\mu=1}^d \sum_{\nu=1}^d \delta^{\mu\nu}_{ij}(e_j - e_i)(g_\nu - g_\mu) q^Z_{\mu\nu} Z^\mu_{s-} X^i_{s-}\Big\} ds \\
&+\int_0^t \sum_{i\in I(k)} \sum_{j\in I(l)} \sum_{\mu=1}^d (e_j - e_i) X^i_{s-} \tilde{q}^X_{ij,\mu} Z^\mu_{s-} \underbrace{Z_s}_{=g_\mu} Y^k_{s-} ds \\
&+\int_0^t \sum_{i\in I(k)} \sum_{j\in I(l)} \sum_{\mu=1}^d \sum_{\nu=1}^d \delta^{\mu\nu}_{ij}(e_j g_\nu - e_i g_\mu) X^i_{s-} Z^\mu_{s-} Y^k_{s-} q^Z_{\mu\nu} Z^\mu_{s-} ds\Big] \\
\overset{(*)}{=}\ & \mathbb{E}\Big[\int_0^t \sum_{i\in I(k)} \sum_{j\in I(l)} \sum_{\mu=1}^d e_i g_\mu q^X_{ij,\mu} X^i_{s-} Z^\mu_{s-} Y^k_{s-} ds \\
&+\int_0^t N^Y_{s-}(k,l)\{\dots\} ds + \int_0^t \sum_{i\in I(k)} \sum_{j\in I(l)} \sum_{\mu=1}^d e_j g_\mu \tilde{q}^X_{ij,\mu} X^i_{s-} Z^\mu_{s-} Y^k_{s-} ds \\
&+\int_0^t \sum_{i\in I(k)} \sum_{j\in I(l)} \sum_{\mu=1}^d \sum_{\nu=1}^d \delta^{\mu\nu}_{ij} e_j g_\nu q^Z_{\mu\nu} X^i_{s-} Z^\mu_{s-} Y^k_{s-} ds \\
&-\int_0^t \sum_{i\in I(k)} \sum_{j\in I(l)} \sum_{\mu=1}^d e_i g_\mu \underbrace{(\tilde{q}^X_{ij,\mu} + \sum_{\nu=1}^d \delta^{\mu\nu}_{ij} q^Z_{\mu\nu})}_{=q^X_{ij,\mu}} X^i_{s-} Z^\mu_{s-} Y^k_{s-} ds\Big] \\
=\ & \mathbb{E}\Big[\int_0^t \sum_{i\in I(k)} \sum_{j\in I(l)} \big(e_j \sum_{\mu=1}^d g_\mu \tilde{q}^X_{ij,\mu} + \sum_{\nu=1}^d \delta^{\mu\nu}_{ij} g_\nu q^Z_{\mu\nu}\big) X^i_{s-} Z^\mu_{s-} Y^k_{s-} ds \\
&+\int_0^t N^Y_{s-}(k,l)\Big\{\sum_{\mu=1}^d (e_i Q^Z g_\mu + (\tilde{Q}^X_\mu + \tilde{Q}^Z_\mu) e_i g_\mu) X^i_s Z^\mu_s \\
&\qquad\qquad + \sum_{i=1}^n \sum_{j=1}^n \sum_{\mu=1}^d \sum_{\nu=1}^d \delta^{\mu\nu}_{ij}(e_j - e_i)(g_\nu - g_\mu) q^Z_{\mu\nu} Z^\mu_{s-} X^i_{s-}\Big\} ds\Big].
\end{aligned}
$$

In $(*)$ we used $Z^\mu_{s-} Z^\mu_{s-} = Z^\mu_{s-} \in \{0,1\}$ by construction of S_Z and $X_{s-} Z_{s-} X^i_{s-} Z^\mu_{s-} = e_i g_\mu X^i_{s-} Z^\mu_{s-}$. □

Next we calculate as in lemma A.3 a representation for $\widehat{X_t Z_t} := \mathbb{E}[X_t Z_t \mid \mathcal{F}^Y_t]$. By writing $\widehat{(X_t Z_t)}_{i\cdot}$ we mean the i-th row of $\widehat{X_t Z_t}$ and by $\widehat{(X_t Z_t)}_{\cdot\mu}$ the μ-th column.

Lemma A.6 *It holds:*

$$\begin{aligned}\widehat{X_tZ_t} &= \widehat{X_0Z_0} + \int_0^t \sum_{i=1}^n \sum_{\mu=1}^d (e_iQ^Z g_\mu + (\tilde{Q}_\mu^X + \tilde{Q}_\mu^Z)e_i g_\mu)(\widehat{X_sZ_s})_{i\mu} ds \\ &\quad + \int_0^t \sum_{i=1}^n \sum_{j=1}^n \sum_{\mu=1}^d \sum_{\nu=1}^d \delta_{ij}^{\mu\nu}(e_j - e_i)(g_\nu - g_\mu)(\widehat{X_sZ_s})_{i\mu} q_{\mu\nu}^Z ds + \widehat{M_t},\end{aligned}$$

where $\widehat{M_t}$ is a $\mathcal{F}_t^Y$-martingale with expectation zero. It has the representation

$$\widehat{M_t} = \int_0^t \sum_{k=1}^m \sum_{l=1}^m \phi_{(k,l)}(s)\big(dN_s^Y(k,l) - q_{kl}^Y(\widehat{X_sZ_s})Y_s^k ds\big),$$

where $\phi_{(k,l)}(t) := \phi_{(k,l)}(\widehat{X_{t-}Z_{t-}})$ is $\mathcal{F}_t^Y$-predictable.

Proof: Using the definition of $\widehat{X_tZ_t}$ and lemma A.3 we get with Itô's formula:

$$\begin{aligned}\widehat{X_tZ_t} &= \mathbb{E}\Big[X_0Z_0 + \int_0^t \sum_{i=1}^n \sum_{\mu=1}^d (e_iQ^Z g_\mu + (\tilde{Q}_\mu^X + \tilde{Q}_\mu^Z)e_i g_\mu)(X_sZ_s)_{i\mu} ds \\ &\qquad + \int_0^t \sum_{i=1}^n \sum_{j=1}^n \sum_{\mu=1}^d \sum_{\nu=1}^d \delta_{ij}^{\mu\nu}(e_j - e_i)(g_\nu - g_\mu) X_{s-}^i dN_s^Z(\mu,\nu) \\ &\qquad + \int_0^t X_{s-} dM_s^Z + \int_0^t Z_{s-} dM_s^X \mid \mathcal{F}_t^Y\Big] \\ &= \mathbb{E}\Big[X_0Z_0 + \int_0^t \sum_{i=1}^n \sum_{\mu=1}^d (e_iQ^Z g_\mu + (\tilde{Q}_\mu^X + \tilde{Q}_\mu^Z)e_i g_\mu)(X_sZ_s)_{i\mu} ds \\ &\qquad + \int_0^t \sum_{i=1}^n \sum_{j=1}^n \sum_{\mu=1}^d \sum_{\nu=1}^d \delta_{ij}^{\mu\nu}(e_j - e_i)(g_\nu - g_\mu) X_{s-}^i (dN_s^Z(\mu,\nu) - q_{\mu\nu}^Z Z_{s-}^\mu ds) \\ &\qquad + \int_0^t \sum_{i=1}^n \sum_{j=1}^n \sum_{\mu=1}^d \sum_{\nu=1}^d \delta_{ij}^{\mu\nu}(e_j - e_i)(g_\nu - g_\mu) X_{s-}^i q_{\mu\nu}^Z Z_{s-}^\mu ds \\ &\qquad + \int_0^t X_{s-} dM_s^Z + \int_0^t Z_{s-} dM_s^X \mid \mathcal{F}_t^Y\Big] \\ &= \widehat{X_0Z_0} + \int_0^t \sum_{i=1}^n \sum_{\mu=1}^d (e_iQ^Z g_\mu + (\tilde{Q}_\mu^X + \tilde{Q}_\mu^Z)e_i g_\mu)(X_sZ_s)_{i\mu} ds \\ &\qquad + \int_0^t \sum_{i=1}^n \sum_{j=1}^n \sum_{\mu=1}^d \sum_{\nu=1}^d \delta_{ij}^{\mu\nu}(e_j - e_i)(g_\nu - g_\mu)(X_sZ_s)_{i\mu} q_{\mu\nu}^Z ds + \widehat{M_t},\end{aligned}$$

where the last equality holds due to Wong and Hajek (1985) and since the three martingales are summarized in $\widehat{M_t}$. The representation of this martingale is standard, see for example Brémaud (1981). □

Now we compute the expectation of $\widehat{X_tZ_t}N_t^Y(k,l)$ and the representation of this expression as in lemma A.5.

Lemma A.7 *It holds:*

$$\begin{aligned}
&\mathbb{E}\Big[\widehat{X_tZ_t}N_t^Y(k,l)\Big] \\
= \;& \mathbb{E}\Big[\int_0^t (\widehat{X_{s-}Z_{s-}} + \phi_{(k,l)}(s))q_{kl}^Y(\widehat{X_{s-}Z_{s-}})Y_{s-}^k ds \\
&\quad + \int_0^t N_{s-}^Y(k,l)\Big\{\sum_{i=1}^n\sum_{\mu=1}^d (e_iQ^Zg_\mu + (\tilde{Q}_\mu^X + \tilde{Q}_\mu^Z)e_ig_\mu)(\widehat{X_sZ_s})_{i\mu} \\
&\qquad\qquad + \sum_{i=1}^n\sum_{j=1}^n\sum_{\mu=1}^d\sum_{\nu=1}^d \delta_{ij}^{\mu\nu}(e_j-e_i)(g_\nu-g_\mu)(\widehat{X_sZ_s})_{i\mu}q_{\mu\nu}^Z ds\Big\}\Big].
\end{aligned}$$

Proof: Using Itô's formula (with $N_0^Y(k,l)=0$) and the fact that $\widehat{X_tZ_t}$ and $N_t^Y(k,l)$ jump only at the same time we get:

$$\begin{aligned}
&\widehat{X_tZ_t}N_t^Y(k,l) \\
= \;& \int_0^t \widehat{X_{s-}Z_{s-}}dN_s^Y(k,l) + \int_0^t N_{s-}^Y(k,l)d(\widehat{X_sZ_s}) + [\widehat{XZ}, N^Y(k,l)]_t \\
= \;& \int_0^t N_{s-}^Y(k,l)\Big\{\sum_{i=1}^n\sum_{\mu=1}^d (e_iQ^Zg_\mu + (\tilde{Q}_\mu^X + \tilde{Q}_\mu^Z)e_ig_\mu)(\widehat{X_sZ_s})_{i\mu} \\
&\qquad\qquad + \sum_{i=1}^n\sum_{j=1}^n\sum_{\mu=1}^d\sum_{\nu=1}^d \delta_{ij}^{\mu\nu}(e_j-e_i)(g_\nu-g_\mu)(\widehat{X_sZ_s})_{i\mu}q_{\mu\nu}^Z ds + d\widehat{M}_s\Big\} \\
&+ \int_0^t \widehat{X_{s-}Z_{s-}}dN_s^Y(k,l) + \sum_{0<s\le t}\phi_{(k,l)}(s)\triangle N_s^Y(k,l).
\end{aligned}$$

Taking expectation on both sides we get with the definition of the $\mathcal{F}_t^Y$-intensity of $N_t^Y(k,l)$:

$$\mathbb{E}\big[\widehat{X_tZ_t}N_t^Y(k,l)\big] = \mathbb{E}\Big[\int_0^t (\widehat{X_{s-}Z_{s-}} + \phi_{(k,l)}(s))q_{kl}^Y(\widehat{X_{s-}Z_{s-}})Y_{s-}^k ds + \int_0^t N_{s-}^Y(k,l)\big\{\dots\big\}ds\Big].$$

□

B Proof of Theorem 5.25

Here we add the detailed proof of theorem 5.25. Define

$$f(i_1,i_2,p) := \begin{cases} 2 & i_1>0, i_2>0 \\ 1 & i_2=0 \\ 2 & i_1=0 \end{cases}$$

and

$$g(i_1,i_2,p) := \begin{cases} 1 & i_1>0, i_2>0, t\in B \\ 2 & i_1>0, i_2>0, t\notin B \\ 1 & i_2=0 \\ 2 & i_1=0 \end{cases}$$

where $B\subset[0,\infty)$. Remember that the existence of an optimal pure control is justified by theorem 5.22. Consider then with $f_t := f(X_t^1, X_t^2, \phi_t^f(p))$ and $g_t := g(X_t^1, X_t^2, \phi_t^g(p))$ the strategies

$$\begin{aligned} \pi &= ((f_t)_{t\geq 0}, (g_t)_{t\geq 0}, (f_t)_{t\geq 0}, \ldots, (f_t)_{t\geq 0}) \in F^n \\ \tilde{\pi} &= ((g_t)_{t\geq 0}, (f_t)_{t\geq 0}, (f_t)_{t\geq 0}, \ldots, (f_t)_{t\geq 0}) \in F^n. \end{aligned}$$

Without loss of generality assume $B=[0,\varepsilon]$. Denote $V_{n-1,gf}$ the expected discounted reward over $(n-1)$-periods under the strategy $((g_t),(f_t),\ldots,(f_t))\in F^{n-1}$ and with the same interpretation for $V_{n-2,f}$ under $((f_t),(f_t),\ldots,(f_t))\in F^{n-2}$. Remember $r(i_1,i_2)=\frac{c_1 i_1+c_2 i_2}{\alpha+\beta}$ and compute

$$\begin{aligned} &V_{n\pi}(i_1,i_2,p) = \mathcal{T}_f V_{n-1,gf}(i_1,i_2,p) \\ &= r(i_1,i_2) + \int_0^\infty e^{-(\alpha+\beta)t}\Big\{V_{n-1,gf}(i_1,i_2,\underbrace{\phi_t^f(p)}_{=p})(\alpha-\mu_2) + V_{n-1,fg}(i_1,i_2-1,\underbrace{\phi_t^f(p)}_{=p})\mu_2\Big\}dt \\ &= r(i_1,i_2) + \int_0^\infty e^{-(\alpha+\beta)t}\Bigg\{r(i_1,i_2)(\alpha-\mu_2) + r(i_1,i_2-1)\mu_2 \\ &\qquad + \Big[\int_0^\varepsilon e^{-(\alpha+\beta)s}\Big\{V_{n-2,f}(i_1,i_2,\phi_s^g(p))(\alpha-\mu_1(\phi_s^g(p))) \\ &\qquad\qquad + V_{n-2,f}(i_1-1,i_2,\phi_s^g(p)+\Phi(\phi_s^g(p)))\mu_1(\phi_s^g(p))\Big\}ds \\ &\qquad + \int_\varepsilon^\infty e^{-(\alpha+\beta)s}\Big\{V_{n-2,f}(i_1,i_2,\phi_\varepsilon^g(p))(\alpha-\mu_2) \\ &\qquad\qquad + V_{n-2,f}(i_1,i_2-1,\phi_\varepsilon^g(p))\mu_2\Big\}ds\Big](\alpha-\mu_2) \\ &\qquad + \Big[\int_0^\varepsilon e^{-(\alpha+\beta)s}\Big\{V_{n-2,f}(i_1,i_2-1,\phi_s^g(p))(\alpha-\mu_1(\phi_s^g(p))) \end{aligned}$$

$$
\begin{aligned}
&\qquad\qquad + V_{n-2,f}(i_1-1, i_2-1, \phi_s^g(p) + \Phi(\phi_s^g(p)))\mu_1(\phi_s^g(p))\Big\} ds \\
&\quad + \int_\varepsilon^\infty e^{-(\alpha+\beta)s} \Big\{ V_{n-2,f}(i_1, i_2-1, \phi_\varepsilon^g(p))(\alpha-\mu_2) \\
&\qquad\qquad + V_{n-2,f}(i_1, i_2-2, \phi_\varepsilon^g(p))\mu_2 \Big\} ds \Big] \mu_2 \Bigg\} dt \\
= r(i_1,i_2) &+ \int_0^\varepsilon e^{-(\alpha+\beta)t} \Big(r(i_1,i_2)\alpha - \frac{c_2\mu_2}{\alpha+\beta} \Big) dt + \int_\varepsilon^\infty e^{-(\alpha+\beta)t} \Big(r(i_1,i_2)\alpha - \frac{c_2\mu_2}{\alpha+\beta} \Big) dt \\
&+ \int_0^\infty e^{-(\alpha+\beta)t} \Bigg\{ \int_0^\varepsilon e^{-(\alpha+\beta)s} \Bigg(\alpha \Big[V_{n-2,f}(i_1,i_2,\phi_s^g(p))(\alpha-\mu_2) \\
&\qquad\qquad + V_{n-2,f}(i_1,i_2-1,\phi_s^g(p))\mu_2 \Big] \\
&\qquad - \mu_1(\phi_s^g(p)) \Big[V_{n-2,f}(i_1,i_2,\phi_s^g(p)) \\
&\qquad\qquad - V_{n-2,f}(i_1-1,i_2,\phi_s^g(p)+\Phi(\phi_s^g(p))) \Big] (\alpha-\mu_2) \\
&\qquad - \mu_1(\phi_s^g(p)) \Big[V_{n-2,f}(i_1,i_2-1,\phi_s^g(p)) \\
&\qquad\qquad - V_{n-2,f}(i_1-1,i_2-1,\phi_s^g(p)+\Phi(\phi_s^g(p))) \Big] \mu_2 \Bigg) ds \Bigg\} dt \\
&+ \int_0^\infty e^{-(\alpha+\beta)t} \Bigg\{ \int_\varepsilon^\infty e^{-(\alpha+\beta)s} \Bigg(\alpha \Big[V_{n-2,f}(i_1,i_2,\phi_\varepsilon^g(p))(\alpha-\mu_2) \\
&\qquad\qquad + V_{n-2,f}(i_1,i_2-1,\phi_\varepsilon^g(p))\mu_2 \Big] \\
&\qquad - \mu_2 \Big[V_{n-2,f}(i_1,i_2,\phi_\varepsilon^g(p)) \\
&\qquad\qquad - V_{n-2,f}(i_1,i_2-1,\phi_\varepsilon^g(p)) \Big] (\alpha-\mu_2) \\
&\qquad - \mu_2 \Big[V_{n-2,f}(i_1,i_2-1,\phi_\varepsilon^g(p)) \\
&\qquad\qquad - V_{n-2,f}(i_1,i_2-2,\phi_\varepsilon^g(p)) \Big] \mu_2 \Bigg) ds \Bigg\} dt.
\end{aligned}
$$

In a complete analogous way we get

$$
\begin{aligned}
& V_{n,\tilde{\pi}}(i_1,i_2,p) = \mathcal{T}_g V_{n-1,ff}(i_1,i_2,p) \\
= \; & r(i_1,i_2) \\
& + \int_0^\varepsilon e^{-(\alpha+\beta)t} \Big\{ V_{n-1,ff}(i_1,i_2,\phi_t^g(p))(\alpha-\mu_1(\phi_t^g(p))) \\
& \qquad\qquad + V_{n-1,ff}(i_1-1,i_2,\phi_t^g(p)+\Phi(\phi_t^g(p)))\mu_1(\phi_t^g(p)) \Big\} dt
\end{aligned}
$$

$$
\begin{aligned}
&+\int_\varepsilon^\infty e^{-(\alpha+\beta)t}\Big\{V_{n-1,ff}(i_1,i_2,\phi_\varepsilon^g(p))(\alpha-\mu_2)+V_{n-1,ff}(i_1,i_2-1,\phi_\varepsilon^g(p))\mu_2\Big\}dt\\
=\;& r(i_1,i_2)+\int_0^\varepsilon e^{-(\alpha+\beta)t}\Big(r(i_1,i_2)(\alpha-\mu_1(\phi_t^g(p)))+r(i_1-1,i_2)\mu_1(\phi_t^g(p))\Big)dt\\
&+\int_\varepsilon^\infty e^{-(\alpha+\beta)t}\Big(r(i_1,i_2)(\alpha-\mu_2)+r(i_1,i_2-1)\mu_2\Big)dt\\
&+\int_0^\varepsilon e^{-(\alpha+\beta)t}\Bigg\{\int_0^\infty e^{-(\alpha+\beta)s}(\alpha-\mu_1(\phi_t^g(p)))\cdot\\
&\qquad\cdot\Big[V_{n-2,f}(i_1,i_2,\phi_t^g(p))(\alpha-\mu_2)+V_{n-2,f}(i_1,i_2-1,\phi_t^g(p))\mu_2\Big]ds\\
&\qquad+\int_0^\infty e^{-(\alpha+\beta)s}\mu_1(\phi_t^g(p))\Big[V_{n-2,f}(i_1-1,i_2,\phi_t^g(p)+\Phi(\phi_t^g(p)))(\alpha-\mu_2)\\
&\qquad\qquad+V_{n-2,f}(i_1-1,i_2-1,\phi_t^g(p)+\Phi(\phi_t^g(p)))\mu_2\Big]ds\Bigg\}dt\\
&+\int_\varepsilon^\infty e^{-(\alpha+\beta)t}\Bigg\{\int_0^\infty e^{-(\alpha+\beta)s}(\alpha-\mu_2)\Big[V_{n-2,f}(i_1,i_2,\phi_\varepsilon^g(p))(\alpha-\mu_2)\\
&\qquad\qquad+V_{n-2,f}(i_1,i_2-1,\phi_\varepsilon^g(p))\mu_2\Big]ds\\
&\qquad+\int_0^\infty e^{-(\alpha+\beta)s}\mu_2\Big[V_{n-2,f}(i_1,i_2-1,\phi_\varepsilon^g(p))(\alpha-\mu_2)\\
&\qquad\qquad+V_{n-2,f}(i_1,i_2-2,\phi_\varepsilon^g(p))\mu_2\Big]ds\Bigg\}dt\\
=\;& r(i_1,i_2)\\
&+\int_0^\varepsilon e^{-(\alpha+\beta)t}\Big(r(i_1,i_2)\alpha-\frac{c_1\mu_1(\phi_t^g(p))}{\alpha+\beta}\Big)dt+\int_\varepsilon^\infty e^{-(\alpha+\beta)t}\Big(r(i_1,i_2)\alpha-\frac{c_2\mu_2}{\alpha+\beta}\Big)dt\\
&+\int_0^\varepsilon e^{-(\alpha+\beta)t}\Bigg\{\int_0^\infty e^{-(\alpha+\beta)s}\Bigg(\alpha\Big[V_{n-2,f}(i_1,i_2,\phi_t^g(p))(\alpha-\mu_2)\\
&\qquad\qquad+V_{n-2,f}(i_1,i_2-1,\phi_t^g(p))\mu_2\Big]\\
&\qquad-\mu_1(\phi_t^g(p))\Big[V_{n-2,f}(i_1,i_2,\phi_t^g(p))\\
&\qquad\qquad-V_{n-2,f}(i_1-1,i_2,\phi_t^g(p)+\Phi(\phi_t^g(p)))\Big](\alpha-\mu_2)\\
&\qquad-\mu_1(\phi_t^g(p))\Big[V_{n-2,f}(i_1,i_2-1,\phi_t^g(p))\\
&\qquad\qquad-V_{n-2,f}(i_1-1,i_2-1,\phi_t^g(p)+\Phi(\phi_t^g(p)))\Big]\mu_2\Bigg)ds\Bigg\}dt
\end{aligned}
$$

$$
\begin{aligned}
&+\int_{\varepsilon}^{\infty} e^{-(\alpha+\beta)t}\Bigg\{\int_{0}^{\infty} e^{-(\alpha+\beta)s}\Bigg(\alpha\Big[V_{n-2,f}(i_1,i_2,\phi_{\varepsilon}^{g}(p))(\alpha-\mu_2)\\
&\qquad\qquad +V_{n-2,f}(i_1,i_2-1,\phi_{\varepsilon}^{g}(p))\mu_2\Big]\\
&\quad -\mu_2\Big[V_{n-2,f}(i_1,i_2,\phi_{\varepsilon}^{g}(p))\\
&\qquad\qquad -V_{n-2,f}(i_1,i_2-1,\phi_{\varepsilon}^{g}(p))\Big](\alpha-\mu_2)\\
&\quad -\mu_2\Big[V_{n-2,f}(i_1,i_2-1,\phi_{\varepsilon}^{g}(p))\\
&\qquad\qquad -V_{n-2,f}(i_1,i_2-2,\phi_{\varepsilon}^{g}(p))\Big]\mu_2\Bigg)ds\Bigg\}dt.
\end{aligned}
$$

Consider then as in (5.1) the difference of $V_{n,\tilde{\pi}}(i_1,i_2,p)$ and $V_{n,\pi}(i_1,i_2,p)$:

$$
V_{n,\tilde{\pi}}(i_1,i_2,p)-V_{n,\pi}(i_1,i_2,p)=\int_{0}^{\varepsilon} e^{-(\alpha+\beta)t}\left(\frac{c_2\mu_2-c_1\mu_1(\phi_t^g(p))}{\alpha+\beta}\right)dt\geq 0. \tag{B.1}
$$

This inequality is true, since $c_1\mu_1(\phi_t^g(p))\leq c_1\mu_1(\mathcal{M}^{i_1-1}p)<c_2\mu_2$. For this assertion we have to use $\phi_t^g(p)\geq p\geq \mathcal{M}p$ (see lemma 5.10) and $\mu_1^A<\mu_1^B$ by (5.10).

We now have to prove that $((f_t)_{t\geq 0},\ldots,(f_t)_{t\geq 0})\in F^n$ is optimal for all n and then we know that $\big(\lim_{n\to\infty}(f_t)_{t\geq 0}\big)^{\infty}=\big((f_t)_{t\geq 0}\big)^{\infty}$ is optimal for the infinite horizon MDP. Starting with $n=0$ and $V_0(i_1,i_2,p)=0$ each decision rule f is a minimizer of $V_0(i_1,i_2,p)$. Assume now that $(f_t)_{t\geq 0}$ is a minimizer of $V_0(i_1,i_2,p),\ldots,V_{n-1}(i_1,i_2,p)$ for $p>p^{>}(i_1)$, thus $((f_t)_{t\geq 0},\ldots,(f_t)_{t\geq 0})\in F^n$ is optimal for the n-period model in (i_1,i_2,p). We have to show that $(f_t)_{t\geq 0}$ is a minimizer of $V_n(i_1,i_2,p)$ for $p>p^{>}(i_1)$ again.

$$
\begin{aligned}
&\mathcal{T}_g V_n(i_1,i_2,p)\\
=\;&\int_{0}^{\infty} e^{-(\alpha+\beta)t}\Bigg\{(c_1i_1+c_2i_2)\\
&\qquad+\int_{0}^{\varepsilon} e^{-(\alpha+\beta)s}\Big\{V_n(i_1,i_2,\phi_s^g(p))(\alpha-\mu_1(\phi_s^g(p)))\\
&\qquad\qquad +V_n(i_1-1,i_2,\phi_s^g(p)+\Phi(\phi_s^g(p)))\mu_1(\phi_s^g(p))\Big\}ds\\
&\qquad+\int_{\varepsilon}^{\infty} e^{-(\alpha+\beta)s}\Big\{V_n(i_1,i_2,\phi_{\varepsilon}^g(p))(\alpha-\mu_2)\\
&\qquad\qquad +V_n(i_1,i_2-1,\phi_{\varepsilon}^g(p))\mu_2\Big\}ds\Bigg\}dt\\
\overset{(*)}{=}\;&\int_{0}^{\infty} e^{-(\alpha+\beta)t}\Bigg\{(c_1i_1+c_2i_2)
\end{aligned}
$$

$$
\begin{aligned}
&+\int_0^{\varepsilon} e^{-(\alpha+\beta)s}\big\{V_{n,(f,\dots,f)}(i_1,i_2,\phi_s^g(p))(\alpha-\mu_1(\phi_s^g(p)))\\
&\qquad +V_{n,(f,\dots,f)}(i_1-1,i_2,\phi_s^g(p)+\Phi(\phi_s^g(p)))\mu_1(\phi_s^g(p))\big\}ds\\
&+\int_{\varepsilon}^{\infty} e^{-(\alpha+\beta)s}\big\{V_{n,(f,\dots,f)}(i_1,i_2,\phi_\varepsilon^g(p))(\alpha-\mu_2)\\
&\qquad +V_{n,(f,\dots,f)}(i_1,i_2-1,\phi_\varepsilon^g(p))\mu_2\big\}ds\Big\}dt
\end{aligned}
$$

$$
\begin{aligned}
&= \mathcal{T}_g V_{n,(f,\dots,f)}(i_1,i_2,p) = V_{n+1,(g,f,\dots,f)}(i_1,i_2,p) \overset{(B.1)}{\geq} V_{n+1,(f,g,f,\dots,f)}(i_1,i_2,p)\\
&= \int_0^{\infty} e^{-(\alpha+\beta)t}\Big\{(c_1 i_1+c_2 i_2)\\
&\qquad +\int_0^{\infty} e^{-(\alpha+\beta)s}\big\{V_{n,(g,f,\dots,f)}(i_1,i_2,\underbrace{\phi_s^f(p)}_{=p})(\alpha-\mu_2)\\
&\qquad\qquad +V_{n,(g,f,\dots,f)}(i_1,i_2,\underbrace{\phi_s^f(p)}_{=p})\mu_2\big\}ds\Big\}dt\\
&= \mathcal{T}_f V_{n,(g,f,\dots,f)}(i_1,i_2,p) \geq \mathcal{T}_f V_n(i_1,i_2,p).
\end{aligned}
$$

Thus (f_t) is a minimizer of $V_n(i_1,i_2,p)$. In $(*)$ we used

$$V_n(i_1,i_2,\phi_s^g(p)) = V_{n,(f,\dots,f)}(i_1,i_2,\phi_s^g(p))$$

which is true by induction hypotheses, since $\phi_s^g(p) \geq p$ and

$$
\begin{aligned}
V_n(i_1-1,i_2,\phi_s^g(p)+\Phi(\phi_s^g(p))) &= V_{n,(f,\dots,f)}(i_1-1,i_2,\phi_s^g(p)+\Phi(\phi_s^g(p)))\\
V_n(i_1,i_2-1,\phi_s^g(p)) &= V_{n,(f,\dots,f)}(i_1,i_2-1,\phi_s^g(p))
\end{aligned}
$$

by the recursion in the state space. The preconditions of this theorem are fulfilled for $n \in \mathbb{N}$ in states $(i_1-1,i_2,\phi_s^g(p)+\Phi(\phi_s^g(p)))$ and $(i_1,i_2-1,\phi_s^g(p))$ as well by lemma 5.28. □

Bibliography

Altman, E., Jimenez, T., Nunez-Queija, R., and Yechiali, U. (2003). Optimal Routing among ./M/1 Queues with Partial Information. *INRIA Research Report No. 4985*.

Altman, E., Marquez, R., and Yechiali, U. (2004). Admission and Routing Control with Partial Information and Limited Buffers. *Working Paper*.

Asmussen, S. (2003). *Applied Probabilities and Queues*. Springer-Verlag.

Bank, P. and ElKaroui, N. (2004). A Stochastic Representation Theorem with Applications to Optimization and Obstacle Problems. *The Annals of Probability*, 32 (18):1030–1067.

Bank, P. and Küchler, C. (2007). On Gittins' Index Theorem in Continuous Time. *Stochastic Processes and their Applications*, 117 (9):1357–1371.

Bellman, R. (1977). *Dynamic Programming*. Princeton University Press.

Bensoussan, A., Cakanyildirim, M., and Sethi, S. (2003). Partially Observed Inventory Systems. *Proceedings of the 44th IEEE Conference on Decision and Control*, pages 1023–1028.

Bertsekas, D. and Shreve, S. (1978). *Stochastic Optimal Control: The Discrete Time Case*. Academic Press.

Bhulai, S. (2002). *Markov Decision Processes - The Control of High-Dimensional Systems*. PhD-thesis, Vrije Universiteit Amsterdam.

Borisov, A. (2007). The Wonham Filter under Uncertainty: a Game-Theoretic Approach. *submitted to Proceedings of NET-COOP 2007*.

Borisov, A. and Stefanovich, A. (2005). Optimal Filtering for HMM Governed by Special Jump Processes. *Proceedings of the 44th Conference on Decision and Control*, pages 5935–5940.

Braun, M. (1993). *Differential Equations and Their Applications*. Springer-Verlag.

Brémaud, P. (1981). *Point Processes and Queues: Martingale Dynamics*. Springer-Verlag.

Ceci, C. and Gerardi, A. (2000). Filtering of a Markov Jump Process with Counting Observation. *Applied Mathematics and Optimization*, 42:1–18.

Clarke, F. (1983). *Optimization and Nonsmooth Analysis*. Wiley.

Davis, M. (1993). *Markov Models and Optimization*. Chapman & Hall.

Dempster, M. (1989). Optimal Control of Piecewise Deterministic Markov Processes. *Applied Stochastic Analysis*, 5:303–325.

Dempster, M. and Ye, J. (1995). Impulse Control of Piecewise Deterministic Markov Processes. *The Annals of Applied Probability*, 5 (2):399–423.

Donchev, D. (1998). On the Two-Armed Bandit Problem with Non-Observed Poissonian Switching of Arms. *Mathematical Methods of Operations Research*, 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. *Journal of Applied Mathematics and Stochastic Analysis*, 12 (2):151–160.

Donchev, D. and Yushkevich, A. (1996). Average Optimality in a Poisson Bandit with Switching Arms. *Mathematical Methods of Operations Research*, 45:265–280.

ElKaroui, N. and Karatzas, I. (1997). Synchronization and Optimization for Multi-Armed Bandit Problems in Continuous Time. *Computational and Applied Mathematics*, 16:117–152.

Elliott, R., Aggoun, R., and Moore, J. (1997). *Hidden Markov Models: Estimation and Control*. Springer-Verlag.

Fleming, W. and Rishel, R. (1975). *Deterministic and Stochastic Optimal Control*. Springer-Verlag.

Fleming, W. and Soner, H. (1993). *Controlled Markov Processes and Viscosity Solutions*. Springer-Verlag.

Forwick, L. (1997). *Optimale Kontrolle Stückweise Deterministischer Prozesse*. PhD-Thesis, Universität Bonn.

Framstad, N., Øksendal, B., and Sulem, A. (2004). Sufficient Stochastic Maximum Principle for the Optimal Control of Jump Diffusions and Applications to Finance. *Journal of Optimization Theory and Applications*, 121:77–98.

Gihman, I. and Skorohod, A. (1979). *Controlled Stochastic Processes*. Springer-Verlag.

Hanson, F. (2007). *Applied Stochastic Processes and Control for Jump Diffusions: Modeling, Analysis, and Computation*. Society for Industrial Mathematics.

Haussmann, U. (1986). *A Stochastic Maximum Principle for Optimal Control of Diffusions*. Longman Scientific & Technical.

Hernandez-Lerma, O. (1989). *Adaptive Markov Control Processes*. Springer-Verlag.

Hernandez-Lerma, O. and Lasserre, J. (1996). *Discrete-Time Markov Control Processes*. Springer-Verlag.

Honhon, D. and Seshadri, S. (2007). Admission Control with Incomplete Information to a Finite Buffer Queue. *Probability in the Engineering and Informational Sciences*, 21 (1):19–46.

Hordijk, A. and Koole, G. (1992). On the Shortest Queue Policy for the Tandem Parallel Queue. *Probability in the Engineering and Informational Sciences*, 6:63–79.

Karatzas, I. and Shreve, S. (2001). *Methods of Mathematical Finance*. Springer.

Kaspi, H. and Mandelbaum, A. (1995). Lévy Bandits: Multi-Armed Bandits Driven by Lévy Processes. *The Annals of Applied Probability*, 5 (2):541–565.

Kaspi, H. and Mandelbaum, A. (1998). Multi-Armed Bandits in Discrete and Continuous Time. *Stochastic Processes and their Applications*, 8 (4):1270–1290.

Kitaev, M. and Rykov, V. (1995). *Controlled Queueing Systems*. CRC-Press.

Koole, G. (1998). A Transformation Method for Stochastic Control Problems with Partial Information. *Systems and Control Letters*, 35:301–308.

Kushner, H. and Dupuis, P. (2001). *Numerical Methods for Stochastic Control Problems in Continuous Time*. Springer-Verlag.

Last, G. and Brandt, A. (1995). *Marked Point Processes on the Real Line*. Springer-Verlag.

Lin, K. and Ross, S. (2003). Admission Control with Incomplete Information of a Queueing System. *INFORMS, Operations Research*, 51:645–654.

Liptser, R. and Shiryayev, A. (2004a). *Statistics of Random Processes I: General Theory*. Springer-Verlag.

Liptser, R. and Shiryayev, A. (2004b). *Statistics of Random Processes II: Applications*. Springer-Verlag.

Massey, W. and Whitt, W. (1998). Uniform Acceleration Expansions for Markov Chains with Time-Varying Rates. *The Annals of Applied Probability*, 8 (4):1130–1155.

Miller, B., Avrachenkov, K., Stepanyan, K., and Miller, G. (2005). Flow Control as a Stochastic Optimal Control Problem with Incomplete Information. *Problems of Information Transmission*, 41 (2):150–170.

Øksendal, B. and Sulem, A. (2005). *Applied Stochastic Control of Jump Diffusions*. Springer-Verlag.

Presman, E. and Sonin, I. (1990). *Sequential Control with Incomplete Information*. Academic press.

Presman, K. (1990). Poisson Version of the Two-Armed Bandit Problem with Discounting. *Theory of Probability and Its Applications*, 35 (2):307–317.

Raman, A., DeHoratius, N., and Ton, Z. (2001). Execution: The Missing Link in Retail Operations. *California Management Review*, 43 (2):136–152.

Rishel, R. (1978). The Minimum Principle, Separation Principle, and Dynamic Programming for Partially Observed Jump Markov Processes. *IEEE Transactions on Automatic Control*, 23:1009–1014.

Rockafellar, R. (1996). *Convex Analysis*. Princeton University Press.

Rogers, L. and Williams, D. (2003). *Diffusions, Markov Processes and Martingales*. Cambridge University Press.

Winter, J. (2007). Finite horizon control problems under partial information. *Proceedings of NET-COOP 2007*, pages 120–128.

Wong, E. and Hajek, B. (1985). *Stochastic Processes in Engineering Systems*. Springer-Verlag.

Yong, J. and Zhou, X. (1999). *Stochastic Controls: Hamiltonian Systems and HJB Equations*. Springer-Verlag.

List of Tables

List of Figures

Zusammenfassung

Informationsbeschaffung und -verarbeitung werden zu immer wichtigeren Bestandteilen eines Entscheidungsprozesses. Durch die wachsende Komplexität, aber auch durch die wachsenden Möglichkeiten Informationen zu beschaffen, steigen entsprechend Beschaffungs- und Verarbeitungskosten (Zeitaufwand, Analyse, Vergleiche, ...) hierfür. Dadurch stehen Entscheidungsträger vor der Frage, welche Informationen für eine Entscheidung relevant sind bzw. wie hoch die erwartete Kostensteigerung beim Fehlen dieser Information ist. Folglich gilt es jeweils abzuwägen, ob die zusätzliche Information überhaupt einen (monetären) Vorteil und Nutzen bringt. Ein klassisches Beispiel aus dem alltäglichen Leben ist der Supermarkt. Um Zeit zu sparen wird an den Kassen nicht mehr jeder Joghurt separat gescannt, sondern es wird nur noch die Gesamtzahl der Joghurts gezählt und dann ein zufälliger Joghurt auf dem Band als Referenz über den Kassenscanner gezogen. Für den Preis des Kunden ist dies nicht entscheidend, aber das Lagerhaltungssystem, das mit dem Abrechnungssystem der Kassen gekoppelt ist, kann nun nicht mehr feststellen, wie viele Erdbeer- oder Pfirsichjoghurts im Laden vorhanden sind. Dadurch müssen für die Nachbestellung entweder alle Joghurts im Laden getrennt nach Sorten gezählt oder die Bestellentscheidung mit einem Informationsdefizit getroffen werden.

Dieses einfache, alltägliche Beispiel illustriert bereits die Problematik und lässt sich auf viele weitere reale Probleme übertragen. Neben dem Lagerhaltungsmodell seien hier nur Fragestellungen aus dem Bereich des Internets und der Telekommunikation, der Steuerung von Call-Centern oder dem Verhalten auf Finanzmärkten genannt. Die unvollständige Information kann dabei nach zwei verschiedenen Einflussarten klassifiziert werden. Die erste ist, dass ein Umweltprozess, der den eigentlichen Zustandsprozess beeinflusst (bspw. hängt das stochastische Verhalten des Zustandsprozesses von der Umwelt ab), nicht beobachtbar ist. Als Beispiel mag hier die weltweite Wirtschaftslage als Einfluss auf das Kreditrating eines einzelnen Unternehmens dienen. Die zweite Variante ist, dass der Zustandsprozess an sich nicht beobachtbar ist. In der Literatur wird bisher meist der Fall betrachtet, dass ein mit dem Zustandsprozess korrelierter (eindimensionaler) Prozess beobachtbar ist. In der Praxis dagegen können jedoch häufig nur Gruppen von Zuständen unterschieden werden. Dieser Fall ist in der Literatur bislang kaum untersucht worden.

Die vorliegende Arbeit schließt diese Lücke für Markov'sche Sprungprozesse. In **Kapitel 2** konstruieren wir zunächst den dreikomponentigen Zustandsprozess mit diskretem Zustandsraum. Die erste Komponente, der Umweltprozess, ist ein Markov'scher Sprungprozess, der auf zwei Arten den eigentlichen Zustandsprozess beeinflusst. Die erste ist, dass der Umweltprozess, den Generator, also die Stochastik, des Zustandsmarkovprozess bestimmt (Hidden-Markov-Model). Der zweite Einfluss ist, dass zugelassen wird, dass Änderungen im Zustand des Umweltprozess zu unmittelbaren Sprüngen im eigentlichen Zustandsprozess führen. Dies ist motiviert durch die Tatsache, dass sich bspw. im Falle einer Verschlechterung der wirtschaftlichen Lage (Umweltprozess) auch die Kreditwürdigkeit eines einzelnen Unternehmens (Zustandsprozess) verschlechtert. Die dritte Komponente des Prozesses wird durch die Informationsstruktur definiert, bei der wir annehmen, dass

nur Gruppen von Zuständen des Zustandsprozesses beobachtbar sind (vgl. Definition 2.1). Wir illustrieren diese Konstruktion anhand verschiedener Spezialfälle, die u.a. das Hidden-Markov-Model und eine 0-1-Beobachtung beinhalten. Anschließend führen wir unser Optimierungsproblem (P) ein. Bei diesem sollen die erwarteten diskontierten Kosten, die von dem Verhalten des dreikomponentigen gesteuerten Prozesses abhängen, über einen unendlichen Horizont minimiert werden. Als zulässige Steuerungen sind nur solche zugelassen, die auf den erhältlichen Beobachtungen beruhen. Dieses Optimierungsproblem ist nicht direkt lösbar, da der Zustandsprozess nicht vollständig beobachtbar ist.

In **Kapitel 3** definieren wir ein zu (P) äquivalentes Problem (P_{red}), welches eines unter vollständiger Information und dadurch direkt lösbar ist. Wir zeigen im Reduktionstheorem 3.13, dass die optimalen Steuerungen und die Optimalwerte der beiden Probleme dieselben sind. Zuvor berechnen wir in Theorem 3.5 eine explizite Darstellung der bedingten Wahrscheinlichkeit, dass Umwelt- und Zustandsprozess in einem Zustand sind unter den bisherigen Beobachtungen und diskutieren Eigenschaften dieser Filtergleichung (3.10). Dieser stückweise-deterministische Schätzprozess ersetzt im reduzierten Problem den unbekannten Umwelt- und Zustandsprozess. Analog können die erwarteten Kosten als Funktion des Schätzprozesses dargestellt werden. Abschließend diskutieren wir in diesem Kapitel die Abhängigkeit zwischen Informationsstruktur und dem Optimalwert und zeigen erste Eigenschaften der Wertfunktion wie bspw. die Konkavität im Schätzer.

Kapitel 4 befasst sich mit der Lösung des reduzierten Problems (P_{red}) unter vollständiger Information. Zunächst zeigen wir in Theorem 4.3, dass die Wertfunktion Lösung einer verallgemeinerten Hamilton-Jacobi-Bellman Gleichung ist. Die Verallgemeinerung besteht dabei darin, dass wir die Differenzierbarkeitsannahme auf differenzierbar im Clarke'schen Sinne abschwächen. Diese Annahme erfüllt die konkave Wertfunktion von (P_{red}). Wir beweisen zudem notwendige und hinreichende Bedingungen an die optimale Steuerung. Letzteres führt in Theorem 4.4 zur verallgemeinerten Verifikationsmethode. Ein zweiter Lösungsvorschlag ist die Formulierung eines zeitdiskreten Markov'schen Entscheidungsproblems (MDP). Hierbei schlagen wir Nutzen aus dem stückweise-deterministischen Verhalten des Schätzprozesses. Der Aktionenraum wird dabei durch einen Funktionenraum beschrieben. Theorem 4.7 beweist den Zusammenhang zwischen diesem MDP und (P_{red}). Insbesondere sind die Optimalwerte gleich und aus der optimalen Politik des MDP, deren Existenz wir in Theorem 4.14 diskutieren, lässt sich eine optimale Steuerung von (P_{red}) herleiten. Zudem zeigen wir den Zusammenhang der beiden Lösungsansätze in Theorem 4.9 auf.

Im abschließenden **Kapitel 5** betrachten wir ein Warteschlangenmodell. Ein Server kann dabei seine Servicekapazität auf zwei Warteschlangen aufteilen, wobei für jeden wartenden Kunden eine Kostenrate c_i, abhängig von Schlange i, anfällt. Die zufälligen Bedienzeiten hängen von der Bedienrate μ_i ab. Im Falle vollständiger Information ist bekannt, dass die $c\mu$-Regel optimal ist, d.h. es ist optimal, die Schlange zu bedienen, bei der $c_i\mu_i$ größer ist, sofern dort ein Kunde wartet. Wir zeigen zunächst, dass die $c\mu$-Regel ebenfalls optimal bleibt, sofern die Informationsstruktur hinreichend genügend fein ist. Anschließend analysieren wir den Fall, dass jedes μ_i zwei Werte μ_i^A bzw. μ_i^B annehmen kann

(Bayes'scher Fall), wobei dem Server nicht bekannt ist, welcher Wert angenommen wird. Wir leiten basierend auf den Grundlagen aus Kapitel 3 zunächst eine explizite Darstellung des Schätzprozesses her und diskutieren Eigenschaften des Schätzers. Danach definieren wir das (uniformisierte) zeitdiskrete Markov'sche Entscheidungsproblem im Sinne von Abschnitt 4.2. Damit beweisen wir für die Wertfunktion die Separationseigenschaft in Theorem 5.13.

Anschließend beweisen wir mit Hilfe der verallgemeinerten Verifikationsmethode aus Kapitel 4, dass die optimale Strategie fast immer eine Schlange exklusiv bedient. Für den symmetrischen Fall, d.h. $\mu_1^A = \mu_2^B$ und $\mu_1^B = \mu_2^A$, zeigen wir unter $c_1 = c_2 = 1$ die Optimalität einer Kontrollgrenzenregel mit Kontrollgrenze $p^* = \frac{1}{2}$. Im Fall, dass eine Bedienrate bekannt ist, beweisen wir, dass es stets optimal ist, nur eine Warteschlange zu bedienen, und geben hinreichende Bedingungen für die optimale Steuerung. Wie im symmetrischen Fall erhalten wir die aus der Banditen-Theorie bekannte stay-on-a-winner Eigenschaft für die optimale Steuerung. Diese Resultate werden durch Simulationsergebnisse veranschaulicht. Abschließend betrachten wir anstelle der Minimierung der Kosten ein Modell mit Gewinnfunktion für jeden bedienten Kunden. Dieses Problem lösen wir mit Hilfe eines Gittins-Index vollständig und ergänzen dabei Resultate aus der zeitstetigen Banditenprobleme-Theorie. Ist dagegen nicht beobachtbar, ob ein Kunde in einer Schlange wartet oder nicht, so ist eine geschlossene Lösungsformel für den Schätzprozess nicht verfügbar. In Abschnitt 5.3 vergleichen wir zwei plausible Steuerungen für eine solche 0-1-Informationsstruktur numerisch.

Danke

- Prof. Dr. Ulrich Rieder für Ihre Anregungen und Tipps, für die fruchtbaren Diskussionen mit Ihnen, für den enormen Rückhalt im Endspurt, für die Möglichkeit meine Dissertation an Ihrem Lehrstuhl zu schreiben und die Freiheiten in Forschung und Lehre, die mir selbstständiges und eigenverantwortliches Arbeiten ermöglichten
- Prof. Dr. Dieter Kalin für Ihr Interesse an meiner Arbeit und die Übernahme des Zweitgutachtens, sowie die gemeinsame Zeit am Institut
- Frank, Marc, Thomas, Harald und allen anderen Kollegen in der Fakultät für die gute Arbeitsatmosphäre, den Austausch und den Spaß außerhalb der Promotion
- meiner Familie für Eure Unterstützung und Euren Rückhalt
- Euch Freunden für das Leben außerhalb der Universität, was mir Abwechslung und immer wieder neuen Schwung für meine Promotion brachte
- Daniela für Deine Liebe, Geduld, Motivation und aufbauenden Ermunterungen