Karlsruher Institut für Technologie

Cost Optimal Control of Piecewise Deterministic Markov Processes under Partial Observation

Zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

von der Fakultät für Mathematik des
Karlsruher Instituts für Technologie (KIT)
genehmigte
Dissertation
von

Dipl.-Math. Dirk Klaus Lange
aus Stuttgart

Tag der mündlichen Prüfung:	15. Februar 2017
Referentin:	Prof. Dr. Nicole Bäuerle
Korreferent:	Prof. Dr. Günter Last

Gratefully dedicated to
my parents and to
Katrin.

Preface

This PhD thesis has been written during my employment as research and teaching assistant at the Institute for Stochastics at Karlsruhe Institute of Technology (KIT). I wish to thank several people at this place for various kinds of support.

My most sincere and deepest gratitude belongs to my advisor Prof. Dr. Nicole Bäuerle. She gave me the opportunity to „come back“ to research and suggested a topic that perfectly matched my interest for control theory of stochastic processes with applications to technical or Engineering problems. With her constant and always helpful guidance and supervision, she created a wonderful working atmosphere that was motivating, inspiring and demanding at the same time. Many fruitful discussions with her broadened my perspective on the topic of my thesis as well as on many other domains of stochastics.

Moreover, I want to thank Prof. Dr. Günter Last for being available as second referee for this thesis and for very helpful feedback on it.

I would like to thank the whole Institute for Stochastics for creating an excellent working atmosphere over the years. I thank Prof. Dr. Norbert Henze and Dr. Bruno Ebner for an inspiring collaboration when teaching together. Furthermore, I want to thank my colleagues for always being available and having an open ear in times of need. Special thanks go to Jan Weis and Dennis Müller for last minute help on graphics and simulations.

This thesis, however, would not have been possible if on a long way over many years, good friends and teachers had not been supporting me. The list would be too long to mention you all. We had great years at Ecole Polytechnique, where Pierre-Louis Lions brought me back to mathematics when I was almost leaving for Engineering. The years at Ecole Polytechnique truly enriched my life and I gratefully memorize „l'X“ as an awesome place to learn and work.

Finally, I want to thank the three most important persons in my life for year long support over many difficult projects and challenges. Without your support, this thesis and much more would never have been possible. Thank you, mum, dad and Katrin.

Dirk Lange Karlsruhe, January 2017

Contents

Chapter 0

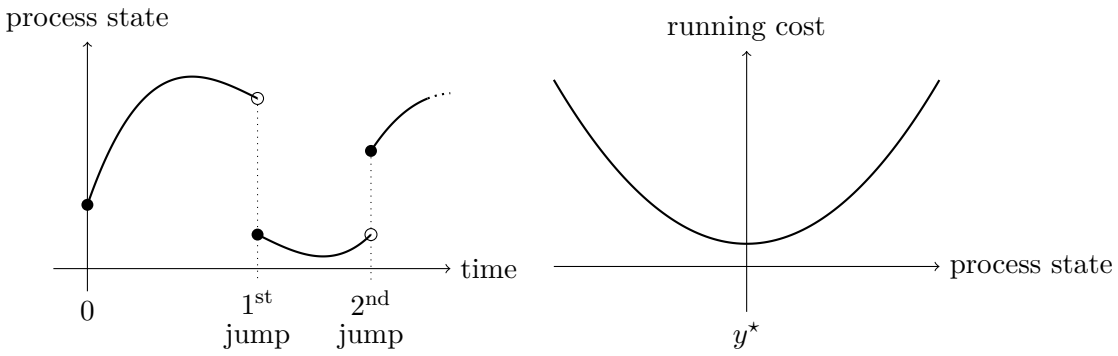
Introduction

Motivation and aims of this thesis

The aims of this thesis are (i) to develop a model for a *controlled Piecewise Deterministic Markov Process (PDMP) under Partial Observation (PO)*, (ii) to prove results on *existence of optimal control policies* that solve the optimization problem of *minimizing total expected discounted cost over lifetime* for a PO-PDMP and (iii) to *characterize optimal policies* for a concrete application example under convexity assumptions on the running cost function.

To make these aims more tangible for readers without mathematical background or for mathematicians without deeper knowledge in control theory for stochastic processes, we like to motivate these aims by an extremely simplified example. A more detailed mathematical motivation then follows in combination with the literature review in the following Section.

Imagine a process that is running in continuous time. One could think of some queueing example where a state of the process at a given point in time is the work load waiting to be processed in the queue. We assume the development of the process states over time to be a *deterministic* function of the time, i.e. knowing the current state of the process, we can calculate the future states of the process, e.g., by solving an initial value problem. This process becomes a *piecewise deterministic process* if we now add jumps to the path of this process while still assuming a deterministic behavior *piecewise* between these jumps.



(a) Process states over time

(b) Running cost as function of current state

Figure 1: Examples for path of a PDMP and for running cost function

This process becomes a *stochastic* process, by assuming that these jumps occur at random points in time. Let the probability distribution of these jump times governed by some known intensity and the state transition at a jump time described, in a very simple example, by some fixed probability distribution on the set of possible process states. See Figure 1(a) for an example of a typical path of such a process. In the queueing example, we would simply assume bulk arrivals (or drop outs) of randomly distributed work loads arriving (dropping out) at random times with known intensity of these events.

Imagine now, that there are agents who can control this process: By executing a control action at every point in continuous time, they can influence, say, the deterministic movement of the process states between two jump times. This could be by controlling the processing speed in the queueing example. If there is some running cost, created at every point in time depending on the current process state, then the challenge for the agents is: Control the process such that total expected discounted running cost over lifetime is minimal. See Figure 1(b) for an example of a running cost function with a unique minimum in the state labeled $y^\star$. Hence, in that situation, a canonical idea for an optimal control policy could be to „always steer the process as quickly as possible to state $y^\star$“.

However, if jumps from state $y^\star$ lead with very high probability to states with very high running cost and if bringing the process then back into state $y^\star$ is taking several hours, this policy might not be an optimal one. Even more complex situations arise, when the intensity, i.e. average expected number of jumps per unit of time, depends on the current state. In a queueing example, this might apply: Think of cash desk queues where fewer customers would join the queue the longer it is. If now, the agent can even control the intensity or the probability distribution that determines where jumps of states lead to, the situation becomes even more complex. This could happen in case the agents can, e.g., influence the demand for their products by means of advertisement.

This is a typical example of a control problem for a PDMP. A broad range of results for these models and associated optimization problems exist already, see following Section.

Imagine now, that the current state of the process cannot be observed. However, whenever a jump to a new state occurs, some noisy measurement right after a jump can be performed to get a noisy observation of this „post-jump state“. This is the setting of a *partially observable* PDMP: only partial, here noisy, information about the actual process state is available. Such situations can arise, e.g., when work load arriving at a queue can only be estimated.

General models of controlled PO-PDMPs have not been published yet. In this thesis, we develop such a model covering this and much more general applications of partially observable PDMPs. We formulate the associated optimization problem of minimizing total expected discounted cost over lifetime and derive results on existence of optimal policies. For some concrete cases with convex running cost functions we even derive uniqueness of such an optimal policy and determine it concretely.

Overview on control theory for Piecewise Deterministic Markov Processes

In [23], Davis introduced the class of *Piecewise Deterministic Markov Processes (PDMP)* as a „general class of non-diffusion models“. This statement is to be understood in the context of the following result of Çinlar and Jacod [18] from 1981: „Every strong Markov process with values in $\mathbb{R}^d$ and continuous paths of locally bounded total variation is deterministic.“ In that sense, a non-trivial strong Markov process with states in $\mathbb{R}^d$ either has continuous

paths and allows for locally infinite total variation or has paths with locally bounded total variation that are not continuous. The first case is leading to the theory of diffusion models. If for the second case, explosion of the jumps is excluded, the movement of the process between two jumps is necessarily deterministic, see [18]. This leads to the theory of PDMPs.

PDMPs are characterized by three local characteristics: The drift, describing the deterministic movement between two jumps of the process, the jump intensity, governing the density of the probability distribution of the inter-jump times as well as the jump transition kernel, the probability distribution on the set of possible post-jump states given the current state of the process right before the jump. A PDMP thus starts in an initial state to then follow the deterministic path defined by the drift up to the first jump time. Then, the process jumps to a new state. This transition is described by the jump transition kernel. The process then follows again the deterministic drift and this scheme repeats. For a rigorous development of the PDMP theory and a summary of the basic results on optimal control problems for PDMPs see the book of Davis [23] or even an early article of Yushkevich [64] from 1987, where PDMPs also were called „Markov decision deterministic drift processes“.

Classical optimization problems can be formulated for PDMPs such as reward maximization or cost minimization. Both versions, minimum expected average cost (see, e.g., [5], [19] or [20]) as well as minimum expected total discounted cost problems are classical for PDMP control problems. In [23], Davis shows existence of optimal control policies for a problem of minimum expected total discounted cost with bounded running cost and constant discount factor. These optimal policies are in general relaxed controls, i.e. a control action is a probability distribution on the action space. Yushkevich is only treating PDMP control problems with uncontrolled drift. However, the idea of reducing the continuous time control problem of a PDMP to a discrete time Markov Decision Process (MDP) is also due to Yushkevich, see [61]. Actually, as the movement of the process between two jumps is deterministic, a pure post-jump consideration is sufficient for the treatment of optimal control problems for PDMPs.

The range of possible applications of the general PDMP control theory is broad. There are applications in finance [57], communication networks [15] or [40], neurosciences [50] and biochemics [49] to only list a very short overview that illustrates the huge variety of domains of application for optimal control problems for PDMPs.

In terms of pure mathematical treatment of PDMP control problems, the status up to 1993 can be found in [23]. Since then, important steps in the further development of this theory were, amongst others: In [21], Costa and Raymundo consider impulse control of PDMPs without continuity or differentiability assumptions on the state. In [1], the control problem in continuous time is reduced to a problem in discrete time while working under even lower regularity assumptions. General conditions such as semi-analytic value functions or universally measurable selectors are applied. Forwick [35] then considers, in contrast to the earlier work of Davis (see above), problems with only locally bounded running cost functions. He shows absolute continuity for the value function and that the value function is a (weak) solution of the Hamilton-Jacobi-Bellmann equation. In addition, he derives sufficient conditions for the existence of optimal deterministic feedback controls.

Later, with [25] and [13], a few new results on numerical methods for optimal stopping problems for PDMPs appeared. In both works, the embedded process of the underlying PDMP is discretized by quantization. Runggaldier's work [52] gave here inspiration for the quantization approach. A further discretization of the time between jumps is then necessary to determine the optimal stopping time, which is not necessarily a jump time of the process. The remaining problem is treated by solving a discrete time version of the

Bellmann equation. Analytical approaches to optimal stopping for PDMPs were published even before, e.g., by Gugerli [37].

Remarkable about the paper of Brandejski et al. [13] is, however, that they treat an optimal stopping problem for a PDMP under partial observation. There are only very few works treating PDMP control problems under partial observation. In [46], a special convex hedging problem on a financial market with price processes with geometric Poisson-distribution is considered. In the second part of this work, partial observation is modeled by assuming to have an unknown jump intensity. In [7], a problem of optimal inventory management is considered. Here, partial observation is modeled by assuming censored observations.

General works on PDMP control problems under partial observation do not exist yet. For their stopping problem, Brandejski et al. suggest to model partial observation by assuming perfect observation of inter-jump times but only noisy measurement of the post-jump state of the PDMP which for other times then jump times, is assumed completely unobservable. Stopping, however is a very special control problem with only two control actions: „stop“ or „continue“.

The model for partial observation introduced by Brandejski et al. is now the base for this thesis, where the first aim is to define a general model of a controlled PDMP under partial observation. Compact and metric action spaces shall be allowed for this model, as this is typical for applications of completely observable PDMP control problems as well. The approach for then solving a problem of minimal expected total discounted cost in this model combines techniques from both: The theory of controlled PDMPs under complete observation as well as the theory of Markov Decision Processes (MDP) under partial observation. A very good summary of the latter, together with applications to finance, provides the book of Bäuerle and Rieder [6]. For further references on results of the MDP theory, we refer to the introduction of Section 2.1.

Main results and outline of this thesis

There are three main results of this thesis and each of them is strongly connected to one central equation of this thesis. Having these three equations in mind when reading through this thesis shall provide a good guidance in order to understand the outline and all intermediate results of this thesis. However, it has to be remarked clearly here: It is not the following three equations that represent the main results of this thesis but they are good starting points to explain these.

The first main result of this thesis is the *definition of a model* for a controlled *Partially Observable Piecewise Deterministic Markov Process (PO-PDMP).*

Models of controlled Piecewise Deterministic Markov Processes (PDMP) under complete observation already exist and have been analyzed and discussed by many authors, see previous Section. General models of controlled PDMPs under partial observation, however, have not been studied yet. The work of Brandejski et al. [13] was taken as a starting point to formulate such a general control model for a PO-PDMP. The resulting optimization problem of minimizing total expected discounted (with discount rate β) cost c over lifetime is defined in Definition 1.37. The value function for this optimization problem is the first equation a reader should try to understand when reading this thesis. It is the function providing, for each initial observation x of a PO-PDMP, the minimal

expected total discounted cost over lifetime:

$$J(x) := \inf_{\pi \in \Pi^P} \mathbb{E}_x^\pi \left[\int_0^{T_\infty} e^{-\beta t} \int_A c(X_t, Y_t, a)\ \pi_t(da)\ dt \right]. \tag{1}$$

Based on this equation, the challenge to define a general control model for a PO-PDMP can be described: We consider a model where jump times $T_1, T_2, \dots$ of the underlying PDMP can be observed perfectly but observation of the unobservable states Y_t of the underlying PDMP is only possible by noisy measurements X_{T_n} of the post-jump states Y_{T_n}. On an interval $[T_n, T_{n+1})$ we keep $X_t = X_{T_n}$ constant as no new observation takes place. Admissible control actions are then given by the set of so-called *relaxed controls*: At each point in time, an admissible control action is a probability measure on a compact action space A. The well-established theory for optimal control of completely observable PDMPs showed that in general, optimal controls do not exist in the set of so-called *deterministic* controls, i.e. executing an element of the compact action space directly instead of taking a probability measure on this space. From the theory of partially observable control problems for other classes of processes, it is known that, in general, decision rules have to be in the set of *history dependent* decision rules, denoted by Π^P above. This means, the decision of what relaxed control to execute at a given point in time has to depend on the history of observed process states X_t and earlier executed control actions up to this point in time.

The challenge for the formulation of a control model for a PO-PDMP was thus to combine relaxed controls and history dependent decision rules in a way such that the resulting controlled process again, has the form of a PDMP. This result can be found in Definition 1.34.

The second main result of this thesis is the developement of a so-called *filter*: A recursive calculation of the conditional probability distribution of the unobservable state of the underlying PDMP given the history of the observable process states as well as of the executed control actions.

Once a control model for a PO-PDMP and an associated optimization problem is defined, the principal approach to show existence of optimal policies is clear: One has to show that the initial optimization problem under partial observation in continuous time is equivalent to an optimization problem for a completely observable MDP, which is a process in discrete time. The key step for this reformulation of the problem is the development of an adequate filter. Such a filter has not yet been developed for a controlled PO-PDMP. It was achieved to develop such a filter for the presented PO-PDMP control model. The development of this filter was inspired from [13], where a filter for an uncontrolled PO-PDMP is developed. The following assumption has been taken for the development of this filter: The set of possible post-jump states of the controlled PO-PDMP is assumed to be of finite cardinality. In view of necessary quantizations of the filter required for numerical methods applicable to determine or to approximate optimal policies, this assumption is no significant restriction.

In Chapter 2 we present all necessary steps and intermediate results for the above mentioned reformulation of the initial optimization problem. The proof of the filter equation of Proposition 2.31 is the central result of this chapter. Based on this filter, the well-known techniques for MDPs can be applied to show that the value function of the optimization problem is a fixed point of the minimum cost operator. This fixed point equation is the

second central equation to have in mind when reading through this thesis:

$$J'(x,\rho) = \inf_{r\in\mathcal{R}} \left\{ g'(x,\rho,r) + \int_{E'} e^{-\beta s'} J'(x',\rho')\ q'_{SXM}(ds',dx',d\rho' \mid \rho,r) \right\}. \tag{2}$$

Even though the filter is only entering this equation implicitly via the definition of the measure appearing in the integral above, this equation could not have been proven without the development of the filter. It is from this fixed point equation later, that we derive existence of optimal policies for the initial optimization problem. We start from this equation as well, when discussing a concrete application of the theory in Chapter 5, where we derive existence of a unique optimal policy and characterize it as of „bang-bang“ type.

The third main result of this thesis is the discussion, including proof of existence of optimal policies, of a *second model of partial observation for a PDMP*: In this model, the *inter-jump time is no longer observable*. Existence of optimal policies can be shown even for models with controlled jump intensity and controlled jump transition kernel.

Based on the above mentioned filter, a sufficient condition for existence of optimal policies could be derived. This sufficient condition contains, amongst others, the assumption of having a PO-PDMP model with *uncontrolled* jump intensity and *uncontrolled* jump transition kernel. The restriction to such models was necessary, as it turned out that the filter developed is not continuous in the argument representing the executed relaxed control action. This is due to properties of the Young topology which is the topology selected for the space of relaxed controls. A detailed discussion of this topology can be found in the Annex of this thesis, see A. In Section 3.2.2, we explain why the filter is not continuous in the mentioned argument. This leads to the third equation, a reader should bear in mind when reading this thesis: The filter equation of Proposition 2.31, and more precisely the following part of this equation:

$$\chi_i^j(\mu,s,x,r) := \mu^i \exp\left(-\Lambda^r(y^i,s)\right) \int_A \lambda^A\left(\Phi^r(y^i,s),a\right) Q^A\left(\Phi^r(y^i,s),a;\{y^j\}\right)\ r_s(da) f_\epsilon(x-\psi(y^j)). \tag{3}$$

A reader familiar with characterizations of convergence in the Young topology will recognize that in the above equation, there is no integral w.r.t. the time parameter s.

For deeper reasons coming from the classical theory of completely observable PDMP control problems, we have to stick with this Young topology on the space of relaxed controls. The latter space is compact under this topology and in order to apply selection theorems for measurable optimizers, we need this compactness property.

An analysis of this continuity issue of the filter brought up the idea that for models with unobservable inter-jump time or even inter-jump times observed under noisy measurement, a filter with sufficient continuity properties should exist. Actually, in such a model, an integral w.r.t. the time parameter would enter the equation above. In Chapter 4, we thus study a model of a PO-PDMP with unobservable inter-jump time. Leveraging a very recent result of Feinberg [33] from 2016, it was achieved, to show existence of optimal policies even for *controlled* jump intensities and *controlled* jump transition kernels. In a sense, we were able to show that the PO-PDMP model of Chapter 4 can be seen as an application example of Feinberg's theory for MDPs under partial observation. The importance of Feinberg's result for the MDP theory lies in the fact that he achieved to provide a sufficient condition for the existence of filters that are „sufficiently continuous“ to guarantee a weakly continuous transition kernel for the filtered MDP model arising from a partially observable MDP model.

Further results and **outline** of this thesis:

In **Chapter 1**, we develop step by step the model of a controlled PO-PDMP. We introduce properly the underlying PDMP in Section 1.1, present how we model partial observation in Section 1.2 and define the control model in Section 1.3. The central results are then Definition 1.34 of a controlled PO-PDMP and Definition 1.37 of the initial optimization problem.

Chapter 2 is at the core of this thesis. Here we combine techniques from the control theory for PDMPs with approaches from the theory of partially observable MDPs. We first reformulate the initial optimization problem in continuous time into an equivalent optimization problem for a partially observable MDP, thus into a discrete time problem in Section 2.1. On that way, the Correpondence Theorem 2.11 for history dependent policies is an important result. In Section 2.2 we then reformulate the optimization problem for a partially observable MDP into an equivalent optimization problem for a completely observable MDP. This Section also contains the main result of this Chapter, the development of the above mentioned filter. We also explain why we can restrict the set of admissible policies to the set of Markov policies for the completely observable problem. In Section 2.3, we finally apply the well-known techniques from stochastic dynamic programming in order to derive existence of optimal policies for the filtered model, thus by equivalence, for the PO-PDMP control problem. A fixed point equation for the value function of the optimization problem is derived and under some (lower semi-) continuity assumptions on the filtered model, existence of one step optimizers is proven. Existence of optimal policies for the initial optimization problem is then derived for two time horizons: for an infinite time horizon as well as for a finite time horizon up to the N-th jump of the PO-PDMP.

In **Chapter 3**, we then translate the (lower semi-) continuity assumptions taken on the filtered model in order to derive existence of optimal polices into sufficient conditions on the initial PO-PDMP. It turns out that if we assume a continuous dependence of the drift on the relaxed control action executed, a lower semi-continuous cost function, an uncontrolled jump intensity and an uncontrolled jump transition kernel for the initial PO-PDMP, optimal policies exist. In Section 3.2.2, we also discuss the above mentioned continuity issue of the filter.

Chapter 4 is then dedicated to the discussion of another PO-PDMP model: the above mentioned model with unobservable inter-jump time. Based on Feinberg's result cited above, we can derive sufficient conditions for the existence of optimal policies also for models with controlled jump intensity and jump transition kernel.

In **Chapter 5**, we discuss a concrete application example with strictly convex cost function, constant jump instensity and uncontrolled jump transition kernel. We briefly motivate this example in Section 5.1, before we present the concrete mathematical model in Section 5.2. The fact that the completely observable version of this example is contained in the general partially observable version of the PO-PDMP model for this example is highlighted in Section 5.3. Uniqueness of an optimal policy and its characterization as „bang-bang"-type policy is then derived in Section 5.4 for the completely observable example and in Section 5.5 for the partially observable example.

The model assumptions taken throughout this thesis are summarized in **Annex C** and discussed in view of concrete applications ins **Chapter 6**, where we also discuss possible refinements and extensions of the PO-PDMP control theory presented in this thesis.

A detailed discussion of the Young topology can be found in **Annex A**.

Chapter 1

The PO-PDMP model and the optimization problem to solve

This first chapter is twofold: It is meant to familiarize the reader with the optimization problem we want to solve in the course of this thesis. Thus, it could be understood as a purely introductory chapter. However, a first result of this work is the definition, as such, of a *controlled Piecewise Deterministic Markov Process (PDMP) under Partial Observation (PO)*. To the best of our knowledge, there has not yet been a publication providing a general definition of a controlled PDMP under PO (also referred to as PO-PDMP in the sequel). Brandejski et al. were among the first to investigate optimal stopping problems for PO-PDMPs in [13]. Stopping, however, is a very simple control action and thus, can be understood as a particular case of the general control model we will introduce in this chapter. The general control model incorporating the space of observable histories shall thus be seen as a first result of this thesis.

In order to make this thesis accessible for readers not familiar with PDMPs in general, we start this chapter with a very brief introduction to the class of Piecewise Deterministic Markov Processes in Section 1.1. Readers familiar with PDMPs under complete observation might want to skip that section. As a next step, in Section 1.2, we introduce how partial observation of a PDMP shall be modeled in this thesis. There is a variety of possible approaches to model partial observation of PDMPs. We will briefly give an overview but then stick, for the rest of this thesis, to the partial observation model introduced in Section 1.2. We do so, as we believe that this model is closest to a huge set of applications. Familiar with the latter aspect of the model, we can introduce, in Section 1.3, the way how the underlying PDMP can be controlled based on partial observation.

Finally, we will end this chapter by stating the optimization problem of *minimizing the total discounted cost over lifetime* for a controlled PO-PDMP in Section 1.4.

1.1 The general PDMP model

In [22], Davis introduced the class of PDMPs as „general class of non-diffusion stochastic models“. He gave a definition of a PDMP based on its infinitesimal generator, and thus, strongly emphasizing the fact that a PDMP is a priori a time-continuous process. Recent publications such as [13] or [34] introduce a PDMP following an axiomatic approach[1] stating a set of properties of a PDMP. We will follow the latter approach in this thesis, for three reasons that are: First, we believe this approach is easier to follow for readers

[1] clearly, leading to an equivalent definition of a PDMP

not familiar with this class of stochastic processes. Second, this definition very obviously reveals parallels to the general theory of discrete-time Markov decision processes (MDP) that we aim to apply here. Third, the domain of definition of the infinitesimal generator associated with a PDMP is in most cases difficult to characterize.

We first define an uncontrolled PDMP before we show in Section 1.1.2, that this stochastic process in continuous time can in fact be fully described by a marked point process in discrete time. Clearly, one could also take the inverse perspective and start with a marked point process in discrete time and, in a second step, define how a PDMP is constructed out of it. This perspective is taken by Jacobsen in section 7.3 of his book [44] which also provides a very good introduction to Piecewise Deterministic Processes (PDP) in general, not only to the subclass of the Markovian PDPs.

1.1.1 Defining a PDMP: A stochastic process in continuous time

As later in this work, we will work with three different processes $(Y_t), (X_t)$ and (S_t), we start from now on to use sub- and sometimes superscripts for state spaces, intensities, densities etc. to inidcate the underlying processs. The following definition follows mainly the approach of Forwick (see [34]) with the difference that we do not restrict our definition to drifts defined by initial value problems.

Definition 1.1 (uncontrolled PDMP). *Let E_Y a Polish space*[2]*. A* piecewise deterministic Markov process (PDMP) $(Y_t)_{t\geq 0}$ *with* local characteristics (Φ_Y, λ_Y, Q_Y) *and* state space E_Y *is a stochastic process in continuous time that satisfies the following properties (i)-(ix):*

(ia) The drift $\Phi_Y : E_Y \times \mathbb{R}^+ \to E_Y$ *is continuous and fulfills*

(ib) The mapping $t \mapsto \Phi_Y(\cdot, t)$ is a semi-group w.r.t. concatenation of mappings, i.e. for all $y \in E_y$:

$$\Phi_Y(y, t+s) = (\Phi_Y(\cdot, t) \circ \Phi_Y(\cdot, s))(y) = \Phi_Y(\Phi_Y(y, s), t).$$

(ii) The jump rate *or* intensity $\lambda_Y : E_Y \to (0, \infty)$ *is a measurable mapping.*

(iii) $Q_Y : E_Y \to \mathbb{P}(E_Y)$ is a transition kernel.

In addition to properties (i)-(iii), there exists a measurable space $(\Omega, \mathcal{F})$, a family $(\mathbb{P}_y)_{y\in E_Y}$ of probability measures on $(\Omega, \mathcal{F})$ and an isotonic sequence $T_0, T_1, \ldots$ of $\overline{\mathbb{R}}$-valued random variables with the following properties:

(iv) The mapping $\mathbb{P}_\bullet(B) : E_Y \to [0,1], y \mapsto \mathbb{P}_y(B)$ is measurable $\forall\ B \in \mathcal{F}$.

(v) $\forall\ y \in E_Y :\ \mathbb{P}_y(Y_0 = y) = \mathbb{P}_y(T_0 = 0) = 1$.

(vi) $T_n \uparrow \infty$ $(n \to \infty)$ $\mathbb{P}_y$-a.s. $\forall y \in E_Y$.

(vii) $\forall y \in E_Y, t \geq 0, n \in \mathbb{N}_0$:

$$\mathbb{P}_y(T_{n+1} - T_n > t | T_0, Y_{T_0}, \ldots, T_n, Y_{T_n}) = \exp\{-\int_0^t \lambda_Y(\Phi_Y(Y_{T_n}, s))\ ds\}.$$

[2] i.e. E_Y is endowed with a topology $\mathcal{T}_{E_Y}$ such that the topological space $(E_Y, \mathcal{T}_{E_Y})$ is complete, separable and metrizable.

(viii) $Y_t = \Phi_Y(Y_{T_n}, t - T_n)$, *for* $T_n \leq t < T_{n+1}, n \in \mathbb{N}_0$.

(ix) $\mathbb{P}_y(Y_{T_{n+1}} \in B | T_0, Y_{T_0}, \ldots, Y_{T_n}, T_{n+1}) = Q_Y(\Phi_Y(Y_{T_n}, T_{n+1} - T_n); B)$
$\forall B \in \mathcal{B}(E_Y), n \in \mathbb{N}_0$.

The so defined process starts at $T_0 = 0$ in $Y_0 = y \in E_Y$ to then follow the path defined by $\Phi_Y(y,t)$ for $t < T_1$. At time T_1, the process jumps from[3] $\Phi_Y(y, T_1^-)$ to $Y_{T_1} \in E_Y$. This transition is described by $Q_Y(\Phi(y, T_1); \cdot) \in \mathbb{P}(E_Y)$. The probability distribution of T_1 is defined by λ_Y as detailed in property (vii) of the above definition. Once the process reaches Y_{T_1}, it restarts from this point following the same logic.

The process (Y_t) is thus „piecewise deterministic“ in the sense that its trajectory is completely determined by Φ_Y and the last post-jump state Y_{T_n} as long as $T_n \leq t < T_{n+1}$, so, as long as no new jump occurs. The stochastic properties of this process can be found in the jumps of the process. More precisely in both, in the jump time (see property (vii) of the above definition) as well as in the post jump state (see property (ix) of the above definition). A closer look on both, property (vii) and (ix), shows the „Markov property“ of the process. A formal proof is required to show the Markov property, but the reader familiar with Markov processes shall see, that the conditional probabilities appearing in these two defining properties of (Y_t) only depend on the last post jump state and not on the full history. Davis even showed the strong Markov property of the PDMP in the jump times, see [23, Thm. 25.5].

Property (i) of the previous definition being a very general requirement for the drift Φ_Y, this property is actually the minimum requirement we need to guarantee a Markovian behavior of (Y_t) even during the time intervals where it follows the deterministic path defined by the drift: Wherever you arrive after following $\Phi_Y(y, \cdot)$ for a time s, there is only one path you follow once you are in $y^\star := \Phi_Y(y,s)$, namely $\Phi_Y(y, \cdot)$ for $t \geq s$. This also means that, if starting in another state y' also leads you to the same point $y^\star$ after a time s', i.e. $\Phi_Y(y', s') = y^\star$, the path $\Phi_Y(y', t)_{t \geq s'}$ is the same as $\Phi_Y(y,t)_{t \geq s}$. Hence, if we knew that after a time period of deterministic behavior of (Y_t), the process will not have a jump during the next, say s seconds, then the (deterministic) development of (Y_t) during those s seconds only depends on the current state, not on the history how this state was reached. Note: The probability of facing a jump during the next s seconds depends on the time since the last jump.

Finally, we like to remark here that often, PDMPs are defined with a boundary jump condition: whenever the drift reaches the boundary of the state space, a jump is initiated and executed under the transition kernel Q. This leads in a very natural way to the notion of a „deterministic exit time“ that one can calculate at every state of the process supposing no jump occurs until the process following the drift reaches the boundary. We do not follow this approach in this thesis.

One important way to construct drifts satisfying the Markovian behavior of property (i) of the previous definition is to describe the drift by an ODE of first order, or more precisely by an initial value problem as detailed in the next example:

Example 1.2 (ODE defined drift). *Let* $E_Y \subset \mathbb{R}^d$ *and* $b : \mathbb{R}^d \to \mathbb{R}^d$ *a vector field guaranteeing for all* $y \in E_Y$ *a unique componentwise continuous solution* $\Phi(y, \cdot) : [0, \infty) \to E_Y$ *of the initial value problem*

$$\frac{d}{dt}\Phi(y,t) = b(\Phi(y,t)), \qquad \Phi(y,0) = y. \tag{1.1}$$

Then Φ *satisfies property (ib) of Definition 1.1.*

[3] we denote the left hand limit $\lim_{t \to s, t<s}$ by t^-

Many publications on PDMPs directly restrict their results to the case of „ODE defined drifts“. We do not need this restriction in this work and thus, have chosen to work under the general framework of property (i) as stated. One of the main fields of application of the theory, however, will remain cases of „ODE defined drifts“.

1.1.2 Describing a PDMP: An embedded Markov Chain in discrete time

Given that the states Y_t for $T_n \leq t < T_{n+1}$ are entirely described (and deterministic) by $\Phi_Y(Y_{T_n}, t - T_n)$ according to property (viii) of Definition 1.1, it is in enough to know the drift Φ_Y as well as the post jump states $(Y_{T_n})_{n\geq 0}$ and the jump times $(T_n)_{n\geq 0}$ to fully describe the path of $(Y_t)_{t\geq 0}$.

This concept of describing the full stochastic behavior of a PDMP $(Y_t)_{t\geq 0}$, which is a priori a stochastic process in *continuous time*, by the marked point process $(T_n, Y_{T_n})_{n\geq 0}$, which is a process in *discrete time*, is a crucial concept when working with PDMPs. We will now further detail this concept also being at the core of the development of a solution to the considered optimization problem in this thesis. For more details on marked point processes, we refer to the books of Last and Brandt [48] or of Brémaud [14]. They not only provide a very good and comprehensive introduction to this class of processes but also present an overview of filtering techniques for these processes under partial observation, thus, for tools we will need later in this thesis.

Definition 1.3 (embedded process - jump times). *Let $Z_n^Y := Y_{T_n}$ for $n \in \mathbb{N}_0$. The marked point process $(T_n, Z_n^Y)_{n\in\mathbb{N}_0}$ is called the* embedded process *of $(Y_t)_{t\geq 0}$.*

By replacing the jump time T_n (measured from the start of the process) by the n-th inter jump time $S_n := T_n - T_{n-1}$ (time between $(n-1)$-th and n-th jump) one can create another version of the embedded process, also called embedded process depending on the literature one is using:

Definition 1.4 (embedded process - inter jump times). *Let $S_0 := 0$ and $S_n := T_n - T_{n-1}$ for $n \geq 1$. The sequence $(S_n, Z_n^Y)_{n\in\mathbb{N}_0}$ is called the* underlying discrete Markov chain *of (Y_t).*

Clearly, the Markov property has to be proven for this latter process. This is what we will do for the rest of this section by developing the transition law for the process (S_n, Z_n^Y) in order to prove its Markov property. Remember that we denote by $\mathbb{P}_y$ the probability measure on the underlying measurable space $(\Omega, \mathcal{F})$ where $\mathbb{P}_y(Y_0 = y) = \mathbb{P}_y(T_0 = 0) = 1$.

Lemma 1.5 (Density of inter jump time distribution). *For $n \in \mathbb{N}$, the density of the inter jump time distribution $\mathbb{P}_y(S_n \leq t \mid S_0, Z_0^Y, \ldots, S_{n-1}, Z_{n-1}^Y)$ is given by*

$$f_n^Y(Z_{n-1}^Y, t) := e^{-\Lambda(Z_{n-1}^Y, t)}\ \lambda(\Phi(Z_{n-1}^Y, t)), \tag{1.2}$$

where we define $\Lambda(y,t) := \int_0^t \lambda(\Phi(y,s))\ ds$.

Proof. According to Definition 1.1, point (vii), we have (note that knowing $S_0, \ldots, S_{n-1}$, one also knows $T_0, \ldots, T_{n-1}$)

$$\mathbb{P}_y(S_n > t \mid S_0, Z_0^Y, \ldots, S_{n-1}, Z_{n-1}^Y) = \exp(-\Lambda(Z_{n-1}^Y, t)).$$

The density f_n^Y is derived by differentiation $\frac{d}{dt}$ of the distribution function $1 - \exp(-\Lambda(Z_{n-1}^Y, t))$ and we obtain the result. □

This density allows to describe the transition law of the embedded process in a way where the Markov property is then obvious.

Lemma 1.6 (Transition law for the embedded process). *For $n \in \mathbb{N}$ and $Z_0^Y = y \in E_Y$ and $B \in \mathcal{B}(E_Y)$, the transition law for the process (S_n, Z_n^Y) is given by*

$$\mathbb{P}_y(S_n \leq t, Z_n^Y \in B \mid S_0, Z_0^Y, \ldots, S_{n-1}, Z_{n-1}^Y) = \int_0^t \exp\left(-\Lambda(Z_{n-1}^Y, s)\right)\ \lambda\left(\Phi_Y(Z_{n-1}^Y, s)\right)\ Q_Y\left(\Phi_Y(Z_{n-1}^Y, s); B\right)\ ds.$$

Proof. For $n \in \mathbb{N}_0$, let $\mathcal{F}_n^Y := \sigma(S_0, Z_0^Y, \ldots, S_n, Z_n^Y)$. We then can write

$$\mathbb{P}_y(S_n \leq t, Z_n^Y \in B \mid S_0, Z_0^Y, \ldots, S_{n-1}, Z_{n-1}^Y) = \mathbb{P}_y(S_n \leq t, Z_n^Y \in B \mid \mathcal{F}_{n-1}^Y)$$

The last expression equals:

$$\begin{aligned}&\int_0^t \mathbb{P}_y(Z_n^Y \in B \mid \mathcal{F}_{n-1}^Y, S_n = s) \cdot \mathbb{P}_y(S_n \in ds \mid \mathcal{F}_{n-1}^Y)\ ds \\ &= \int_0^t f_n^Y(Z_{n-1}^Y, s)\ Q_Y(\Phi(Z_{n-1}^Y, s); B)\ ds.\end{aligned}$$

□

Corollary 1.7. *The process $(S_n, Z_n^Y)_{n\geq 0}$ is a discrete time Markov chain on the state space $E_{SY} := \mathbb{R}^+ \times E_Y$. For $n \in \mathbb{N}, B \in \mathcal{B}(E_Y)$ and $(0, y, s_1, y_1, \ldots, s_{n-1}, y_{n-1}) \in (\mathbb{R}^+ \times E_Y)^n$ its time homogeneous transition law Q_{SY} is given by:*

$$\begin{aligned}&Q_{SY}([0,t] \times B \mid 0, y, s_1, y_1, \ldots, s_{n-1}, y_{n-1}) \\ &= Q_{SY}([0,t] \times B \mid y_{n-1}) \\ &= \int_0^t \exp\left(-\Lambda(y_{n-1}, s)\right)\ \lambda\left(\Phi_Y(y_{n-1}, s)\right)\ Q_Y\left(\Phi_Y(y_{n-1}, s); B\right)\ ds.\end{aligned}$$

Proof. Follows directly from the proof of the previous lemma as only the las post jump state Z_{n-1}^Y intervenes in the formula for the transition law. □

1.2 Modeling partial observation of PDMPs

PDMPs under complete observation, i.e. with observable post jump states and known drift Φ, have been studied under a variety of aspects over the last thirty years[4]. For discrete-time Markov decision processes, partial observation has been studied extensively as well. A good introduction with references to further books and articles provides for example [6], where one can find a variety of applications to Finance as well.

Partial observation of PDMPs, however, has not played an important role in publications so far. Especially for *optimal control* of PDMPs under partial observation, the author is not aware of any publication treating the general control problem under discounted cost over lifetime. A first step towards this direction was made by Brandejski et al. in 2013, though: In their publication [13], they investigate the optimal stopping problem for a PDMP under partial observation. Stopping is a very simple control, thus,

[4] see introduction of this thesis

they are not yet introducing a general control model. They do, however, suggest a way of modeling partial observation by getting noisy measurements of the post jump state of a PDMP. We will follow the approach of Brandejski et al. of modeling partial observation by noisy measurements of the post jump state of the PDMP.

By no means, however, this approach shall be seen as the only possible way to model partial observation for a PDMP. A brief discussion of the pros and cons of alternative ways to model partial observation shall thus follow in Section 1.2.1 before we start to define and build the observation process used throughout the rest of this thesis in Section 1.2.2. Consequences of this way of modeling partial observation on the underlying probability space are then discussed in Section 1.2.3.

1.2.1 The question how to model partial observation

The notion of „partial observation“ does not impose one clearly and uniquely defined modification of the completely observable version of a given stochastic process. The term „observation“ might even be misleading sometimes as observation might seem as being linked to states of a process. However, the notion of partial observation has been used in recent research for both, situation where (parts of) states of a process are not observable as well as for situations where model parameters are unknown. The latter being perhaps better described as a situation of „partial information“. We need to clarify thus, what we mean by partial observation and motivate why we work under this condition.

The theory of Markov Decision Processes (MDP) has come up with a very general model for so-called PO-MDPs, meaning partially observable MDPs: One considers an MDP where the state has two components, say (X, Y). One of these states is observable by an agent controlling this process, the other not. This very general framework even allows for, e.g., Y being stochastically independent from X. Observing X does not allow to conclude anything about Y then. Preferable are thus situations where X and Y are depending on each other somehow. One being a noisy measurement of the other certainly is one example of such a dependence.

For the case of PDMPs, Kirch and Runggaldier [46] looked at a specific convex hedging problem on a financial market model with price processes under geometric Poisson distribution. In the second part of their work, they assume an unknown jump intensity. This is thus a case of partial information in the sense of unknown model parameters. One could also think of problems where the transition kernel Q of a PDMP is unknown.

This thesis tries to contribute to the case of „partial observation“ in the pure sense that an agent trying to control a PDMP cannot observe the state of the PDMP. Several models might arise from this imperfect information about the system state. In view of applications to problems from telecommunications, engineering, supply chain or finance[5], the idea is to assume that one can at least measure (or estimate) the true state of the system with some measurement noise. A choice has then to be made on how or when one can measure the state of the system: One could think of recurrent measurements in fixed time intervals, pre-planned deterministic measurement times or even measurements at random points in time.

In the case of PDMPs, knowing the last post jump state is sufficient for calculating the current state until the next jump occurs. If it is not possible to observe the state correctly, then the hope is that measuring the post jump state as good as possible is „good enough“ to take meaningful control decisions. Sure, more measurement points, also at non jump times, might help to improve the estimation of the true current status, but most of the application examples we have in mind would not or only under huge cost allow to do

[5] see Chapter 6 for more details on examples

additional measurements at other times than jump times.

Think of a production network that is running. Measuring the state of the whole network is something very complex. As soon as there is a break down of the network, however, good diagnostic tools are available to „measure“ the state of the system in terms of identifying the source of the breakdown.

Alternatively, think of the workload still waiting in a queue to be processed. You probably cannot do better than measure or estimate every new portion of workload entering the queue (at random points in time) and knowing the processing speed at every moment in time to have the best estimate possible for the workload currently waiting in the queue.

Both are examples potentially worth to be modeled as PDMP. A fundamental decision has though to be taken as one can see when comparing these examples: Shall the model account for measurements of the current status of the process at jump times (first example) or shall the model account for measurements of the change in process status at jump times (i.e. the jump height, second example). Both is certainly possible and meaningful in view of applications. For this thesis, we decided for the first version: Measuring the state of the process at jump times. This also implies that jump times are observable and we will now give the mathematical definitions of the observation process.

1.2.2 The observation process

We aim to work under a condition of *partial observation* where we only have

(i) full information about the inter-jump times $(S_n)_{n\geq 0}$ of $(Y_t)_{t\geq 0}$ and

(ii) noisy information about the post-jump states $(Z_n^Y)_{n\geq 0}$ of $(Y_t)_{t\geq 0}$.

We model this by introducing an observation process:

Definition 1.8 (Observation space). *Let $(E_X, +, 0_X)$ be a Polish space endowed with a commutative group structure with neutral element 0_X and $\psi : E_Y \to E_X$ a homeomorphism of topological spaces. We call $E_{SX} := \mathbb{R}^+ \times E_X$ the* observation space*.*

Definition 1.9 (Observation noise). *Let $(\epsilon_n)_{n\geq 0}$ be a sequence of E_X-valued i.i.d. random variables $\epsilon_n : \Omega \to E_X$, that are independent from $(S_n, Z_n)_{n\geq 0}$. We call ϵ_n* observation noise *and denote its distribution by Q_ϵ.*

Remark 1.10. *For all $y \in E_Y$, the PDMP (Y_t) with $Y_0 = y$ induces a probability measure $\mathbb{P}_y$ on $(\Omega, \mathcal{F})$. As the random variable ϵ_n is $(\mathcal{F}, \mathcal{B}(E_X))$-measurable, its distribution Q_ϵ on $\mathcal{B}(E_X)$ can be understood as the induced distribution $\mathbb{P}_y^{\epsilon_n}$, i.e. $Q_\epsilon(B) = \mathbb{P}_y^{\epsilon_n}(B) = \mathbb{P}_y(\epsilon_n \in B)$ for all $B \in \mathcal{B}(E_X)$.*

As we intend to model a pure observation noise (which in real life examples might be due to, e.g., noisy measurements of Z_n^Y) we required the independence of $(\epsilon_n)_{n\geq 0}$ from $(S_n, Z_n)_{n\geq 0}$ in the previous definition.

The fact that we do not get any new information between two jumps (except about time elapsed) is summarized in the following definition of the so-called observation process.

Definition 1.11 (Observation process). *Define $Z_n^X := \psi(Z_n^Y) + \epsilon_n$ for $n \geq 0$. We then define the* observation process *on E_{SX} for $t \geq 0$ by $(S_t, X_t)_{t\geq 0}$ and (with slight abuse*

of notation)

$$S_t := \sum_{n=0}^{\infty} \mathbb{1}_{[T_n, T_{n+1})}(t) \cdot (t - T_n), \tag{1.3}$$

$$X_t := \sum_{n=0}^{\infty} \mathbb{1}_{[T_n, T_{n+1})}(t) \cdot Z_n^X. \tag{1.4}$$

This means we have perfect observation of the *time elapsed since the last jump* S_t. In addition, we have a noisy measurement Z_n^X of the post-jump state of (Y_t). From this information[6] we can deduct

(i) The number $N_t := \#\{s \in (0,t] | S_s = 0\}$ of jumps occured until time t,

(ii) the jump times $T_0 := 0, T_1 := \inf\{t > 0 | S_t = 0\}, \ldots, T_n := \inf\{t > T_{n-1} | S_t = 0\}, \ldots$ and

(iii) the n-th inter-jump time $S_n := T_n - T_{n-1}$ denoted with a double abuse of notation: We use here a notation where S_n neither stands for the process (S_t) at time $t = n$ nor for the process (S_t) at time $t = T_n$. The latter being 0 according to the above definition of the process S_t. The simple rule to remember here is: Whenever we write S_t , we talk about a time since the last jump (that might be 0 at a jump time T_k), whenever we write S_n for an integer index n, we talk about the time between the $(n-1)$-th and n-th jump time.

Our PO-PDMP model can now be understood in the following way: Let $E := E_{SX} \times E_Y$ be the state space and (S_t, X_t, Y_t) a continuous time stochastic process with values in E where (S_t, X_t) and Y_t are defined as outlined before. (Y_t) is then the underlying, unobservable PDMP and (S_t, X_t) the observable part.

Remark 1.12. *The above defined observation process is slightly more general than the one presented in [13]. Brandejski et al. assume perfect observation of the inital state of the process, i.e.* $Z_0^X = y = Y_0$. *We assume a noisy measurement of* Y_0, *thus,* $Z_0^X = \psi(Y_0) + \epsilon_0$. *The situation of Brandejski et al. is included in our model, as we will explain in the next section.*

1.2.3 The underlying probability space

As long as we only had the completely observable PDMP (Y_t), we had a family of probability measures $(\mathbb{P}_y)_{y \in E_Y}$ depending on the initial state $Y_0 = y$ of the process. We derived a transition law for the embedded process (S_n, Z_n^Y) and showed its Markov property.

Now, looking at the process (S_t, X_t, Y_t), we can still argue that the only states of this process that are of interest are the states of the embedded process (S_n, Z_n^X, Z_n^Y). The reasoning being still: (Y_t) is fully described by Φ_Y and the combination of the inter-jump times (S_n) with the post-jump states (Z_n^Y), while X_t is constant between two jump times and S_t is a straight line between two jump times.

In order to describe this time-discrete process (S_n, Z_n^X, Z_n^Y), we can leverage the probability measure $\mathbb{P}_y$ and the distribution Q_ϵ of ϵ_n. For the latter we will assume, for the rest of this thesis, the existence of a bounded density w.r.t. some σ-finite measure on E_X.

[6] A real life example could consist of a noisy measurement of Z_n^Y being triggered every time we face a jump of (Y_t) and thus providing Z_n^X and setting the time counter for S_t to zero.

Assumption 1.13 (Bounded density of noise). *We assume the distribution Q_ϵ of the noise ϵ_n to have a bounded density function $f_\epsilon : E_X \to \mathbb{R}$ with respect to some σ-finite measure ν on $(E_X, \mathcal{B}(E_X))$, i.e. $Q_\epsilon(B) = \int_B f_\epsilon(x)\nu(dx)$ for all $B \in \mathcal{B}(E_X)$.*

In order to simplify notations, we will sometimes write $Q_\epsilon(B)$ instead of writing the full integral $\int_B f_\epsilon(x)\nu(dx)$.

Let Q_ϵ be the distribution of the i.i.d. random variables $(\epsilon_n)_{n\in\mathbb{N}_0}$.

Lemma 1.14 (Transition law for the embedded process of the PO-PDMP). *For $n \in \mathbb{N}$ and $Z_0^Y = y \in E_Y$ it then holds:*

$$\mathbb{P}_y(S_n \le t, Z_n^X \in B, Z_n^Y \in C \mid S_0, Z_0^X, Z_0^Y, \dots, S_{n-1}, Z_{n-1}^X, Z_{n-1}^Y) = \int_0^t \exp(-\Lambda(Z_{n-1}^Y, s))\lambda(\Phi_Y(Z_{n-1}^Y, s)) \int_C Q_\epsilon(B - \psi(y))\ Q_Y(\Phi_Y(Z_{n-1}^Y, s); dy)ds$$

Proof. The proof follows along the same lines as the proof for Lemma 1.6 taking into account that ϵ_n is independent of (S_n, Z_n^Y) as well as the definition of $Z_n^X := \psi(Z_n^Y) + \epsilon_n$.□

Corollary 1.15. *The process $(S_n, Z_n^X, Z_n^Y)_{n\ge 0}$ is a Markov process on the state space $E = \mathbb{R}^+ \times E_X \times E_Y$ with time-homogeneous transition law*

$$\begin{aligned} &Q_{SXY}([0,t] \times B \times C \mid s_{n-1}, x_{n-1}, y_{n-1}) \\ &= Q_{SXY}([0,t] \times B \times C \mid y_{n-1}) \\ &= \int_0^t \exp(-\Lambda(y_{n-1}, s))\lambda(\Phi_Y(y_{n-1}, s)) \int_C Q_\epsilon(B - \psi(y))\ Q_Y(\Phi_Y(y_{n-1}, s); dy)ds. \end{aligned}$$

Proof. Follows directly from the proof of the previous lemma as only the last post jump state Z_{n-1}^Y intervenes in the formula for the transition law. □

As we now take the perspective of an observer that cannot observe the states of (Y_t), especially not the initial state Y_0, a meaningful probability measure on $(\mathbb{R}^+ \times E_X \times E_Y)^\infty$ should not depend on the inital state $Y_0 = y$. Though, the so far discussed probability measure $\mathbb{P}_y$ on $(\Omega, \mathcal{F})$ does. To solve this inconvenience, we add an additional model parameter: We take an assumption on the initial distribution Q_0 of Y_0 and add this to our model parameters.

In the most general case we do not know anything about the underlying law that determines the initial state Y_0 of the unobservable process (Y_t). A reasonable assumption is to assume Y_0 to be generated by a random experience with probability distribution Q_0^-. We write the "$-$" as superscript to emphasize the fact that this is the distribution prior to any observation. A case where we know that the process (Y_t) will start in some fix $Y_0 = y \in E_Y$ could be modeled by assuming $Q_0^- = \delta_y$.

According to our model, we observe $X_0 = \psi(y_0) + \epsilon_0$ at the same time when $Y_0 = y_0$ is realized. Knowing the probability distribution Q_ϵ of ϵ_0 we can apply Bayes' formula to derive the conditional distribution Q_0^+ of Y_0 given the initial observation of $X_0 = x$, i.e. $Q_0^+(\cdot) := \mathbb{P}(Y_0 \in \cdot \mid X_0 = x)$. Note the superscript "$+$" to emphasize the fact that this is the assumed conditional distribution *after observation* of X_0. We summarize this as follows:

Assumption 1.16 (Initial distribution of Y_0). *We assume Y_0 to be determined by a random experiment with probability distribution Q_0^- on E_Y.*

Lemma 1.17 (Initial conditional distribution of Y_0). *The initial conditional distribution of Y_0 given $X_0 = x$ is given by*

$$Q_0^-(Y_0 \in C \mid X_0 = x) = \frac{\int_C f_\epsilon(x - \psi(y))\ Q_0^-(dy)}{\int_{E_Y} f_\epsilon(x - \psi(y))\ Q_0^-(dy)} \qquad \forall\ C \in \mathcal{B}(E_Y). \tag{1.5}$$

Proof. This is a simple application of Bayes' formula taking into account that $X_0 = \psi(y_0) + \epsilon_0$ and thus $Q_o^-(X_0 = x, Y_0 \in C) = \int_C f_\epsilon(x - \psi(y))\ Q_0^-(dy)$. □

Corollary 1.18. *Let Q_0^- the initial distribution of Y_0. Then, the right hand side of* (1.5) *defines a transition kernel Q_0^+ from E_X to $\mathbb{P}(E_Y)$.*

Thus, making an assumption on the initial *unconditional* distribution Q_0^- of Y_0 implies an assumption on a family of *conditional* distributions $(Q_0^+(x;\cdot))_{x \in X_0}$. Instead of making an assumption on the unconditional distribution to then derive the conditional distribution we will be working with, we can directly make an assumption on the conditional distribution. Therefore, to simplify notations, we will make the following assumption throughout the rest of this document:

Assumption 1.19. *Let $(Q_0(x;\cdot))_{x \in E_X}$ a family of probability measures on E_Y. We assume in our model, that the initial conditional distribution of Y_0 given the observation of $X_0 = x$ is described by $Q_0(x;\cdot)$.*

To summarize this discussion, we can define two probability measures on the space $(\mathbb{R}^+ \times E_X \times E_Y)^\infty$.

Taking the perspective of an observer that has not yet observed the initial state $X_0 = x$, we get: The transition laws Q_{SXY} of Corollary 1.15, the distribution Q_ϵ of ϵ_n as well as the initial (unconditional) distribution Q_0^- of Y_0 define a probability measure $\mathbb{P}$ on the space $(\mathbb{R}^+ \times E_X \times E_Y)^\infty$ according to the theorem of Ionescu-Tulcea. More precisely, $\mathbb{P}(\cdot) = \int \mathbb{P}(\cdot \mid S_0 = 0, X_0 = x, Y_0 = y)\ Q_\epsilon(dx - \psi(y))\ Q_0^-(dy)$.

Taking the perspective of an observer that has already observed the initial state $X_0 = x$, we find: Adding to the model parameters a family $(Q_0(x;\cdot))_{x \in E_X}$ of possible initial (conditional) distributions of Y_0 given the initial observation of $X_0 = x$, we get a probability measure $\mathbb{P}_x(\cdot)$ on $(\mathbb{R}^+ \times E_X \times E_Y)^\infty$ by $\mathbb{P}_x(\cdot) = \int \mathbb{P}(\cdot \mid S_0 = 0, X_0 = x, Y_0 = y)\ Q_0(x; dy)$. This again follows from the theorem of Ionescu-Tulcea based on the transition kernels Q_{SXY} of corollary 1.15.

This latter perspective is the one we will take for the rest of this work. We will often only write $Q_0(\cdot)$ instead of $Q_0(x,\cdot)$ if the context is clear.

1.3 Controlling the process based on partial observation

After having introduced the observation process in the previous section, we will now develop the control model, or, more precisely, define the model of a *controlled PO-PDMP* (Partially Observable Piecewise Deterministic Markov Process). This section thus contains the first result of this thesis, as to our best knowledge, no general model for controlled PDMPs under partial observation has been published yet. The model we develop here is inspired from three areas of research: classicical models of controlled PDMPs under

complete observation (see, e.g., [23] or [20]), the theory of partially observed MDPs (discrete time problems, see, e.g., [6]) and recent research on optimal stopping of partially observable PDMPs (see, e.g., [13]).

We will use history dependent relaxed piecewise open loop policies as a collection of (pre-defined) decision rules an agent will use to decide on the control action to apply to the process. As our PO-PDMP is a time-continuous stochastic process, the control applied to the process should also be time-continuous, i.e. modeled by a time-continuous process $(A_t)_{t\geq 0}$ of so-called *control actions*. A control action A_t should act on all three characteristics λ_Y, Φ_Y and Q_Y of the underlying PDMP (Y_t).

A *control policy* π is then a set of upfront defined decision rules that shall be applied to decide what control action to execute at time t. As we only have partial observation of our underlying PDMP (Y_t), a decision rule at time t shall follow the *separation principle of estimation and control*: We first observe all information available until time t to then estimate the current status of the process before the agent takes a control decision based on the available information. Thus, a decision rule has to be a function of the observable information up to time t.

The available information up to time t not only consisting of the observed time elapsed, the jump times and the noisy measurements of the post jump states, but also the control actions applied to the process until time t, decision rules enter the model twice: Clearly, they are applied when deciding on the current control action to execute. They are, however, as well part of the observable history, as knowing the decision rules the agent had used in the past and the observations of time and measurements he had at that moment defines the control actions that have been executed until time t. This is a point where giving a correct definition becomes difficult: one needs to define a decision rule as a function of (amongst others) decision rules. We suggest here a definition by recursion and first look at a one period model before we define a multi period model.

An interesting question is then to find out at what point in time the agent really needs to take a decision: Really at every point in time if there is no relevant new information as long as no new jump occurs? To clarify this question, we start this chapter with a very brief introduction to general control theory in Section 1.3.1. We will explain the difference between closed loop (feedback) controls and open loop controls before we start building the model for history dependent relaxed piecewise open loop policies in Section 1.3.2. This is, where we will introduce the space of observable histories together with the notion of decision rules. Readers not familiar with the Young topology might want to first take a look at Annex A, even though we will not need more than to know the name of this topology for the moment. Finally, we will define the model of a controlled PO-PDMP in Section 1.3.3, where we will also define the controlled jump rate, transition kernel and drift.

1.3.1 The concepts of open and closed loop controls

Control theory is an interdisciplinary field of research between (electrical) engineering and mathematics. It is the study of the (change in) behavior of dynamical systems getting inputs from a controller while also providing feedback to the controller. The main concepts of control of a dynamical system are (i) closed loop controls (also known as feedback controls), (ii) open loop controls and (iii) impulse controls.

For closed loop controls, there is a feedback loop between the controller and the dynamical system to control. The controller provides input to the dynamical system, i.e. is executing a control action. The system will provide feedback to the controller, e.g. about the new state of the system after the control action executed. This new information will be

processed by the controller in order to decide on the new control action to execute and the loop starts again. An example would be a cruise control in a car of recent generation: The driver activates cruise control and selects the desired speed. The controller is measuring the current speed of the car and comparing current speed to desired speed. The difference providing information to the controller whether an accelaration or a breaking is required. The controller will execute this action and right after measure the current speed again to compare it with the desired speed. The loop is starting again. If ever the car is climbing a hill, the controller will then accelerate as the speed of the car is falling.

An open loop control is a control where a controller is once taking a decision on what control action to execute. Once the action executed, there will be no feedback from the controlled dynamic system. An example would be cruise control of very early generations that basically only locked in the current throttle position of the engine. This would guarantee to hold the current speed of the car as long as there is no hill the car would need to climb up or roll down. As the controller would not get any feedback from the car, the change in speed while climbing up a hill cannot be detected by the controller and thus no new control action can be enganged in order to correct the speed of the car.

Both, open and closed loop controls, when applied to PDMPs, shall be understood as „continuous“ controls in a sense that we need to specify later. Roughly speaking, the control acts „continuously“ at every point in time on the jump rate, the transition kernel and the drift. Thus, the directions of the drift, the probability of obtaining a new jump and the post jump state can be controlled in a sense we will specify later. However, not directly allowed for open and closed loop controls, are control actions that intervene in the evolution of the process by provoking a jump of the process at a stopping time specified by the controller in order to move the process to a controller-specified post jump state.

This is what so-called *impulse control* is about. Bensoussan and Lions [10] were pioneers of this subject in the context of diffusion processes. Many publications and applications followed their initial article and Davis [23] has developed a „self-contained theory of impulse control for PDMPs“ under complete observation. A more recent work of Costa and Raymundo [21] is also studying impulse control besides classical continuous control of PDMPs.

This thesis is not covering impulse control. The question, however, whether open or closed loop controls are adequate for the models investigated is a crucial one and shall be briefly discussed in the sequel. For PDMPs under complete observation, people usually use so-called *relaxed piecewise open loop policies* for the class of admissible control policies. The term „relaxed“ will be explained later in this work, important to notice here is the fact that *open* instead of *closed* loop controls are used in classical PDMP control problems under complete observation. There are at least two reasons for this, one of them being of more analytical nature concerning existence of solutions to initial value problems for ODEs, as we will explain once we introduced the concept of *relaxed* controls. The other reason is easy to understand from a stochastic point of view: If we have complete observation of the process, it is enough to know the last jump time T_n as well as the last post jump state Y_{T_n} to decide what control action to execute at time t with $T_n \leq t < T_{n+1}$. There is actually no new information until time t except of the fact that no new jump has occurred since T_n. The state Y_t is thus not required as input parameter to a decision rule at time t as one can calculate it based on Y_{T_n} and t by $Y_t = \Phi(Y_{T_n}, t)$. A controller can thus take a control decision at time T_n on how to control the process until the next jump occurs. This is an open loop policy then as no feedback from the process is considered in the control decision made „upfront“ at time T_n for all $T_n \leq t < T_{n+1}$. It is a *piecewise* open loop policy then, as the controller will only follow the decision taken at time T_n until the next jump occurs at time T_{n+1}.

In the setting of this thesis, where the post jump states of the process (Y_t) are not observable directly but only via a noisy measurement, one could argue that closed loop controls might make sense: Making a control decision at time t depending on a current noisy measurement of the state Y_t might make sense as the measurement in continuous time of the unobservable state would improve the current estimate of the true, but unobservable state of the process. It would require, however, the ability of executing these measurements in continuous time. For the application examples we have in mind, this would not work either for technical reasons that do not allow continuous-time measurements or for huge associated costs. We develop a theory for applications, where measurements of the unobservable state are only possible at jump times.

We thus focus on piecewise open loop policies in this thesis and will develop the necessary theory to define a controlled PO-PDMP under relaxed piecewise open loop policies in the following sections. We start by defining the space of observable histories and history dependent relaxed piecewise open loop policies in Section 1.3.2. A separation into a one period and a muli period model will be presented there to develop all necessary definitions correctly in an approach by recurrence. We will then define how relaxed controls will act on the characteristics of the PO-PDMP in Section 1.3.3. At the end of that Section, we will present the model of a controlled PO-PDMP on an extended state space. The extended state space will be required to formulate the model such that the structure of a PDMP with only state dependent intensities and transition kernels is conserved.

1.3.2 The concept of history dependent relaxed piecewise open loop policies

The goal of this section is to define the set of admissible control policies (sometimes also called „strategies“) for a PO-PDMP. A control policy shall be understood as a set of upfront defined decision rules that shall be applied to decide how to control a PO-PDMP at any point in time $t \geq 0$.

The action space and the relaxed action space :
Applying a decision rule at time t shall deliver an action A_t to be executed at time t and thus, we start by defining the *action space*.

Definition 1.20 (Action space). *Let A a compact metric space called the* action space. *An element $a \in A$ is called* (control) action. *Denote by d_A the metric on A and by $\mathcal{T}[d_A]$ the topology on A enduced by d_A. Let $\mathcal{B}_A$ the Borel-σ-algebra on A generated by $\mathcal{T}[d_A]$, thus $(A, \mathcal{B}_A)$ is a measurable space.*

From earlier research on optimal control of *completely* observable PDMPs (see, e.g., [34]), we know that optimal control policies, in general, can only be found in the set of so-called *relaxed* or *randomized* controls. Therefore we expect optimal policies for our PO-PDMP to also be of this type and define the relaxed action space:

Definition 1.21 (Relaxed action space). *Let $\mathbf{P}(A)$ the space of probability measures on $(A, \mathcal{B}_A)$ and endow $\mathbf{P}(A)$ with the weak topology $\mathcal{T}[\mathbf{C}(A)]$. Denote by $\mathcal{B}_{\mathbf{P}(A)}$ the Borel-σ-algebra on $\mathbf{P}(A)$ generated by $\mathcal{T}[\mathbf{C}(A)]$, then $(\mathbf{P}(A), \mathcal{B}_{\mathbf{P}(A)})$ is a measurable space, called* relaxed action space.

Remark 1.22. *The reader not familiar with the weak topology on a space of probability measures may refer to section A.1.2 and especially to Definition A.3.*

The way how one should understand the space $\mathbf{P}(A)$ as action space is, that, instead of executing a (deterministic) control action $a_t \in A$ at time t, we will execute a *relaxed*

control action $p_t \in \mathbf{P}(A)$ at time t. Identifying an element $a \in A$ with $\delta_a \in \mathbf{P}(A)$ shows, that, in this sense $\mathbf{P}(A)$ can be understood as an extension of the determinstic action space A.

We refer to section 1.3.3 for more details on how a relaxed control action acts on the model parameters λ_Y, Φ_Y and Q_Y.

The one period model :
We start by defining the notion of a *decision rule* for the one period model, that is, for the time interval $[0, T_1)$. A decision rule shall be understood as an upfront defined rule that tells us, what relaxed control action to execute at time $t \in [0, T_1)$ given the observable history up to time t.

The observable history up to time $t \in [0, T_1)$ basically consists of three parts:

1. The observed path of $(S_s)_{s \leq t}$,
2. the observed path of $(X_s)_{s \leq t}$ and
3. the executed (relaxed) control actions $(p_s)_{s<t}$ (note the sign "$<$" instead of "$\leq$").

For a point in time t with $0 \leq t < T_1$, the observable history up to time t can basically be separated into two components:

a) the initial observation at time $T_0 = 0$ and

b) the observable history since $T_0 = 0$, that is the observable history on $(0, t]$.

The important point here is, that the only observation we make on $(0, t]$ is that there was no jump of the process (Y_t) since $T_0 = 0$. This means, (S_s) is the straight line $S_s := s$ and $X_s := X_0$ is constant for $s \in (0, t]$. Thus, the time $S_t = t$ (elapsed time since process start) is sufficient to describe the observation of (S_s, X_s) on $(0, t]$.

In this sense, a decision rule for the first period of our PO-PDMP should be a measurable function of the initial observation as well as of the time elapsed since process start and we formalize this with the following definition.

Definition 1.23. *Let $\mathcal{H}_0 := \mathbb{R}^+ \times E_X$ the space of possible initial observations (or of observable histories up to time T_0) endowed with the product σ-algebra. A decision rule for the first period of the PO-PDMP is then a measurable mapping*

$$\pi_0 : \mathcal{H}_0 \times [0, \infty) \to \mathbf{P}(A).$$

Applying this decision rule on $[0, T_1)$ then means executing the relaxed control action $\pi_0(h_0, t - T_0) \in \mathbf{P}(A)$, where $h_0 \in \mathcal{H}_0$ is the initial observation we made at time $T_0 = 0$.

Remark 1.24. *Two remarks on the previous definition:*

a) Clearly, this definition ensures that the relaxed controls executed on $[0, T_1)$ are a measurable function of the observed paths of (S_s) and (X_s). However, implicitly, this definition also ensures that the relaxed control action executed at time $t \in [0, T_1)$ is a function of the (observed) path of executed relaxed control actions $(\pi_0(h_0, s))_{s<t}$. This is intrinsic to π_0 being a function defined on $\mathcal{H}_0 \times [0, \infty)$.

b) As we require $(Y_t)_{t \geq 0}$ to satisfy $\mathbb{P}(T_0 = 0) = 1$, an initial observation h_0 will w.l.o.g. always be of the form $h_0 = (0, x)$ for some $x \in E_X$.

The multi period model :
In order to control a multi period model, i.e. control a PO-PDMP for countably many time intervals $[T_n, T_{n+1})$, we can generalize the definitions of the one period model by recursion.

Actually, for any time interval $[T_n, T_{n+1})$ and a point in time t within this interval, we find analogously to the one period model: The observable history up to time t can be separated into (i) the observable history up to time T_n and (ii) the observable history from T_n to t. The latter one is, as for the one period model, sufficiently described by the time $\tau := t - T_n$. The observable history up to time T_n can be separated into the observable history up to time T_{n-1} and the observable history from T_{n-1} to T_n. This is where we make use of a definition by recursion:

Suppose the space $\mathcal{H}_{n-1}$ of observable histories up to jump time T_{n-1} defined in a way such that the information coded in an observed history $h_{n-1} \in \mathcal{H}_{n-1}$ is enough to fully describe the paths of $(S_s)_{s \leq T_{n-1}}, (X_s)_{s \leq T_{n-1}}$ as well as the path of the relaxed control actions $R_s \in \mathbf{P}(A)$ that have been executed for $0 \leq s < T_{n-1}$. Suppose futher, that a decision rule for the period $[T_{n-1}, T_n)$ is a measurable mapping $\pi_{n-1} : \mathcal{H}_{n-1} \times [0, \infty) \to \mathbf{P}(A)$.

Under these assumptions, we need to define how to describe the paths of (S_s) and (X_s) on $(T_{n-1}, T_n]$ as well as the path in $\mathbf{P}(A)$ of the executed relaxed control actions for the time interval $[T_{n-1}, T_n)$. We start with an important observation on how to describe the executed relaxed control actions:

A decision rule π_{n-1} together with an observed history $h_{n-1} \in \mathcal{H}_{n-1}$ up to time T_{n-1} define a measurable mapping

$$\pi_{n-1}(h_{n-1}, \cdot) : [0, \infty) \to \mathbf{P}(A)$$

to be executed on the time interval $[T_{n-1}, T_n)$ as $\pi_{n-1}(h_{n-1}, t - T_{n-1})$. Thus, to reconstruct the path in $\mathbf{P}(A)$ of the executed relaxed control actions on $[T_{n-1}, T_n)$, it is enough to "remember" the function $\pi_{n-1}(h_{n-1}, \cdot)$ as an element of a function space and to observe the n-th inter-jump time $S_n = T_n - T_{n-1}$ that indicates how long we execute this function. This leads to the following definition for an adequate function space for this purpose:

Definition 1.25. *We define the space $\mathcal{R}$ of* relaxed controls *as*

$$\mathcal{R} := \{[r] \mid r : [0, \infty) \to \mathbf{P}(A), r \text{ measurable}\},$$

where $[r]$ denotes the λ^1-equivalence class of r.

Remark 1.26. *Some remarks regarding this definition:*

a) We use λ^1 equivalence classes in this definition, that means $\tilde{r} \in [r] \Leftrightarrow \tilde{r} = r$ for λ_1-almost all $t \in [0, \infty)$.

b) r shall be measurable w.r.t. the Borel-σ-algebras $\mathcal{B}([0, \infty))$ and $\mathcal{B}_{\mathbf{P}(A)}$, see Annex A.1.2 for further details on the weak topology on $\mathbf{P}(A)$ as well as on $\mathcal{B}_{\mathbf{P}(A)}$. Remind that $\mathbf{P}(A)$ is separable and metrizable and thus $\mathcal{B}_{\mathbf{P}(A)} = \sigma(\mathcal{V}(\mathbf{C}(A)))$, i.e.

$$r^{-1}(V_\epsilon(p, f)) \in \mathcal{B}([0, \infty)) \quad \forall \epsilon > 0, f \in \mathbf{C}(A), p \in \mathbf{P}(A),$$

where we use the notation of Annex A.1.2, Definition A.3.

With S_n we also observe X_n at time T_n and these observations allow to reconstruct the full path of (S_s) and (X_s) on $[T_{n-1}, T_n]$. Putting all these aspects together, the following definition shall be motivated sufficiently:

Definition 1.27. *For $n \geq 1$ we define the space of observable histories up to time T_n by recursion as*

$$\mathcal{H}_n := \mathcal{H}_{n-1} \times \mathcal{R} \times \mathbb{R}^+ \times E_X,$$

and endow this space with the product σ-algebra of the usual Borel σ-algebras on $\mathbb{R}^+$ and E_X as well as of the Borel σ-algebra $\mathcal{B}_{\mathcal{R}}$ deduced from the Young topology on $\mathcal{R}$.

An element $h_n = (s_0, x_0, r_0, \dots, s_n, x_n) \in \mathcal{H}_n$ is called observed history up to time T_n.

A decision rule for the period $[T_n, T_{n+1})$ is a measurable mapping

$$\pi_n : \mathcal{H}_n \times [0, \infty) \to \mathbf{P}(A).$$

For $n \in \mathbb{N}_0$, the space of all decision rules for the period $[T_n, T_{n+1})$ is denoted by Π_n^P and the space of all history dependent relaxed piecewise open loop policies is defined as

$$\Pi^P := \times_{n \in \mathbb{N}_0} \Pi_n^P.$$

Remark 1.28. *The Young topoloy on $\mathcal{R}$ as well as the resulting Borel σ-algebra are discussed in detail in the Annex of this work. For instance, we do not need any further details than to know what σ-algebra to use in the measurability requirement on π_n.*

Executing a history dependent relaxed piecewise open loop policy $\pi = (\pi_0, \pi_1, \dots) \in \Pi^P$ means executing, at time $t \geq 0$

$$\pi_t := \sum_{n=0}^{\infty} \mathbb{1}_{\{T_n^\pi \leq t < T_{n+1}^\pi\}}(t) \cdot \pi_n(H_n^\pi, t - T_n^\pi), \tag{1.6}$$

where T_n^π stands for the n-th jump time of the π -ontrolled PO-PDMP and H_n^π is the $\mathcal{H}_n$-valued random variable describing the observable history up to time T_n^π of the π-controlled PO-PDMP. It is defined recursively by

$$\begin{aligned}
&H_0^\pi := (S_0^\pi, X_0^\pi), \\
&R_0^\pi := \pi_0(H_0^\pi), \\
&H_n^\pi := (S_0^\pi, X_0^\pi, R_0^\pi, \dots, S_{n-1}^\pi, X_{n-1}^\pi, R_{n-1}^\pi, S_n^\pi, X_n^\pi) \qquad \forall n \geq 1, \\
&R_n^\pi := \pi_n(H_n^\pi) \qquad \forall n \geq 1,
\end{aligned}$$

where we put again the superscript π to indicate the fact that the random variables come from the π-controlled PO-PDMP.

1.3.3 The controlled PO-PDMP

We turn now into the definitions on how a history dependent relaxed piecewise open loop policy π - and more precisely, the resulting relaxed control actions $\pi_t \in \mathbf{P}(A)$ at each time t - acts on the model parameters Φ_Y, λ_Y and Q_Y of the underlying PDMP (Y_t).

The controlled drift :
At each point in time, the relaxed control action $\pi_t \in \mathbf{P}(A)$ should act on the drift Φ_Y and thus create the controlled drift Φ^π. Our main requirement will be that the controlled drift still fullfills the characteristic properties of a drift of a PDMP, especially $\Phi^\pi(y, t+s) = \Phi^\pi(\Phi^\pi(y,t), s)$. The latter being a property „over the course of time“, the following consideration turns out to be useful:

Given a history dependent relaxed piecewise open loop policy $\pi = (\pi_0, \pi_1, \dots)$, we actually control the drift „over time“ on a period $[T_n, T_{n+1})$ with $\pi_n(h_n, \cdot) \in \mathcal{R}$.

This means, it is enough to define how a relaxed control $r \in \mathcal{R}$ will act on Φ_Y and thus create the controlled drift Φ^r.

Assumption 1.29 (Controlled drift). *Let $[r] \in \mathcal{R}$ an arbitrary relaxed control. We assume that the drift Φ^r of the controlled PDMP (Y_t^r) is a continuous mapping $\Phi^r : E_Y \times \mathbb{R}^+ \to E_Y$ and satisfies:*

(i) The mapping $t \mapsto \Phi^r(\cdot, t)$ is a semi-group, i.e. for all $y \in E_Y$:

$$\Phi^r(y, s+t) = \Phi^r(\Phi^r(y, s), t).$$

(ii) The controlled drift Φ^r is λ^1-a.e. independent of the choice of a representative of $[r]$, i.e., for all $r' \in [r]$ and all $y \in \mathbb{E}_Y$:

$$\Phi^r(y, t) = \Phi^{r'}(y, t) \text{ for } \lambda^1-\text{almost all } t \in [0, \infty).$$

For an ODE defined drift Φ as of example 1.2, the natural way how a relaxed control policy should act on Φ is by influencing the time derivative of Φ. In that case, a standard assumption to model the response of Φ on a relaxed control is the following assumption:

Example 1.30 (ODE controlled Drift). *Let E_Y as in example 1.2 and A a compact metric space. Further, let $b : \mathbb{R}^d \times A \to \mathbb{R}^d$ a vector field such that for all $y \in E_Y$ and all relaxed controls $r \in \mathcal{R}$ the initial value problem*

$$\frac{d}{dt}\Phi^r(y, t) = \int_A b(\Phi^r(y, t), a)\ r_t(da), \qquad \Phi^r(y, 0) = y \tag{1.7}$$

has a unique componentwise continuous solution $\Phi^r(y, \cdot) : [0, \infty) \to E_Y$.

Remark 1.31. *For the controlled ODE defined drift, property (ii) of the definition of a controlled drift is satisfied, as here, $\int_0^t \int_A b(\Phi^r(y, t), a)\ r_t(da)\ dt$, is independent of the choice of a representative of $[r]$.*

One should understand the way how the relaxed control policy acts on the time derivative of Φ as *"creating an expected time derivative"*. The same calculation as done after example 1.2 shows that the so defined ODE controlled drift Φ^r satisfies property (i) of Assumption 1.29.

In Section 1.3.1, we mentioned a second reason why for completely observable PDMPs, closed loop controls are not applied in general. Knowing history dependent relaxed controls, we can detail this further now:

For completely observable PDMPs, a relaxed closed loop control would thus be a measurable mapping

$$\pi : E_Y \to \mathbf{P}(A).$$

In the case of an ODE defined drift of the underlying PDMP, the controlled drift should then be defined as solution to the initial value problem

$$\frac{d}{dt}\Phi^\pi(y, t) = \int_A b(\Phi^\pi(y, t), a)\ \pi(\Phi^\pi(y, t); da), \qquad \Phi^\pi(y, 0) = y.$$

This initial value problem, however, is not easy to solve: Even under restrictive assumptions to the vector field b, one can show that there is no guarantee for the existence of a unique optimal solution Φ^π for every closed loop control π. As ODE defined drifts play one of the most important roles in concrete applications of the PDMP theory, closed loop controls would thus not be the most adequate set of admissible control policies.

The deterministically controlled jump rate and transition kernel :
In order to define the controlled jump rate and transition kernel, we start by defining how a deterministic control action $a \in A$ acts on those:

Definition 1.32 (Controlled jump rate). *Let $\lambda^A : E_Y \times A \to (0, \infty)$ a continuous and bounded function satisfying for all $n \in \mathbb{N}, \pi_n \in \Pi_n^P, h_n \in \mathcal{H}_n, y \in E_Y$:*

$$\lim_{t\to\infty} \int_0^t \int_A \lambda^A(\Phi^{\pi_n(h_n,\cdot)}(y,s),a)\ \pi_n(h_n,s)(da)\ ds = \infty \tag{1.8}$$

Definition 1.33 (Controlled transition kernel). *Let $Q^A : E_Y \times A \to \mathbf{P}(E_Y)$ a weakly continuous transition kernel.*

The controlled PO-PDMP model on an extended state space :
The definition of how a relaxed control acts on the model parameters λ_Y and Q_Y will now be given together with the full model of the controlled PO-PDMP model. In the classical PDMP model, the jump rate λ is a function defined on the state space of the process. The intensity of a controlled PO-PDMP shall as well have the state space of the controlled process as domain of definition. As we are using history dependent control policies, we therefore need to extend the state space of the PO-PDMP by the observable histories $\mathcal{H}_n$. As these spaces are different for each $n \in \mathbb{N}$, we also need to introduce the concept of external states of a PO-PDMP and finally get an extended state space with countably many external states. This formalizes as follows:

Definition 1.34 (derived π-controlled PO-PDMP). *Let $(S_t, X_t, Y_t)_{t\geq 0}$ a PO-PDMP as introduced before and $\pi \in \Pi^P$. The derived π-controlled PO-PDMP is a PDMP in the sense of Definition 1.1 on the extended state space $\tilde{E}$ with parameters $\lambda^\pi, \tilde{\Phi}^\pi, Q^\pi$ where we define:*

(i) The state space $\tilde{E}$ tracks the number of already observed jumps of the process as external states and is defined as

$$\tilde{E} := \{(n, z) \mid n \in \mathbb{N}_0, z \in E_n\},$$

where the state space for the external state n is defined as

$$E_n := \mathbb{R}^+ \times E_X \times E_Y \times \mathcal{H}_n.$$

(ii) The drift $\tilde{\Phi}^\pi$ of the π-controlled PO-PDMP is defined as

$$\tilde{E}\times[0,\infty) \ni (n, s, x, y, h_n, \tau) \mapsto \tilde{\Phi}^\pi(n, s, x, y, h_n, \tau) := (n, s+\tau, x, \Phi^{\pi_n(h_n,\cdot)}(y, \tau), h_n) \in \tilde{E}.$$

(iii) The jump rate λ^π is defined as

$$\lambda^\pi : \tilde{E} \to (0, \infty), (n, s, x, y, h_n) \mapsto \int_A \lambda^A(y, a)\ \pi_n(h_n, s)(da).$$

(iv) The transition kernel Q^π from $\tilde{E}$ to $\mathbb{P}(\tilde{E})$ is defined as

$$Q^\pi(n, s, x, y, h_n; \cdot) := \frac{(\lambda Q)^\pi(n, s, x, y, h_n; \cdot)}{\int_A \lambda^A(y, a)\ \pi_n(h_n, s)(da)},$$

where we define for $B := B_{\mathbb{N}} \times B_{\mathbb{R}^+} \times B_{E_X} \times B_{E_Y} \times B_{\mathcal{H}_{n+1}}$ *being a product of respective Borel sets of the underlying spaces:*

$$(\lambda Q)^\pi(n,s,x,y,h_n;B) := \mathbb{1}_{B_{\mathbb{N}}}(n+1)\cdot\mathbb{1}_{B_{\mathbb{R}^+}}(0)\int_A\int_{B_{E_Y}}\int_{B_{E_X}}\mathbb{1}_{B_{\mathcal{H}_{n+1}}}(h_n,\pi_n(h_n,\cdot),s,x')\cdots \cdots Q_\epsilon(dx'-\psi(y'))\ \lambda^A(y,a)\ Q^A(y,a;dy')\ \pi_n(h_n,s)(da).$$

This means, the controlled PO-PDMP starts in $(0,0,x,y_0,(0,x))$ to then follow the path $(0,\tau,x,\Phi^{\pi_0((0,x),\cdot)}(y_0,\tau),(0,x))$ for $\tau\in[0,T_1)$. At time T_1, the process jumps from $(0,T_1,x,\Phi^{\pi_0((0,x),\cdot)}(y_0,T_1),(0,x))$ to the state $(1,0,X_{T_1},Y_{T_1},(0,x,\pi_0((0,x),\cdot),T_1,X_{T_1}))$ and this transition is happening according to Q^π.

Every post jump state has the form $(n,0,x_n,y_n,h_n)$ where h_n is of the form $h_n = (h_{n-1},\pi_{n-1}(h_{n-1},\cdot),s_n,x_n)$.

Analogously to the uncontrolled PDPM, for every $\pi\in\Pi^P$, there is a family of probability measures $(\mathbb{P}^\pi_{(0,z)})_{z\in E_0}$ on $(\Omega,\mathcal{F})$. As $z\in E_0$ has the form $(0,x,y,(0,x))$, this family of probability measures is only depending on $(x,y)\in E_X\times E_Y$. Thus, the initial observation $x\in E_X$ together with the initial (conditional) distribution Q_0 of Y_0 define a probability measure $\mathbb{P}^\pi_x(\cdot)=\int\mathbb{P}^\pi_{x,y}(\cdot)\ Q_0(dy)$.

The probability distribution of the first jump time T_1 is here determined depending on λ^π according to the following law:

Lemma 1.35. *Let* $(x,y)\in E_X\times E_Y$ *determine an initial state of the PO-PDMP* (S_t,X_t,Y_t) *and* $\pi\in\Pi^P$. *The (conditional) density of the probability distribution for the first jump time* T_1^π *of the derived* π*-controlled PO-PDMP is given by*

$$f^\pi_{T_1}(t\mid y) := \exp(-\Lambda^\pi(0,y,h_0,t))\int_A\lambda^A(\Phi^{\pi_0(h_0,\cdot)}(y,t),a)\ \pi_0(h_0,t)(da),$$

where $h_0=(0,x)$ *and where we define for* $n\in\mathbb{N}_0$

$$\Lambda^\pi(n,y,h,t) := \int_0^t\int_A\lambda^A(\Phi^{\pi_n(h,\cdot)}(y,\tau),a)\ \pi_n(h,\tau)(da)\ d\tau.$$

Proof. According to definition/construction, we have

$$\mathbb{P}^\pi_{x,y}(T_1>t)=\exp\left(-\int_0^t\lambda^\pi(\tilde\Phi^\pi((0,0,x,y,(0,x)),\tau))\ d\tau\right).$$

Applying the definitions of $\tilde\Phi^\pi$, λ^π and Λ^π, this becomes

$$\mathbb{P}^\pi_{x,y}(T_1>t)=\exp\left(-\int_0^t\int_A\lambda^A(\Phi^{\pi_0(h_0,\cdot)}(y,\tau),a)\ \pi_0(h_0,\tau)(da)\ d\tau\right)=\exp(-\Lambda^\pi(0,y,h_0,t)).$$

The density being the derivative $\frac{d}{dt}$ of the distribution function $1-\exp(-\Lambda^\pi(0,y,h_0,t))$, the result follows. □

Remark 1.36. *Looking at this density and especially at the definition of* Λ^π, *the necessity of condition (1.8) becomes clear in order to guarantee that* $T_n\uparrow\infty$ $(n\to\infty)$ *which implicitly states* $T_n<\infty$ *a.s. for all* $n\in\mathbb{N}$.

1.4 The initial optimization problem in continuous time

In this section we state the initial continuous-time optimization problem of minimizing the total expected discounted cost over lifetime of a PO-PDMP. This optimization problem, characterization of its value function and the question of existence of optimal policies are the core subjects of this thesis in the sequel. We end this section by briefly summarizing the approach followed in order to prove existence of optimal policies for this optimization problem.

1.4.1 The optimization problem

We will state the initial, time-continuous optimization problem for cost optimal control of a PO-PDMP under history dependent relaxed piecewise open loop policies. Two main criteria for *cost optimal* control are usually investigated: (i) Minimal average cost and (ii) minimal discounted cost over lifetime. The focus of this thesis is only on minimal discounted cost over lifetime.

We give the general formulation of our optimization problem for a time interval $[0; T_\infty)$ where T_∞ can represent either a deterministic time horizon in which case one has to distinguish $T_\infty < \infty$ from $T_\infty = \infty$ or a stochastic time horizon given by a stopping time T_∞. In this thesis, we will only be treating the cases $T_\infty = \infty$ and the case of $T_\infty = T_n$ for any upfront selected n-th jump time. The case of a fixed finite time horizon will not be covered by this thesis.

Definition 1.37 (Optimization Problem). *Let $\beta \in \mathbb{R}^+$ and $c : E_X \times E_Y \times A \to \mathbb{R}^+$ a measurable function. We call β* Discount rate *and c* cost function*. Let further $(S_t, X_t, Y_t)_{t\geq 0}$ a PO-PDMP. For $\pi \in \Pi^P$, we define the* cost of policy π *under an initial observation $x \in E_X$ as*

$$J(x, \pi) := \mathbb{E}_x^\pi \left[\int_0^{T_\infty} e^{-\beta t} \int_A c(X_t^\pi, Y_t^\pi, a)\ \pi_t(da)\ dt \right], \tag{1.9}$$

where we use the notation $\mathbb{E}_x^\pi$ for the expectation under the probability measure $\mathbb{P}_x^\pi$ and X_t^π, Y_t^π denote the components for the observable and unobservable state of the derived π-controlled PO-PDMP.

The value function *of the control model gives the minimal cost under an initial observation $x \in E_X$ and is defined as*

$$J(x) := \inf_{\pi \in \Pi^P} J(x, \pi) \quad \forall\ x \in E_X. \tag{1.10}$$

The optimization problem *is then to find, for $x \in E_X$, a policy $\pi^\star \in \Pi^P$ such that we get*

$$J(x) = J(x, \pi^\star).$$

1.4.2 The solution approach

In order to prove the existence of optimal policies for the above defined time-continuous optimization problem under partial observation, we will follow a four steps approach:

Step 1: Reformulation of the time-continuous optimization problem into a time-discrete optimization problem, see Section 2.1

Step 2: Reformulation of this time-discrete but only partially observable optimization problem into a completely observable time-discrete problem, see Section 2.2

Step 3: Restriction of the history dependent control policies to Markovian control policies by proofing that this will not change the value of the value function $J(x)$, see Section 2.2.4

Step 4: Proof of existence of optimal policies for the equivalent completely observable time-discrete optimization problem under Markovian control policies, see Section 2.3

This approach is inspired from the general theory of optimal control of PDMPs under complete observation (Step 1 and 4) as well as from the general theory of optimal control of partially observed MDPs (Step 2 and 3). However, all of these steps had to be developed properly for this concrete model of a controlled PO-PDMP. The general theory for completely observable PDMPs does not use the space of observable histories we introduced, thus we had to adapt the approach used there for step 1 and step 4. The general theory for MDPs under partial observation requires a filter for the unobservable state of the process (step 2). This filter has to be developed for every concrete model and thus, we developed an adequate filter for an POMDP coming from a PO-PDMP. The filter we developed is largely inspired from the work of Brandejski et al. in [13]. Their approach, however, had to be combined with the concept of history dependent relaxed piecewise open loop policies and with a controlled PO-PDMP in general. Finally, we kept Brandejski's assumption of a finite set of post jump states in order to get a finite dimensional filter. For computational purposes as well as in view of possible applications of the theory, this assumption seemed most adequate to us.

Chapter 2

Reformulation of the problem and existence of optimal policies

This Chapter is at the core of this thesis. Having defined the model for a controlled PO-PDMP in the first Chapter, we will now approach the question of existence of optimal policies for the optimization problem stated at the end of Chapter 1. The general approach we will follow is quite standard for PDMPs: We will first reformulate the optimization problem for a PO-PDMP, i.e. for a stochastic process in continuous time into an optimization problem for a PO-MDP, that is for a partially observable Markov Decision Model, thus for a stochastic process in discrete time. This idea of reducing the continuous time problem to an MDP goes back to Yushkevich [61]. We will detail this step in Section 2.1. Although this approach is standard for PDMPs, some measurability questions arise in the case of a PO-PDMP where history dependent policies are used. One key result on this way will be the correspondence between continuous time policies (for the PO-PDMP) and discrete time policies (for the PO-MDP) as summarized in the Correspondence Theorem 2.11. We will fully develop all necessary steps up to an equivalent reformulation into a PO-MDP.

The second step is to pass from partial observation to complete observation, thus from a PO-MDP to a classical MDP. This step as such is also standard for PO-MDPs and presented in Section 2.2. In the concrete case here, however, we have to develop a so-called filter that is adapted to our concrete PO-PDMP model. Filtering is a classical technique and subject of standard books such as, e.g. [3], [36] or for the special case of Hidden Markov Models [31]. For the concrete case of our PO-PDMP model, however, we cannot simply apply any standard filter[1]. We have to develop an adequate filter for our concrete model. To do so, we follow the approach of Brandejski et al. [13]: In their paper, they develop a filter for an uncontrolled PO-PDMP under the assumption of only having a finite number of possible post-jump states. We now adapt their approach to the case of a controlled PO-PDMP. To the best of our knowledge, this filter has not yet been developed or published earlier and thus presents one of the main results of this thesis. Based on this filter, we can define a so-called derived filtered process, a completely observable MDP. We show equivalence of the corresponding optimization problem to the initial optimization problem for the PO-PDMP.

The third step is then, to prove existence of optimal policies for the optimization problem formulated for the derived filtered process of step 2. As we showed the equivalence of this optimization problem to the initial optimization problem for the PO-PDMP, existence of optimal policies for the initial problem follow. We show existence of optimal policies

[1] as it would be the case for a Hidden Markov Model with unknown jump intensity for example

for both time horizons considered in this thesis: A finite time horizon up to the N-th jump time of the PO-PDMP as well as for $T_\infty = \infty$, hence, for an infinite time horizon. Again, the approach used is quite standard in terms of tools from stochastic dynamic programming applied. Although, we have again to adapt earlier methodologies published for PDMPs under complete observation to the concrete setting of a controlled PO-PDMP with history dependent policies. The resulting existence results are new and have not been published earlier. We derive these results under additional measurability and continuity assumptions on the derived filtered process as summarized in Section 2.3.1. As these are assumptions made on the derived filtered process, we discuss how these assumptions translate into assumptions on the initial PO-PDMP. This discussion will be presented in Chapter 3.

2.1 First reformulation: From continuous time to discrete time

The goal of this section is to derive a formulation of a time-discrete optimization problem that is equivalent to the initial time-continuous optimization problem of Definition 1.37. More precisely, we aim to formulate a partially observable (PO) Markov Decision Process (MDP) having a value function $\tilde{J}$ that satisfies $\tilde{J}(x) = J(x)$ for all $x \in E_X$, where J as of (1.10). Such a reformulation would then allow to apply the full toolkit available from earlier publications on PO-MDPs to solve the optimization problem, i.e. to determine an optimal policy.

The theory of MDPs goes back to Bellmann [8] (for a reprint see [9]) and Howard [43]. Shiryaev [58] and Hinderer [41] contributed essentially to a more rigorous treatment. Bertsekas and Shreve [11] as well as Dynkin and Yushkevich [30] generalized the models further and investigated the basic measurability questions that arose in the context of this theory. More recent books providing a good introduction as well as overviews of the most important recent results are the books of Bäuerle and Rieder [6](with applications to Finance), Puterman [53] and Feinberg [32] (with recent state-of-the-art contributions).

We will develop, in the course of the following sections, all intermediate results necessary for the understanding of the proof of existence of optimal policies for our problem. Thus, the reader is not required to be knowledgeable about PO-MDP theory.

A proper definition of a PO-MDP contains the definitions of the state space, the action space, the set of admissible control actions, the transition kernel, the initial distribution of the unobservable state and the reward or cost function. All these points will be properly developed in this section, hence it is organized as follows: In Section 2.1.1, we motivate the attempt of reformulating the optimization problem as PO-MDP. Inspired from the fact that a PDMP can be described by its drift and its embedded time-discrete process, we develop a representation of the cost of a policy π that only depends on the post jump states $(X_{T_k}, Y_{T_k})_{k\geq 0}$, the inter jump times and the drift Φ^π of the π-controlled PO-PDMP. This observation leads to an investigation of the transition law for the (essential components of the) embedded time-discrete process of a π-controlled PO-PDMP in Section 2.1.2. In Section 2.1.3 we broach the issue, w.r.t. a proper reformulation as PO-MDP, of the admissible control policies $\pi \in \Pi^P$ depending on a continuous-time argument: They depend on the time since the last jump. This issue is resolved then in Section 2.1.4, where we define a slightly modified version of the embedded process: we endow it with modified admissible decision rules only depending on the observable parts of the discrete-time states of the process. Finally, we prove the equivalence of the so-created PO-MDP to the initial

optimization problem in Section 2.1.5.

2.1.1 Motivation and one period cost function

In Chapter 1, we explained how to describe a completely observable PDMP by its embedded marked point process. Looking closer at the definition of the *cost of a policy* π as given in (1.9), one will recognize, that $J(x,\pi)$ actually only depends on the time-discrete, embedded process $(S_k, X_{T_k}, Y_{T_k})_{k\geq 0}$ of the π-controlled PO-PDMP as well as on its probability distribution under the control policy π.

We will develop this properly and start with the definition of the one-period cost function:

Definition 2.1 (One period cost function). *For $x \in E_X, y \in E_Y$ and $[r] \in \mathcal{R}$ we define the undiscounted one period cost function g as*

$$g(x,y,r) := \mathbb{E}^r_{x,y}\left[\int_0^{T_1} e^{-\beta t}\int_A c(x,\Phi^r(y,t),a)\ r_t(da)\ dt\right], \tag{2.1}$$

and for $t \in \mathbb{R}^+$ we also define the discounted one period cost function G as

$$G(t,x,y,r) := e^{-\beta t}\ g(x,y,r). \tag{2.2}$$

Remark 2.2. *As a policy π together with an initial observation $x \in E_X$ define a relaxed control $\pi((0,x),\cdot) \in \mathcal{R}$ to be executed on $[0,T_1)$. We can thus use the notation $\mathbb{P}^r_x$ and $\mathbb{E}^r_x$ instead of $\mathbb{P}^\pi_x$ and $\mathbb{E}^\pi_x$ in the previous definition.*

Based on the one period cost function g, we can re-write the cost $J(x,\pi)$ of policy π as a function of the post-jump times T_k as well as of the post-jump states X_{T_k}, Y_{T_k} of the π-controlled PO-PDMP.

Lemma 2.3 (Post-jump state representation for cost of a policy). *Let $T_\infty = \infty$. Then, for $x \in E_X$ and $\pi \in \Pi^P$, the cost $J(x,\pi)$ of policy π can be written as*

$$J(x,\pi) = \mathbb{E}^\pi_x\left[\sum_{k=0}^{\infty} e^{-\beta T_k} g(X_{T_k}, Y_{T_k}, \pi_k(H_k,\cdot))\right] = \mathbb{E}^\pi_x\left[\sum_{k=0}^{\infty} G(T_k, X_{T_k}, Y_{T_k}, \pi_k(H_k,\cdot))\right].$$

Proof. Based on the definition of the one period cost function we apply rules for iterated conditional expectations and obtain:

$$\begin{aligned}
J(x,\pi) &= \mathbb{E}^\pi_x\left[\int_0^\infty e^{-\beta t}\int_A c(X_t,Y_t,a)\ \pi_t(da)\ dt\right] \\
&= \int_{E_Y} \mathbb{E}^\pi_{x,y}\left[\int_0^\infty e^{-\beta t}\int_A c(X_t,Y_t,a)\ \pi_t(da)\ dt\right]\ Q_0(x;dy) \\
&= \int_{E_Y} \mathbb{E}^\pi_{x,y}\left[\sum_{k=0}^{\infty} e^{-\beta T_k} g(X_{T_k},Y_{T_k},\pi_k(H_k,\cdot))\right]\ Q_0(x;dy) \\
&= \mathbb{E}^\pi_x\left[\sum_{k=0}^{\infty} e^{-\beta T_k} g(X_{T_k},Y_{T_k},\pi_k(H_k,\cdot))\right].
\end{aligned}$$

The third equality above is explained as follows, where we apply (1.6), and denote by $\mathcal{F}_{T_k}$ the full (observable and unobservable states) filtration up to time T_k:

$$
\begin{aligned}
&\mathbb{E}^{\pi}_{x,y}\left[\int_0^{\infty} e^{-\beta t}\int_A c\left(X_t,Y_t,a\right)\ \pi_t(da)\ dt\right]\\
&=\mathbb{E}^{\pi}_{x,y}\left[\sum_{k=0}^{\infty}\int_{T_k}^{T_{k+1}} e^{-\beta t}\int_A c\left(X_{T_k},\Phi^{\pi_k(H_k,\cdot)}(Y_{T_k},t-T_k),a\right)\ \pi_k(H_k,t-T_k)(da)\ dt\right]\\
&=\mathbb{E}^{\pi}_{x,y}\left[\sum_{k=0}^{\infty}\int_{0}^{T_{k+1}-T_k} e^{-\beta T_k}e^{-\beta t}\int_A c\left(X_{T_k},\Phi^{\pi_k(H_k,\cdot)}(Y_{T_k},t),a\right)\ \pi_k(H_k,t)(da)\ dt\right]\\
&=\mathbb{E}^{\pi}_{x,y}\left[\sum_{k=0}^{\infty} e^{-\beta T_k}\ \mathbb{E}^{\pi}_{x,y}\left[\int_{0}^{T_{k+1}-T_k} e^{-\beta t}\int_A c\left(X_{T_k},\Phi^{\pi_k(H_k,\cdot)}(Y_{T_k},t),a\right)\pi_k(H_k,t)(da)dt|\mathcal{F}_{T_k}\right]\right]\\
&=\mathbb{E}^{\pi}_{x,y}\left[\sum_{k=0}^{\infty} e^{-\beta T_k}\ \mathbb{E}^{\pi_k(H_k,\cdot)}_{X_{T_k},Y_{T_k}}\left[\int_0^{T_1} e^{-\beta t}\int_A c\left(X_{T_k},\Phi^{\pi_k(H_k,\cdot)}(Y_{T_k},t),a\right)\ \pi_k(H_k,t)(da)\ dt\right]\right]\\
&=\mathbb{E}^{\pi}_{x,y}\left[\sum_{k=0}^{\infty} e^{-\beta T_k} g\left(X_{T_k},Y_{T_k},\pi_k(H_k,\cdot)\right)\right]
\end{aligned}
$$

Applying the definition of G leads to the result. □

In view of this last result, we are very close to the formulation of a PO-MDP having the same value function as the initial optimization problem of Definition 1.37. Taking the one period cost function g as cost function for a PO-MDP, the representation of $J(x,\pi)$ of Lemma 2.3 already looks very much like a typical „cost under policy π“ of a π-controlled PO-MDP with time-discrete states $(X_{T_k},Y_{T_k})_{k\geq 0}$. Thus, it looks like a PO-MDP with non-uniform time steps (compare, e.g., equation (5.2) in [6]). In the following two section, we will therefore shed a closer look on the transition law of as well as on admissible decision rules for the controlled process $(X_{T_k},Y_{T_k})_{k\geq 0}$.

2.1.2 The embedded time-discrete process of a controlled PO-PDMP

Given Definition 1.34 of a π-controlled PO-PDMP, its embedded process could be understood as the process with states $(k,0,X_{T_k},Y_{T_k},H_k)$ at the k-th jump time. Note that the second component must be 0 as right at the jump time T_k, the time elapsed since the last jump is 0. As we have seen in the previous section, the post-jump state representation of the cost of policy π only depends on T_k, X_{T_k}, Y_{T_k} and H_k. As further, $T_k=\sum_{i=0}^{k} S_i$ (k-th jump time is sum of previous inter-jump stimes), and, as H_k only depends on[2] $(S_0,X_0,\ldots,S_k,X_{T_k})$, it is enough to further investigate the time-discrete process $(S_k,X_{T_k},Y_{T_k})_{k\in\mathbb{N}_0}$. To simplify notations, we will write X_k and Y_k instead of X_{T_k} and Y_{T_k} in the sequel.

Lemma 2.4. *Let $n\in\mathbb{N}_0$. The transition law to stage $n+1$, given the history up to stage n, of the embedded time-discrete process $(S_k,X_k,Y_k)_{k\in\mathbb{N}_0}$ of the π-controlled PO-PDMP*

[2] see equation (1.6) and explanations thereafter

$(S_t, X_t, Y_t)_{t\in\mathbb{R}^+}$ is given by

$$
\begin{aligned}
&Q^{\pi,n}_{SXY}([0,t]\times B_X\times B_Y \mid s_0,x_0,y_0,\dots,s_n,x_n,y_n) = \\
&\int_0^t \exp\left(-\int_0^{s'}\int_A \lambda^A(\Phi^{\pi_n(h_n,\cdot)}(y_n,\tau),a)\ \pi_n(h_n,\tau)(da)\ d\tau\right)\int_A\int_{B_Y} Q_\epsilon(B_X-\psi(y'))\cdots \\
&\qquad\cdots\lambda^A\left(\Phi^{\pi_n(h_n,\cdot)}(y_n,s'),a\right)\ Q^A\left(\Phi^{\pi_n(h_n,\cdot)}(y_n,s'),a;dy'\right)\ \pi_n(h_n,s')(da)\ ds',
\end{aligned}
$$

where $B_X\in\mathcal{B}(E_X)$ and $B_Y\in\mathcal{B}(E_Y)$.

Proof. Knowing the full history of the states $s_0,x_0,y_0,\dots,s_n,x_n,y_n$ of the embedded process allows to reconstruct the full path of all components of the π-controlled PO-PDMP $(N_t,S_t,X_t,Y_t,H_t)_{0\le t\le T_n}$. This is equivalent to knowing the full filtration $\mathcal{F}_{T_n}$ of the π-controlled PO-PDMP. Thus, the transition law satisfies

$$
\begin{aligned}
&Q^{\pi,n}_{SXY}([0,t]\times B_X\times B_Y \mid S_0,X_0,Y_0,\dots,S_n,X_n,Y_n) = \\
&\qquad\mathbb{P}^\pi_x(N_{T_{n+1}}=n+1, S_{n+1}\le t, X_{n+1}\in B_X, Y_{n+1}\in B_Y, H_{n+1}\in\mathcal{H}_{n+1}\mid\mathcal{F}_{T_n}).
\end{aligned}
$$

The latter can be calculated using the conditional density $f^\pi_{S_{n+1}}(s'\mid\mathcal{F}_{T_n})$ of the $(n+1)$-th inter jump time S_{n+1} given $\mathcal{F}_{T_n}$ under policy π (see Definition 1.1 part (vii) and Lemma 1.35) as:

$$
\begin{aligned}
&\mathbb{P}^\pi_x(N_{T_{n+1}}=n+1, S_{n+1}\le t, X_{n+1}\in B_X, Y_{n+1}\in B_Y, H_{n+1}\in\mathcal{H}_{n+1}\mid\mathcal{F}_{T_n}) \\
&=\int_0^t \mathbb{P}^\pi_x\left(N_{T_{n+1}}=n+1, X_{n+1}\in B_X, Y_{n+1}\in B_Y, H_{n+1}\in\mathcal{H}_{n+1}\mid\mathcal{F}_{T_n}, S_{n+1}=s'\right)\cdots \\
&\cdots\mathbb{P}^\pi_x(S_{n+1}\in ds'\mid\mathcal{F}_{T_n})\ ds' \\
&=\int_0^t Q^\pi\left((n,s',X_n,\Phi^{\pi_n(H_n,\cdot)}(Y_n,s'),H_n);\{n+1\}\times\{0\}\times B_X\times B_Y\times\mathcal{H}_{n+1}\right)\cdots \\
&\cdots f^\pi_{S_{n+1}}(s'\mid\mathcal{F}_{T_n})\ ds' \\
&=\int_0^t \exp\left(-\Lambda^\pi(n,Y_n,H_n,s')\right)\int_A\int_{B_Y} Q_\epsilon\left(B_X-\psi(y')\right)\ \lambda^A\left(\Phi^{\pi_n(H_n,\cdot)}(Y_n,s'),a\right)\cdots \\
&\cdots Q^A\left(\Phi^{\pi_n(H_n,\cdot)}(Y_n,s'),a;dy'\right)\ \pi_n(H_n,s')(da)\ ds'
\end{aligned}
$$

Applying the definition of Λ^π terminates the proof. □

Corollary 2.5. *The transition laws $Q^{\pi,n}_{SXY}$ of the embedded process of the π-controlled PO-PDMP only depend on the last post-jump state Y_n of the unobservable process (Y_t), not on the full history of $Y_0,\dots,Y_n$, i.e.*

$$
Q^{\pi,n}_{SXY}(\cdots\mid s_0,x_0,y_0,\dots,s_n,x_n,y_n) = Q^{\pi,n}_{SXY}(\cdots\mid s_0,x_0,\dots,s_{n-1},x_{n-1},s_n,x_n,y_n).
$$

Proof. Follows directly from previous lemma as only y_n intervenes in the formula for the transition law. The full history of $s_0,x_0,\dots,s_n,x_n$, however, is necessary to construct h_n used in $\pi_n(h_n)$. □

The previous Corollary can be seen as the counterpart of Corollary 1.7: In the case of an uncontrolled completely observable PDMP, we obtained a discrete-time Markov chain as embedded process. Here, under partial observation, we do not obtain a Markovian structure for the embedded process of a controlled PO-PDMP as the transition law depends

on the selected relaxed control. The decision which relaxed control to execute from stage n to $n+1$ is taken according to π_n which itself depends on the observable history up to stage n.

The embedded π-controlled process $(S_k, X_k, Y_k)_{k\in\mathbb{N}_0}$ induces a probability measure on $(\mathbb{R}^+ \times E_X \times E_Y)^\infty$: According to Ionescu-Tulceas theorem, the transitions laws $(q^{\pi,n}_{SXY})_{n\geq 0}$ together with the initial observation $x \in E_X$ as well as with the initial conditional distribution[3] $Q_0(x,\cdot)$ of Y_0 define a probability measure on $(\mathbb{R}^+ \times E_X \times E_Y)^\infty$.

As long as the context is clear, we will denote this probability measure as well with $\mathbb{P}^\pi_x$ and the corresponding expectation with $\mathbb{E}^\pi_x$.

2.1.3 The principal reformulation issue

Although Lemma 2.3 provides a representation of the cost $J(x,\pi)$ of policy π that looks like a function of the embedded time-discrete process[4] investigated in Lemma 2.4, the reformulation into an equivalent classical PO-MDP is slightly more difficult.

The principal issue is the following: For a classical PO-MDP, history dependent decision rules have to be measurable functions of the observable history $\mathcal{H}_n$ only. Our decision rules for the control of the time-continuous PO-PDMP, however, have a time argument and thus are measurable functions π_n on $\mathcal{H}_n \times [0,\infty)$. Thus, to fully transform our problem into a PO-MDP, we somehow have to „get rid“ of this time component as argument of the decision rule. The principal idea is to go from measurable mappings $\pi_n : \mathcal{H}_n \times \mathbb{R}^+ \to \mathbf{P}(A)$ to measurable mappings $\pi^D_n : \mathcal{H}_n \to \mathcal{R}$. Although this seems to be the „canonical“ approach, it is not trivial as some measurability questions arise that have to be solved.

Hence, the outline of the next two sections is the following: We will define a PO-MDP with suitable admissible decision rules in Section 2.1.4. Finally, we will prove the equivalence of the resulting optimization problem for this PO-MDP and the initial continuous-time optimization problem of our PO-PDMP in Section 2.1.5.

2.1.4 The pseudo-embedded π^D-controlled process

The goal of this section is to define a controlled time-discrete stochastic process as well as a set of admissible decision rules forming together a PO-MDP. This PO-MDP shall be our candidate for the formulation of an equivalent optimization problem in discrete time. In order to satisfy this equivalence condition, it would seem at least promising (if not necessary) if the probability distribution on $(\mathbb{R}^+ \times E_X \times E_Y)^\infty$ induced by this PO-MDP was the same as the probability distribution induced by the embedded time-discrete process $(S_k, X_{T_k}, Y_{T_k})_{k\in\mathbb{N}_0}$ of an „equivalently controlled“ PO-PDMP. The meaning of „equivalently controlled“ will be detailed later.

We start with the definition of a *pseudo-embedded* $\mathcal{R}$-controlled process, that is, a time-discrete process on $\mathbb{R}^+ \times E_X \times E_Y$ which is controlled by some $[r] \in \mathcal{R}$. In a next step, we will then detail how the selection of such controls shall be executed based on decision rules for each stage of the process.

The term „pseudo-embedded“ will be explained in detail below. The basic idea is to endow this process with a transition law that is „basically the same“ as the transition law of an embedded discrete time process of an $[r]$-controlled PO-PDMP.

Definition 2.6. *For an initial observation $x \in E_X$ and an initial conditional distribution $Q_0(x,\cdot)$ of Y_0, a* pseudo-embedded $\mathcal{R}$-controlled process *is a time-discrete stochastic*

[3] note that by definition, we have $S_0 = 0$ in our model, so to be correct, we shall also mention δ_0 as initial distribution for S_0

[4] remember $T_k = \sum_{n=0}^k S_n$

process $(\tilde{S}_k, \tilde{X}_k, \tilde{Y}_k)_{k\geq 0}$ *on the state space* $E_{SXY} := \mathbb{R}^+ \times E_X \times E_Y$ *with the following properties:*

(i) The process starts in $(0, x, \tilde{Y}_0)$ *where the distribution of* $\tilde{Y}_0$ *is given by* $Q_0(x, \cdot)$.

(ii) For a chosen relaxed control action $[r] \in \mathcal{R}$*, the transition law* $\tilde{Q}_{SXY}$ *of the process is given by*

$$\begin{aligned}\tilde{Q}_{SXY}\left([0,t] \times B_X \times B_Y \mid s_n, x_n, y_n[r]\right) := \\ \int_0^t \exp\left(-\int_0^{s'} \int_A \lambda^A(\Phi^r(y_n,\tau), a)\ r_\tau(da)\ d\tau\right) \int_A \int_{B_Y} Q_\epsilon\left(B_X - \psi(y')\right) \\ \lambda^A\left(\Phi^r(y_n, s'), a\right)\ Q^A\left(\Phi^r(y_n, s'), a; dy'\right)\ r_{s'}(da)\ ds',\end{aligned}$$

where $B_X \in \mathcal{B}(E_X)$ *and* $B_Y \in \mathcal{B}(E_Y)$.

Lemma 2.7. *The transition laws of the pseudo-embedded* $\mathcal{R}$*-controlled process are well-defined, i.e.*

$$\tilde{Q}_{SXY}([0,t] \times B_X \times B_Y \mid s_n, x_n, y_n, [r])$$

is independent from the choice of a representative of $[r]$.

Proof. Follows from the fact that in the definition of $\tilde{Q}_{SXY}$, the integral w.r.t. $r_\tau(da)$ appears within an integral w.r.t. $d\tau$ and the integral w.r.t. $r_{s'}(da)$ appears within an integral w.r.t. ds' as well as from the fact that $\Phi^{r'}(y_n, \tau) = \Phi^r(y_n, \tau)$ for λ^1-almost all τ for $r' \in [r]$ (see definition of controlled drift). □

The question is now how to select, for each stage n of the process, a control $[r] \in \mathcal{R}$ to execute. In order to define a partially observable MDP that shall be the candidate for an equivalent formulation of our initial optimization problem, we will use the following decision rules:

Definition 2.8. *A* time-discrete history dependent relaxed control policy *is a sequence* $\pi^D := (\pi_0^D, \pi_1^D, \dots)$ *of time-discrete history dependent decision rules defined as follows:*

(i) The spaces of observable histories remain defined as previously introduced as $\mathcal{H}_0 := \mathbb{R}^+ \times E_X$ *and for* $n \geq 1$ *we set* $\mathcal{H}_n := \mathcal{H}_{n-1} \times \mathcal{R} \times \mathbb{R}^+ \times E_X$.

(ii) For $n \geq 0$*, a* time discrete history dependent decision rule at stage n *is defined as a measurable mapping*

$$\pi_n^D : \mathcal{H}_n \to \mathcal{R} \tag{2.3}$$

We write Π_n^D *for the set of all time discrete history dependent decision rules at stage* n *and define the set of all time-discrete history dependent relaxed control policies as* $\Pi^D := \times_{n\geq 0} \Pi_n^D$.

An observable history at stage n is thus an $\mathcal{H}_n$-valued random vector

$$\tilde{H}_n = (\tilde{S}_0, \tilde{X}_0, \tilde{R}_0, \dots, \tilde{S}_{n-1}, \tilde{X}_{n-1}, \tilde{R}_{n-1}, \tilde{S}_n, \tilde{X}_n)$$

and under a given control policy π^D, we define $\tilde{R}_n := \pi_n^D(\tilde{H}_n)$ for $n \geq 0$.

Definition 2.9. *For $x \in E_X$ and $Q_0(x,\cdot)$ an initial conditional distribution of Y_0, a* pseudo-embedded π^D-controlled process *is a pseudo-embedded $\mathcal{R}$-controlled process where at each stage n of the process, the control $[r]$ to execute is selected according to the decision rule π_n^D. Its transition laws are thus given by*

$$\tilde{Q}_{SXY}^{\pi^D,n}([0,t]\times B_X\times B_Y \mid s_0,x_0,y_0,\dots,s_n,x_n,y_n) := \\ \tilde{Q}_{SXY}\left([0,t]\times B_X\times B_Y \mid s_n,x_n,y_n,[\pi_n^D(h_n)]\right).$$

Remark 2.10. *As only $s_0,x_0,\dots,s_n,x_n$ are necessary for the construction of h_n under π^D, the transition laws of a pseudo-embedded π^D-controlled process at stage n do not depend on $y_0,\dots,y_{n-1}$.*

According to Ionescu-Tulcea's theorem, the transition probabilities $\tilde{Q}_{SXY}^{\pi^D,n}$ together with the initial distribution $Q_0(x,\cdot)$ of $\tilde{Y}_0$ define a probability measure $\mathbb{P}_x^{\pi^D}$ on $(\mathbb{R}^+\times E_X\times E_Y)^\infty$ for any given initial observation $x\in E_X$. We will denote the expectation w.r.t. this probability measure by $\mathbb{E}_x^{\pi^D}$.

2.1.5 Equivalent time-discrete optimization problem under partial observation

In this Section, we will prove that we can formulate an optimization problem for a pseudo-embedded π^D-controlled process that is equivalent to the initial continous-time optimization problem of Definition 1.37.

The first result is now crucial in two senses: First, it further explains why we call the above defined process „pseudo-embedded“ process and second, it is at the core of the resolution of the earlier stated reformulation issue.

Theorem 2.11 (Correspondence theorem). *Let $n\in\mathbb{N}_0$. For every $\pi_n^P\in\Pi_n^P$ there exists $\pi_n^D\in\Pi_n^D$ such that*

$$\pi_n^P(h_n,\cdot)=\pi_n^D(h_n)(\cdot)\quad \lambda^1\text{−a.e. on }\mathbb{R}^+\ \ \forall h_n\in\mathcal{H}_n \tag{2.4}$$

and vice-versa.

Proof. Looks like a trivial statement but we need to prove measurability requirements which we do in detail in the Annex, see proof of Theorem A.27. □

With this correspondence theorem in mind, it becomes clear what we meant with „equivalently controlled“ process in the introduction to the previous section.

The λ^1-a.e. equality in the correspondence theorem basically states that $\pi_n^P(h_n,\cdot)=\pi_n^D(h_n)(\cdot)$ as equality of λ^1-equivalence classes in $\mathcal{R}$. As previously shown in Lemma 2.7, the transition law $\tilde{q}_{SXY}$ of the $\mathcal{R}$-controlled pseudo-embedded process is independent from the choice of a representative of the executed relaxed control $[r]\in\mathcal{R}$ and thus, the next result is an immediate consequence of this invariance and the definition of a pseudo-embedded process.

Corollary 2.12. *Let π^P the corresponding history dependent relaxed piecewise open loop policy to π^D according to the correspondence theorem. Then, for all $n\in\mathbb{N}_0$, for all*

$(s_0, x_0, \dots, s_n, x_n) \in (\mathbb{R}^+ \times E_X)^n$ and $y_n \in E_Y$, the following equality of transition laws holds:

$$\tilde{Q}_{SXY}^{\pi^D,n}([0,t] \times B_X \times B_Y \mid s_0, x_0, \dots s_n, x_n, y_n) = Q_{SXY}^{\pi^P,n}([0,t] \times B_X \times B_Y \mid s_0, x_0, \dots s_n, x_n, y_n).$$

With other words: the transition laws of the embedded process of the π^P-controlled PO-PDMP are the same as the transition laws of the pseudo-embedded π^D-controlled time-discrete process.

Proof. Follows immediately from the correspondence theorem and Lemma 2.7: By the correspondence theorem, we get $\pi_0^P((s_0,x_0),\cdot) = \pi_0^D((s_0,x_0))(\cdot)$ as equality in $\mathcal{R}$ and thus, $h_1^{\pi^P} = (s_0, x_0, \pi_0^P((s_0,x_0),\cdot), s_1, x_1) = (s_0, x_0, \pi_0^D((s_0,x_0))(\cdot), s_1, x_1) = h_1^{\pi^D}$. A simple induction shows that for all $k \leq n$ we have $h_k^{\pi^P} = h_k^{\pi^D}$. Consequently, we have $\pi_n^P(h_n^{\pi^P},\cdot) = \pi_n^D(h_n^{\pi^D})(\cdot)$ as equality in $\mathcal{R}$ and the result follows from Lemma 2.7. □

Corollary 2.13. *On $(\mathbb{R}^+ \times E_X \times E_Y)^\infty$, the probability measure induced by the embedded process of the π^P-controlled PO-PDMP coincides with the probability measure induced by the pseud-embedded π^D-controlled process, i.e.*

$$\mathbb{P}_x^{\pi^P} = \mathbb{P}_x^{\pi^D}.$$

Proof. Follows from previous corollary and Ionescu-Tulcea's theorem. □

The definition of π^D-controlled pseudo-embedded time-discrete processes induced a probability measure $\mathbb{P}_x^{\pi^D}$ on $(\mathbb{R}^+ \times E_X \times E_Y)^\infty$. The correspondence theorem then is at the core of the fact that this probability measure actually coincides with the probability measure $\mathbb{P}_x^{\pi^P}$ induced by the embedded process of the π^P-controlled PO-PDMP. It is acutally corollary 2.12 that illustrates why we have chosen the name "pseudo-embedded processes": They have the same transition laws as embedded processes of with corresponding policies controlled PO-PDMPs.

These are all ingredients we need for the complete reformulation into a classical time-discrete optimal control problem under partial observation as known from the theory of PO-MDPs. We will now first define an optimal control problem under partial observation for a π^D-controlled pseudo-embedded process $(\tilde{S}_n, \tilde{X}_n, \tilde{Y}_n)_{n\geq 0}$. Then, we will prove the equivalence of this problem to our initial optimal control problem for the π^P-controlled PO-PDMP where π^P is the corresponding policy to π^D, in the sense of the correspondence theorem.

Definition 2.14. *For a policy $\pi^D \in \Pi^D$, an initial observation $x \in E_X$ together with an initial conditional distribution $Q_0(x,\cdot)$ and a π^D-controlled pseudo-embedded process $(\tilde{S}_n, \tilde{X}_n, \tilde{Y}_n)_{n\geq 0}$ we define the cost of policy π^D as*

$$\tilde{J}(x,\pi^D) := \mathbb{E}_x^{\pi^D}\left[\sum_{k=0}^{\infty} e^{-\beta\sum_{n=0}^k \tilde{S}_n} g(\tilde{X}_k, \tilde{Y}_k, \pi_k^D(\tilde{H}_k))\right] = \mathbb{E}_x^{\pi^D}\left[\sum_{k=0}^{\infty} G(\sum_{n=0}^k \tilde{S}_n, \tilde{X}_k, \tilde{Y}_k, \pi_k^D(\tilde{H}_k))\right],$$

where g and G are as of Definition 2.1 (remember $T_1 = S_1$).

The value function *of the time-discrete control model gives the minimal cost under an initial observation $x \in E_X$ and is defined as*

$$\tilde{J}(x) := \inf_{\pi^D \in \Pi^D} \tilde{J}(x,\pi^D) \quad \forall\, x \in E_X. \tag{2.5}$$

The time-discrete optimization problem *is then to find, for $x \in E_X$, a policy $\pi^{\star D} \in \Pi^D$ such that we get*

$$\tilde{J}(x) = \tilde{J}(x, \pi^{\star D}).$$

The time-discrete optimization problem defined, we can now turn to the main result of this section: The proof of the equivalence of the time-discrete optimization problem to our initial, time-continuous optimization problem for the controlled PO-PDMP.

Proposition 2.15. *Let $x \in E_X$ an initial observation, $\pi^P \in \Pi^P$ a history dependend relaxed piecewise open loop control policy for the PO-PDMP and $\pi^D \in \Pi^D$ its corresponding time-discrete policy according to the correspondence theorem. Then, it holds*

$$J(x, \pi^P) = \tilde{J}(x, \pi^D).$$

Proof. As shown in lemma 2.3 we can write the cost of policy π^P as

$$J(x, \pi^P) = \mathbb{E}_x^{\pi^P}\left[\sum_{k=0}^{\infty} e^{-\beta T_k} g(X_{T_k}, Y_{T_k}, \pi_k^P(H_k, \cdot))\right].$$

As g is independent from the choice of a representative of $[\pi_k^P(H_k, \cdot)] \in \mathcal{R}$, we can replace $\pi_k^P(H_k, \cdot)$ by $\pi_k^D(H_k)(\cdot)$ in the argument of g.

Further, we showed that $\mathbb{P}_x^{\pi^P} = \mathbb{P}_x^{\pi^D}$ on $(\mathbb{R}^+ \times E_X \times E_Y)^\infty$ and thus, we can replace $\mathbb{E}_x^{\pi^P}$ by $\mathbb{E}_x^{\pi^D}$ while at the same time replacing $T_k, X_{T_k}, Y_{T_k}, H_k$ (the states and the history of the embedded process of the π^P-controlled PO-PDMP) by $\tilde{T}_k, \tilde{X}_k, \tilde{Y}_k, \tilde{H}_k$ (the states of the pseudo-embedded π^D-controlled process in discrete time).

Finally, applying $\tilde{T}_k = \sum_{n=0}^k \tilde{S}_n$, we get

$$\begin{aligned} J(x, \pi^P) &= \mathbb{E}_x^{\pi^D}\left[\sum_{k=0}^{\infty} e^{-\beta \sum_{n=0}^k \tilde{S}_n} g(\tilde{X}_k, \tilde{Y}_k, \pi_k^D(\tilde{H}_k)(\cdot))\right] \\ &= \tilde{J}(x, \pi^D), \end{aligned}$$

and the definition of G leads to the result. □

With this result and the correspondence theorem in mind, our solution strategy going forward is now: Solve the optimal control problem for the pseudo-embedded π^D-controlled time-discrete process which shall provide an optimal control $\pi^{\star D} \in \Pi^D$. The corresponding history dependent relaxed piecewise open loop policy $\pi^{\star P} \in \Pi^P$ is then an optimal control for the time-continuous optimization problem of the π^P-controlled PO-PDMP.

Having this in mind, we can simplify notations from now on: we will only write $J(x, \pi)$ and π instead of specifying by superscript P or D if we mean π^P or the corresponding π^D together with $\tilde{J}$. The context shall always be clear such that the reader should know if we talk about the embedded process of the controlled PO-PDMP or about the corresponding pseudo-embedded controlled process.

The next step of our solution strategy is the reformulation of our optimization problem into a fully observable time-discrete optimization problem, i.e. passing from a PO-MDP to a MDP under complete observation.

2.2 Second reformulation: From partial to complete observation

This Section is dedicated to steps 2 and 3 of the approach outlined at the end of Chapter 1. The goal of these two steps is, to pass from an optimization problem for a PO-MDP (see previous section) to one for an MDP. With other words, to pass from partial to complete observation of the time-discrete stochastic process that is underlying to the optimization problem formulated. Again, the aim is to get a resulting optimization problem still being equivalent to the original optimization problem.

With respect to the general theory for PO-MDPs, this is a standard approach involving filter techniques: One basically replaces the unobservable component of a state of the process by the conditional distribution of the unobservable state given all information observable up to the current stage. This conditional distribution is *observable* as it can be calculated by an iterative process, involving, at every stage of the process, the previous conditional distribution calculated as well as the new observable information obtained at the current stage of the process. Important is that we apply the so-called *separation principle of estimation and control*, i.e. we first observe the new observable information, then calculate the new conditional distribution to finally select the control action to execute based on this information.

Filtering is a classical technique and subject of standard books such as, e.g. [3], [36] or for the special case of Hidden Markov Models [31]. For the concrete case of our PO-PDMP model, however, we cannot simply apply any standard filter (as it would be the case for a Hidden Markov Model with unknown jump intensity for example). We have to develop an adequate filter and to do so, we follow the approach of Brandejski et al. [13]: In their paper, they develop a filter for an uncontrolled PO-PDMP under the assumption of only having a finite number of possible post-jump states. In view of possible applications, we hold this assumption for meaningful[5]. In view of computational aspects, this assumption is no significant limitation of the model. One would always have to pass by a discretization of the filter in order to apply numerical methods for calculations of optimal policies. Thus, we keep this assumption and will only investigate the resulting *finite dimensional case* in the sequel. As Brandejski et al. have developped their filter only for an uncontrolled PO-PDMP, we have to develop the adequate filter for our controlled model step by step, although our approach will follow theirs in the great lines.

Once the filter is developed, we can pass to the definition of a so-called derived filtered model and formulate the corresponding optimization problem. In order to apply standard techniques to characterize the value function of the derived filtered model, we have to restrict the admissible control policies to Markov policies, i.e. we no longer allow history dependent policies. A fundamental result from Hinderer (1970) [41] finally shows that this restriction will not prevent the resulting MDP under complete observation to be equivalent, in terms of the optimization problem then formulated, to the original optimization problem for the controlled PO-PDMP.

The outline of this Section is thus as follows: In Section 2.2.1 we will briefly state the model assumptions for the finite dimensional case before we develop the filter under these assumptions in Section 2.2.2. We then formulate the optimization problem for the derived filtered process in Section 2.2.3 to finally show, in Section 2.2.4, how to restrict the problem to Markov policies and at the same time still having an equivalent optimization problem.

[5] see also Section 6.1.1 for a more detailed discussion of this aspect

2.2.1 The finite dimensional case: Model assumptions

We will concentrate our investigations throughout the next sections on the so-called *finite dimensional case*. That is a case, where the set of possible post-jump states of the unobservable process $(\tilde{Y}_t)_{t\geq 0}$ is of finite cardinality. Concretely, in addition to all previously made model assumptions, we will assume the following whenever we refer to the *finite dimensional case*:

Assumption 2.16 (Finite set of possible post-jump states). *We assume the set E_Y^0 of possible post-jump states to be finite, i.e. $\exists q \in \mathbb{N} : E_Y^0 := \{y^1, \dots, y^q\} \subset E_Y$ and $Q_Y^A(y, a; E_Y^0) = 1$ for all $y \in E_Y, a \in A$. We further assume $Y_0 \in E_Y^0$.*

Remark 2.17. *Note that in the sequel, we will denote elements of E_Y^0 by y^i, i.e. we will make use of superscripts to emphasize that we talk about post-jump states in E_Y^0. A state at stage n will still be denoted by y_n to emphasize that the point in time, hence stage, is of major interest here. Very often, we will try to follow the convention to denote by y^i a current post-jump state and by y^j the following post-jump state.*

Definition 2.18 (Probability measures on E_Y^0). *We denote by $\mathbf{P}(E_Y^0)$ the set of probability measures on E_Y^0.*

Assumption 2.19 (Initial conditional distribution of Y_0). *We assume the initial conditional distribution of Y_0 given the observation of $X_0 = x$ to be given by some distribution $Q_0(x; \cdot) \in \mathbf{P}(E_Y^0)$.*

Remark 2.20 (Generalization of model in [13]). *We shall remark here, that we have chosen to work under a slightly more general model than the one of [13]. Brandejski et al. assume perfect observation of the initial state y_0, i.e. $x_0 = \psi(y_0)$ with ψ a bijection. Therefore, the initial observation of $X_0 = x$ leads directly to a Dirac distribution for the initial conditional distribution of Y_0 given $X_0 = x$, i.e. $Q_0(x, \{y^j\}) = \mathbb{P}_x^\pi(Y_0 = y^j) = \delta_{\psi^{-1}(x_0)}(y^j)$ independent of the initial unconditional distribution of Y_0. As we do not assume perfect observation of the initial state Y_0, we need to work under an assumption for the initial conditional distribution Q_0 as outlined previously.*

2.2.2 The finite dimensional filter

The goal of this section is to develop a recursive formulation for the conditional distribution of $\tilde{Y}_n$ given the observed history $\tilde{H}_n$ where $\tilde{Y}_n$ is the unobservable state of a pseudo-embedded π-controlled process $(\tilde{S}_k, \tilde{X}_k, \tilde{Y}_k)_{k\geq 0}$. This conditional distribution is called *filter* and and we will develop the recursive formulation for the finite dimensional case defined in the previous section.

Remember the definition of the random vector $\tilde{H}_n = (\tilde{S}_0, \tilde{X}_0, \tilde{R}_0, \dots, \tilde{S}_n, \tilde{X}_n)$ where $\tilde{R}_n = \pi_n(\tilde{H}_n)$ for a π-controlled process. We then define:

Definition 2.21 (Filter M_n). *Let $\pi \in \Pi^D$ and $(\tilde{S}_k, \tilde{X}_k, \tilde{Y}_k)_{k\geq 0}$ a π-controlled pseudo-embedded process with initial transition kernel Q_0 from E_X to $\mathbf{P}(E_Y^0)$. For an initial observation $x \in E_X$ and $n \in \mathbb{N}_0$ we define the filter M_n^{π,x,Q_0} as $\mathbf{P}(E_Y^0)$-valued random vector by*

$$\forall\, i \in \{1, \dots, q\} : \qquad M_n^{\pi,x,Q_0,i} := M_n^{\pi,x,Q_0}(\{y^i\}) := \mathbb{E}_x^\pi[\mathbb{1}_{\{\tilde{Y}_n^\pi = y^i\}} |\, \sigma(\tilde{H}_n)]. \tag{2.6}$$

To simplify notations, we will only write $M_n^{\pi,j}$ instead of M_n^{π,x,Q_0} in the sequel, as long as out of the context, it is clear which initial observation x and conditional distribution Q_0 we refer to.

Remark 2.22 (Initial Filter M_0). *By definition of the filter and the previously outlined construction of the probability measure $\mathbb{P}_x^\pi$, it is clear that the initial filter M_0^{π,x,Q_0} is given for $i \in \{1,\dots,q\}$ by*

$$M_0^{\pi,x,Q_0,i} = \mathbb{E}_x^\pi\left[\mathbb{1}_{\{Y_0^\pi = y^i\}} \mid \sigma(\tilde{S}_0, \tilde{X}_0)\right] = \delta_0(\tilde{S}_0)\cdot Q_0(\tilde{X}_0, \{y^i\}). \tag{2.7}$$

Remark 2.23 (Factorization of filter). *The filter M_n^π being a $\sigma(\tilde{H}_n)$-measurable random variable, there exists for all $n \in \mathbb{N}_0$ a measurable function $\mu_n^\pi : \mathcal{H}_n \to \mathbf{P}(E_Y^0)$ such that $\mathbb{P}_x^\pi$-a.s., we have $M_n^\pi = \mu_n^\pi \circ H_n^\pi$.*

The goal of this section is now to develop a recursive formula for M_n^π as a function of M_{n-1}^π. Informally speaking, we will develop a recursive formula for $\mathbb{P}_x^\pi\left(\tilde{Y}_n = y^j \mid \sigma(\tilde{H}_n^\pi)\right)$. The main idea of the proof for this recursive formula is roughly (and informally) speaking:

- Use Bayes' formula to write $\mathbb{P}_x^\pi\left(\tilde{Y}_n = y^j \mid \sigma(\tilde{H}_n)\right)$ as

$$\mathbb{P}_x^\pi\left(\tilde{Y}_n = y^j \mid \sigma(\tilde{H}_n)\right) = \frac{\mathbb{P}_x^\pi\left(\tilde{S}_n \in ds, \tilde{X}_n \in dx, \tilde{Y}_n = y^j \mid \sigma(\tilde{H}_{n-1})\right)}{\mathbb{P}_x^\pi\left(\tilde{S}_n \in ds, \tilde{X}_n \in dx \mid \sigma(\tilde{H}_{n-1})\right)},$$

- then, in step 1, write $\mathbb{P}_x^\pi\left(\tilde{S}_n \in ds, \tilde{X}_n \in dx, \tilde{Y}_n = y^j \mid \sigma(\tilde{H}_{n-1})\right)$ as a function of f_ϵ and $\mathbb{P}_x^\pi\left(\tilde{S}_n \in ds, \tilde{Y}_n = y^j \mid \sigma(\tilde{H}_{n-1})\right)$ and
- in step 2, write $\mathbb{P}_x^\pi\left(\tilde{S}_n \in ds, \tilde{X}_n \in dx \mid \sigma(\tilde{H}_{n-1})\right)$ as a function of f_ϵ and $\mathbb{P}_x^\pi\left(\tilde{S}_n \in ds, \tilde{Y}_n = y^j \mid \sigma(\tilde{H}_{n-1})\right)$ and,
- finally, in step 3, get a recursive formulation of $\mathbb{P}_x^\pi\left(\tilde{S}_n \in ds, \tilde{Y}_n = y^j \mid \sigma(\tilde{H}_{n-1})\right)$ making intervene $\mathbb{P}_x^\pi\left(\tilde{Y}_{n-1} = \cdot \mid \sigma(\tilde{H}_{n-1})\right)$.

We formalize this by the following step-by-step results. The outlined approach has been inspired by the work of Brandejski et al. in [13]. As they have been working on uncontrolled PO-PDMPs, we adapted their approach to our concrete situation of a controlled PO-PDMP, or more precisely, a controlled pseudo-embedded process. To simplify notations, we introduce first:

Definition 2.24 (Joint conditional distribution of $(\tilde{S}_n, \tilde{Y}_n)$). *For $n \geq 1$, $\pi \in \Pi^D$ and $h_{n-1} \in \mathcal{H}_{n-1}$ we define the joint conditional distribution of $(\tilde{S}_n, \tilde{Y}_n)$ under π given the observed history h_{n-1} by*

$$\gamma_n^\pi(h_{n-1}, \{y^j\}, ds) := \mathbb{P}_x^\pi\left(\tilde{S}_n \in ds, \tilde{Y}_n = y^j \mid \tilde{H}_{n-1} = h_{n-1}\right). \tag{2.8}$$

Under the use of γ_n^π we can explicitly describe how the density of the noise ϵ_n intervenes in the conditional distribution of $(\tilde{S}_n, \tilde{X}_n, \tilde{Y}_n)$ given an observed history h_{n-1}:

Lemma 2.25 (Step 1 - Intervention of noise density). *Let $n \geq 1$. For $\pi \in \Pi^D$ and $h_{n-1} \in \mathcal{H}_{n-1}$ we have the following equality of probability measures on $\mathbb{R}^+ \times E_X \times E_Y^0$, for all $j \in \{1,\dots,q\}$:*

$$\begin{aligned}\mathbb{P}_x^\pi(\tilde{S}_n \in ds, \tilde{X}_n \in dx, \tilde{Y}_n = y^j \mid \tilde{H}_{n-1} = h_{n-1}) \\ = \gamma_n^\pi(h_{n-1}, \{y^j\}, ds)\ f_\epsilon(x - \psi(y^j))\ \nu(dx).\end{aligned} \tag{2.9}$$

Proof. Let h a bounded and measurable real-valued function on $\mathbb{R}^+ \times E_X \times E_Y^0$. We then get

$$\begin{aligned}
&\mathbb{E}_x^\pi \left[h(\tilde{S}_n, \tilde{X}_n, \tilde{Y}_n) \mid \tilde{H}_{n-1} = h_{n-1} \right] \\
&= \int h\left(s, \psi(y) + \epsilon, y\right) \; \mathbb{P}_x^\pi \left(\tilde{S}_n \in ds, \epsilon_n \in d\epsilon, \tilde{Y}_n \in dy \mid \tilde{H}_{n-1} = h_{n-1} \right) \\
&= \sum_{j=1}^{q} \int \int h\left(s, \psi(y^j) + \epsilon, y^j\right) \; f_\epsilon(\epsilon) \; \nu(d\epsilon) \; \mathbb{P}_x^\pi \left(\tilde{S}_n \in ds, \tilde{Y}_n = y^j \mid \tilde{H}_{n-1} = h_{n-1} \right) \\
&= \sum_{j=1}^{q} \int \int h\left(s, x, y^j\right) \; f_\epsilon(x - \psi(y^j)) \; \nu(dx) \; \gamma_n^\pi(h_{n-1}, \{y^j\}, ds)
\end{aligned}$$

where we use the definition of $\tilde{X}_n$ in the first equality, the finite cardinality of E_Y^0 as well as the independence of ϵ_n from $(\tilde{S}_n, \tilde{Y}_n)$ and its density in equality 2 and finally the change of variable $x = \psi(y^j) + \epsilon$ together with the definition of γ_n^π in the last equality. □

Summation with respect to y^j in the previous lemma leads immediately to

Corollary 2.26 (Step 2 - Joint conditional distribution of $(\tilde{S}_n, \tilde{X}_n)$). *Let $n \geq 1$. For $\pi \in \Pi^D$ and $h_{n-1} \in \mathcal{H}_{n-1}$ we have the following equality of probability measures on $\mathbb{R}^+ \times E_X$:*

$$\mathbb{P}_x^\pi \left(\tilde{S}_n \in ds, \tilde{X}_n \in dx \mid \tilde{H}_{n-1} = h_{n-1} \right) = \left[\sum_{j=1}^{q} \gamma_n^\pi(h_{n-1}, \{y^j\}, ds) \; f_\epsilon \left(x - \psi(y^j) \right) \right] \nu(dx). \tag{2.10}$$

In order to simplify notations, we introduce the following notation (compare Lemma 1.35 where notation Λ^π introduced):

Definition 2.27. *For $y \in E_Y$ and $s \in \mathbb{R}^+$ we define*

$$\Lambda^r(y, s) := \int_0^s \int_A \lambda^A \left(\Phi^r(y, \tau), a \right) r_\tau(da) \; d\tau. \tag{2.11}$$

In the next step, we express γ_n^π as a function of μ_{n-1}^π:

Lemma 2.28 (Step 3 - γ_n^π as a function of μ_{n-1}^π). *Let $\pi \in \Pi^D$, then for all $n \geq 1, h_{n-1} \in \mathcal{H}_{n-1}$ and $j \in \{1, \ldots, q\}$ we have*

$$\begin{aligned}
&\gamma_n^\pi(h_{n-1}, \{y^j\}, ds) \\
&\quad = \sum_{i=1}^{q} \mu_{n-1}^{\pi,i}(h_{n-1}) \exp\left(-\Lambda^{\pi_{n-1}(h_{n-1})}(y^i, s) \right) \int_A \lambda^A \left(\Phi^{\pi_{n-1}(h_{n-1})}(y^i, s), a \right) \\
&\qquad\qquad Q^A \left(\Phi^{\pi_{n-1}(h_{n-1})}(y^i, s), a; \{y^j\} \right) \; \pi_{n-1}(h_{n-1})(s)(da) \; ds.
\end{aligned} \tag{2.12}$$

Proof. We define $\mathcal{F}_{T_n}$ as the σ-algebra modelling the full information available (observable and non-observable) up to time T_n by

$$\mathcal{F}_{T_n} := \sigma(\tilde{S}_0, \tilde{X}_0, \tilde{Y}_0, \tilde{R}_0, \ldots, \tilde{S}_n, \tilde{X}_n, \tilde{Y}_n),$$

where $\tilde{R}_k = \pi_k(\tilde{H}_k)$ for $k = 0, 1, \ldots, n-1$. With this notation, and for a bounded and measurable real-valued function h on $\mathbb{R}^+ \times E_Y^0$, we then can write:

$$\mathbb{E}_x^\pi \left[h(\tilde{S}_n, \tilde{Y}_n) \mid \tilde{H}_{n-1} = h_{n-1} \right] = \mathbb{E}_x^\pi \left[\mathbb{E}_x^\pi \left[h(\tilde{S}_n, \tilde{Y}_n) \mid \mathcal{F}_{T_{n-1}} \right] \mid \tilde{H}_{n-1} = h_{n-1} \right].$$

Based on the transition laws for the pseudo-embedded π-controlled process as outlined in Definition 2.9, the inner conditional expectation above can be written as

$$\begin{aligned}
&\mathbb{E}_x^\pi \left[h(\tilde{S}_n, \tilde{Y}_n) \mid \mathcal{F}_{T_{n-1}} \right] \\
&= \int h(s,y)\ \tilde{Q}_{SXY}^{\pi,n-1}(ds \otimes dy \otimes E_X \mid \tilde{S}_0, \tilde{X}_0, \tilde{Y}_0, \ldots \tilde{S}_{n-1}, \tilde{X}_{n-1}, \tilde{Y}_{n-1}) \\
&= \int h(s,y)\ \tilde{Q}_{SXY}(ds \otimes dy \otimes E_X \mid \tilde{S}_{n-1}, \tilde{X}_{n-1}, \tilde{Y}_{n-1}, \pi_{n-1}(\tilde{H}_{n-1})).
\end{aligned}$$

Taking now on this expression the conditional expectation given $\tilde{H}_{n-1} = h_{n-1}$, where h_{n-1} has the form $h_{n-1} = (s_0, x_0, \pi_0(s_0, x_0), \ldots, s_{n-1}, x_{n-1})$, we find

$$\begin{aligned}
&\mathbb{E}_x^\pi \left[h(\tilde{S}_n, \tilde{Y}_n) \mid \tilde{H}_{n-1} = h_{n-1} \right] \\
=&\mathbb{E}_x^\pi \left[\int h(s,y)\ \tilde{Q}_{SXY} \left(ds \otimes dy \otimes E_X \mid \tilde{S}_{n-1}, \tilde{X}_{n-1}, \tilde{Y}_{n-1}, \pi_{n-1}(\tilde{H}_{n-1}) \right) \right. \\
&\qquad \left. \mid \tilde{H}_{n-1} = h_{n-1} \right] \\
=&\int \int h(s,y)\ \tilde{Q}_{SXY}\left(ds \otimes dy \otimes E_X \mid s_{n-1}, x_{n-1}, y_{n-1}, \pi_{n-1}(h_{n-1})\right) \\
&\qquad \mu_{n-1}(h_{n-1})(dy_{n-1}) \\
=&\sum_{i=1}^{q} \mu_{n-1}(h_{n-1})(\{y^i\}) \int h(s,y) \\
&\qquad \tilde{Q}_{SXY}\left(ds \otimes dy \otimes E_X \mid s_{n-1}, x_{n-1}, y^i, \pi_{n-1}(h_{n-1}) \right),
\end{aligned}$$

where the last equation holds because $E_Y^0 = \{y^1, \ldots, y^q\}$ is of finite cardinality. Detailing now $\tilde{q}_{SXY}$ according to Definition 2.6, we obtain

$$\begin{aligned}
&\int h(s,y)\ \tilde{Q}_{SXY}(ds \otimes dy \otimes E_X \mid s_{n-1}, x_{n-1}, y^i, \pi_{n-1}(h_{n-1})) \\
&\qquad = \int_0^\infty \sum_{j=1}^{q} h(s, y^j) \exp\left(-\Lambda^{\pi_{n-1}(h_{n-1})}(y^i, s) \right) \int_{E_X} f_\epsilon(x - \psi(y^j)) \nu(dx) \\
&\int_A \lambda^A \left(\Phi^{\pi_{n-1}(h_{n-1})}(y^i, s), a \right) Q^A \left(\Phi^{\pi_{n-1}(h_{n-1})}(y^i, s), a; \{y^j\} \right)\ \pi_{n-1}(h_{n-1})(s)(da)\ ds.
\end{aligned}$$

Now the result follows as $\int_{E_X} f_\epsilon(x - \psi(y^j)) \nu(dx) = 1$. □

Looking closer at the statement of the previous lemma, the following dependence of γ on π is clear.

Corollary 2.29 (γ's dependence on the control policy). *Let $r = (r_0, r_1, \ldots) \in \mathcal{R}^\infty$. Then we have*

$$\mathbb{P}^r(\tilde{S}_n \in ds, \tilde{Y}_n = y^j \mid \tilde{H}_{n-1} = h_{n-1}) = \gamma_n(h_{n-1}, r_{n-1}, \{y^j\}, ds) \tag{2.13}$$

where $\gamma_n(h_{n-1}, r_{n-1}, \{y^j\}, ds)$ is defined by replacing $\pi_{n-1}(h_{n-1})$ by r_{n-1} in the term of the right hand side of equation (2.12) and $\mu_{n-1}^{r,i}(h_{n-1})$ is not explicitly dependent on r, thus can be written as $\mu_{n-1}(h_{n-1})$ in the term of the right hand side of equation (2.12).

Proof. The representation of γ_n^π as a function of μ_{n-1}^π as stated in equation (2.12) only depends on π via $\pi_{n-1}(h_{n-1})$, i.e. on the executed randomized control policy on the time interval $[T_{n-1}, T_n)$. Actually, $\mu_{n-1}^\pi(h_{n-1})$ does not depend on π_k for $k \geq n-1$ and the relevant information of π_k for $k \leq n-2$ is the executed randomized policy $\pi_k(h_k)$ which can be deducted from the observation $h_{n-1} = (s_0, x_0, \pi_0(h_0), \ldots, s_{n-1}, x_{n-1})$.

The proof of the previous lemma also holding for not history dependent, randomized piecewise open loop policies, we conclude the statement of the corollary. □

We can now prove the main result of this section, the recursive formulation of the filter sequence $(M_n^\pi)_{n\in\mathbb{N}_0}$.

Definition 2.30 (Filter equation). *Let the filter function defined as* $\chi = (\chi^1, \ldots, \chi^q) : \mathbf{P}(E_Y^0) \times \mathbb{R}^+ \times E_X \times \mathcal{R} \longrightarrow \mathbf{P}(E_Y^0)$ *where*

$$\chi^j(\mu, s, x, r) := \frac{1}{\overline{\chi}} \cdot \sum_{i=1}^q \chi_i^j(\mu, s, x, r) \qquad \text{for } j = 1, \ldots, q \text{ and}$$

$$\chi_i^j(\mu, s, x, r) := \mu^i \exp\left(-\Lambda^r(y^i, s)\right) \int_A \lambda^A\left(\Phi^r(y^i, s), a\right) \cdots \cdots Q^A\left(\Phi^r(y^i, s), a; \{y^j\}\right) r_s(da) f_\epsilon(x - \psi(y^j))$$

$$\text{for } i, j = 1, \ldots, q \text{ and}$$

$$\overline{\chi} := \sum_{j=1}^q \sum_{i=1}^q \chi_i^j \qquad \text{for normalization.}$$

Proposition 2.31 (Recursive formulation of filter). *Let* $\pi \in \Pi^D$*, then the filter* M_n^π *satisfies* $M_0^{\pi,j} = \delta_0(\tilde{S}_0)\ Q_0(\tilde{X}_0, \{y^j\}), j = 1, \ldots, q$ *and for* $n \geq 1$*:*

$$M_n^\pi = \chi(M_{n-1}^\pi, \tilde{S}_n, \tilde{X}_n, \pi_{n-1}(\tilde{H}_{n-1})) \qquad \mathbb{P}_x^\pi\text{−a.s..} \tag{2.14}$$

Proof. For $n = 0$ see Remark 2.22. For $n \geq 1$, let $\pi \in \Pi^D, h_{n-1} \in \mathcal{H}_{n-1}$ and $j \in \{1, \ldots, q\}$. First, Bayes' formula yields

$$\begin{aligned}\mathbb{P}_x^\pi(\tilde{S}_n \in ds, \tilde{X}_n \in dx, \tilde{Y}_n = y^j \mid \tilde{H}_{n-1}^\pi = h_{n-1}) \\ = \mathbb{P}_x^\pi\left(\tilde{Y}_n = y^j \mid \tilde{H}_n^\pi = (h_{n-1}, \pi_{n-1}(h_{n-1}), s, x)\right) \\ \times \mathbb{P}_x^\pi(\tilde{S}_n \in ds, \tilde{X}_n \in dx \mid \tilde{H}_{n-1}^\pi = h_{n-1}).\end{aligned} \tag{2.15}$$

Applying Lemma 2.25 and corollary 2.26 this becomes:

$$\begin{aligned}\gamma_n^\pi(h_{n-1}, \{y^j\}, ds) f_\epsilon(x - \psi(y^j)) \nu(dx) \\ = \mathbb{P}_x^\pi\left(\tilde{Y}_n = y^j \mid \tilde{H}_n^\pi = (h_{n-1}, \pi_{n-1}(h_{n-1}), s, x)\right) \\ \times \left[\sum_{k=1}^q \gamma_n^\pi(h_{n-1}, \{y^k\}, ds) f_\epsilon(x - \psi(y^k))\right] \nu(dx).\end{aligned} \tag{2.16}$$

With respect to x, this shows the equality of two absolute continuous measures and we can deduct the equality a.e. of the densities. Thus, for almost all $x \in E_X$ w.r.t. ν,

$$\begin{aligned}\gamma_n^\pi(h_{n-1}, \{y^j\}, ds) f_\epsilon(x - \psi(y^j)) \\ = \mathbb{P}_x^\pi\left(\tilde{Y}_n = y^j \mid \tilde{H}_n^\pi = (h_{n-1}, \pi_{n-1}(h_{n-1}), s, x)\right) \\ \times \left[\sum_{k=1}^q \gamma_n^\pi(h_{n-1}, \{y^k\}, ds) f_\epsilon(x - \psi(y^k))\right].\end{aligned} \tag{2.17}$$

This equality being an equality of two measures of the variable s, we have for all bounded, measurable and real-valued functions F on $\mathbb{R}^+$

$$\int_{\mathbb{R}^+} F(s) f_\epsilon(x-\psi(y^j))\gamma_n^\pi(h_{n-1},\{y^j\},ds) \\ = \int_{\mathbb{R}^+} F(s)\ \mathbb{P}_x^\pi\left(\tilde{Y}_n = y^j \mid \tilde{H}_n^\pi = (h_{n-1},\pi_{n-1}(h_{n-1}),s,x)\right) \\ \times\left[\sum_{k=1}^q f_\epsilon(x-\psi(y^k))\gamma_n^\pi(h_{n-1},\{y^k\},ds)\right]. \quad (2.18)$$

In Lemma 2.28 we showed that $\gamma_n^\pi(h_{n-1},\{y^j\},ds)$ has a density f_γ with respect to the variable s and in corollary 2.29 we argued that this density only depends on π via $\pi_{n-1}(h_{n-1})$ and we can write it

$$f_\gamma(h_{n-1},\pi_{n-1}(h_{n-1}),\{y^j\},s) := \sum_{i=1}^q \mu_{n-1}^i(h_{n-1})\exp(-\Lambda^{\pi_{n-1}(h_{n-1})}(y^i,s)) \\ \int_A \lambda^A(\Phi^{\pi_{n-1}(h_{n-1})}(y^i,s),a)\ Q^A(\Phi^{\pi_{n-1}(h_{n-1})}(y^i,s),a;\{y^j\})\ \pi_{n-1}(h_{n-1})(s)(da) \quad (2.19)$$

Using this density, equation (2.18) implies that almost surely w.r.t. the Lebesgue measure on $\mathbb{R}^+$, we have

$$\mathbb{P}_x^\pi\left(\tilde{Y}_n = y^j \mid \tilde{H}_{n-1}^\pi = (h_{n-1},\pi_{n-1}(h_{n-1}),s,x)\right) \\ = \frac{f_\gamma(h_{n-1},\pi_{n-1}(h_{n-1}),\{y^j\},s)f_\epsilon(x-\psi(y^j))}{\sum_{k=1}^q f_\gamma(h_{n-1},\pi_{n-1}(h_{n-1}),\{y^k\},s)f_\epsilon(x-\psi(y^k))}. \quad (2.20)$$

From the equalities a.e. (2.17) and (2.20) we conclude that there exists a measurable set $\mathcal{N}_x \subset E_X$ with $\nu(\mathcal{N}_x) = 0$ and a borel subset $\mathcal{N}_s \subset \mathbb{R}^+$ negligible w.r.t. the Lebesgue measure on $\mathbb{R}^+$, such that for all $x \in E_X \setminus \mathcal{N}_x$ and all $s \in \mathbb{R}^+ \setminus \mathcal{N}_s$

$$\mu_n^\pi(h_{n-1},\pi_{n-1}(h_{n-1}),s,x) = \chi(\mu_{n-1}^\pi(h_{n-1}),s,x,\pi_{n-1}(h_{n-1})). \quad (2.21)$$

Further, $\mathbb{P}_x^\pi(\tilde{X}_n \in \mathcal{N}_x) \le \sum_{j=1}^q \mathbb{P}_x^\pi(\psi(y^j)+\epsilon_n \in \mathcal{N}_x) = 0$ because ϵ_n is absolute continuous w.r.t. the measure ν on E_X and $\mathbb{P}_x^\pi(\tilde{S}_n \in \mathcal{N}_s) = 0$ because the distribution of $\tilde{S}_n$ is absolute continuous on $\mathbb{R}^+$. Thus, we conclude

$$\mu_n^\pi(\tilde{H}_{n-1}^\pi,\pi_{n-1}(\tilde{H}_{n-1}^\pi),\tilde{S}_n,\tilde{X}_n) = \chi(\mu_{n-1}^\pi(\tilde{H}_{n-1}^\pi),\tilde{S}_n,\tilde{X}_n,\pi_{n-1}(\tilde{H}_{n-1}^\pi)) \qquad \mathbb{P}_x^\pi\text{−a.s.} \quad (2.22)$$

and the statement of the proposition follows as

$$M_n^\pi = \mu_n^\pi(\tilde{H}_{n-1}^\pi,\pi_{n-1}(\tilde{H}_{n-1}^\pi),\tilde{S}_n,\tilde{X}_n) \qquad \mathbb{P}_x^\pi\text{−a.s.} \quad (2.23)$$

by the factorization of conditional expected values (see Remark 2.23). □

2.2.3 The derived filtered process and the corresponding optimization problem

The main result of this section will be the formulation of a completely observable MDP together with a corresponding optimization problem. This problem shall then be the candidate for the equivalent reformulation of the initial optimization problem for the PO-PDMP into a time-discrete problem under complete observation. We will construct this completely

observable MDP based on a pseudo-embedded π-controlled process $(\tilde{S}_k, \tilde{X}_k, \tilde{Y}_k)_{k\in\mathbb{N}_0}$ while replacing the unobservable component $\tilde{Y}_k$ by the previously developed filter M_k. Thus, our goal is, to define a so-called *derived filtered process* $(\tilde{S}_k, \tilde{X}_k, M_k)_{k\in\mathbb{N}_0}$ with adequate state spaces and transition laws. After explaining why we can allow Π^D as set of admissible control policies for the derived filtered process, we finally will define an optimization problem to solve for this process.

Definition of derived filtered model :
We start by defining the derived filtered model.

Definition 2.32 (Derived filtered model). *Let $(\tilde{S}_k, \tilde{X}_k, \tilde{Y}_k)_{k\geq 0}$ a pseudo-embedded process with transition law $\tilde{q}_{SXY}$ and initial conditional distribution Q_0. Let further χ the corresponding filter function. The* derived filtered $[r]$-controlled model *consists of a set of data (E', q'_{SXM}, g', G') with the following meaning:*

- *$E' := E_S \times E_X \times \mathbf{P}(E_Y^0)$ is the state space. An element is denoted by (s, x, ρ) where s and x should be understood as the observable part of the state of the pseudo-embedded process $(\tilde{S}_k, \tilde{X}_k, \tilde{Y}_k)_{k\geq 0}$ and ρ is the conditional distribution of the unobservable state. We endow the space E' with the product topology and the corresponding Borel-σ-algebra, where $E_S = \mathbb{R}^+$ is endowed with the standard topology, E_X is endowed with the topology coming from E_Y via ψ and $\mathbf{P}(E_Y^0)$ is endowed with the weak topology.*

- *Q'_{SXM} is a stochastic kernel which determines the distribution of the new states as follows: For fixed $(s, x, \rho) \in E'$ and $[r] \in \mathcal{R}$ as well as for $t \in \mathbb{R}^+$ and Borel subsets $B_X \subset E_X, C \subset \mathbf{P}(E_Y^0)$ we define*

$$Q'_{SXM}([0,t] \times B_X \times C \mid s, x, \rho, [r]) := \sum_{j=1}^{q} \int_{[0;t]\times B_X} \mathbb{1}_C(\chi(\rho, s', x', r))\ \tilde{Q}_{SXY}(ds' \otimes dx' \otimes E_Y^0 \mid s, x, y^j, [r])\ \rho(\{y^j\})$$

- *g' is the undiscounted one step cost function defined by*

$$g'(x, \rho, r) := \sum_{j=1}^{q} g(x, y^j, r)\rho(\{y^j\}) \tag{2.24}$$

- *G' is the discounted one step cost function defined by*

$$G'(t, x, \rho, r) := e^{-\beta t} g'(x, \rho, r). \tag{2.25}$$

An initial state of this filtered process typically has the format $(0, x, Q_0(x; \cdot))$, where $Q_0(x, \cdot)$ is the initial conditional distribution of $\tilde{Y}_0$ as given in the model of the pseudo-embedded process. We call the above process *derived* filtered model, because its transition law is *derived* from the transition law of the underlying pseudo-embedded process and the corresponding filter function.

We could look at the derived filtered process as a process where the states are realized by an *arbitrary* random experience with transition law Q'_{SXM}. What we will do in the sequel, however, is fixing a *concrete* random experience to realize the stages of the derived filtered process: we start in $(0, x, Q_0(x\cdot))$, select a relaxed control $r \in \mathcal{R}$ and run the underlying pseudo-embedded process with initial observation x and initial conditional distribution Q_0 under this relaxed control. We then note the observed states s_1 and x_1 of

the pseudo-embedded process and get the stage one states of the derived filtered process by keeping s_1 and x_1 as first two components and calculating ρ_1 as $\rho_1 = \chi(Q_0(x,\cdot), s_1, x_1, r)$. Selecting a new relaxed control we iterate this procedure.

With other words: we *derive* the states of the *derived filtered process* from the pseudo-embedded process by running the pseudo-embedded process and, at the same time, calculating after each observation, the current state of the filter ρ_n. The (straight forward) proof that this procedure is actually creating a process with transition law q' is left to the reader.

History dependent control policies for the derived filtered model :
The derived filtered process generating states $(\tilde{S}_k)_{k\geq 0}$ and $(\tilde{X}_k)_{k\geq 0}$ it also generates an observable history $\tilde{H}_k = (\tilde{S}_0, \tilde{X}_0, \tilde{R}_0, \ldots, \tilde{S}_k, \tilde{X}_k)$ for each $k \in \mathbb{N}_0$ as long as the process is controlled, at each stage k, by some relaxed control $\tilde{R}_k \in \mathcal{R}$. Having this in mind, we simply can allow history dependent time-discrete relaxed control policies $\pi^D \in \Pi^D$ for the derived filtered process: at each stage k, we select the relaxed control $\tilde{R}_k$ to be executed by the decision rule $\tilde{R}_k = \pi^D_k(\tilde{H}_k)$ as we also did it for the pseudo-embedded π^D-controlld processes.

Analogously to the pseudo-embedded process, we get a transition law $Q'^{\pi^D,n}_{SXM}$ for the π^D-controlled derived filtered process. This transition law is given by

$$\begin{aligned} Q'^{\pi^D,n}_{SXM}([0,t]\times B_X \times C \mid s_0, x_0, \ldots, s_n, x_n, \rho_n) := \\ \sum_{j=1}^{q} \int_{[0;t]\times B_X} \mathbb{1}_C(\chi(\rho_n, s', x', \pi^D_n(h_n))) \\ \tilde{Q}^{\pi^D,n}_{SXY}(ds' \otimes dx' \otimes E^0_Y \mid s_0, x_0, \ldots, s_n, x_n, y^j)\ \rho_n(\{y^j\}). \end{aligned}$$

Analogously to the case of the pseudo-embedded process, this transition law is explicitly depending on π^D, on the history of $\tilde{S}_0, \tilde{X}_0, \ldots, \tilde{S}_n, \tilde{X}_n$ as well as on the component ρ_n of the current state. It is not explicitly depending on the history $\rho_0, \ldots, \rho_{n-1}$ of the conditional distributions. Implicitly, however, these are coded in the current state ρ_n, which basically is calculated by iteration of χ, and thus, contains the information about the history of the conditional distributions.

An initial observation of a derived filtered process has always the form $(0, x, Q_0(x,\cdot))$ with $x \in E_X$ and Q_0 the initial transition kernel from E_X to $\mathbf{P}(E^0_Y)$ of the underlying pseudo-embedded process. Thus, knowing x and Q_0, one knows the full initial observation. For a given policy $\pi^D \in \Pi^D$, the theorem of Ionescu-Tulcea implies that the transition laws $Q'^{\pi^D,n}_{SXM}$ together with x and Q_0 of an initial observation determine a probablity distribution $\mathbb{P}^{\pi^D}_{xQ_0}$ on the space $(\mathbb{R}^+ \times E_X \times \mathbf{P}(E^0_Y))^\infty$. By restriction to $(\mathbb{R}^+ \times E_X)^\infty$, this probability distribution induces a probability distribution on $(\mathbb{R}^+ \times E_X)^\infty$. As long as the context is clear, we will again denote this probability distribution with $\mathbb{P}^{\pi^D}_{xQ_0}$ by slight abuse of notation.

Whenever we look at a pseudo-embedded process together with its derived filtered process, the next result is as trivial as important for the sequel.

Lemma 2.33. *Let $(\tilde{S}_n, \tilde{X}_n, \tilde{Y}_n)_{n\geq 0}$ a pseudo-embedded process with initial transition kernel Q_0 and transition law $\tilde{Q}_{SXY}$. Then, for $\pi^D \in \Pi^D$ and an initial observation $x \in E_X$, the pseudo-embedded process as well as its derived filtered process induce probability measures on $(\mathbb{R}^+ \times E_X)^\infty$ denoted by $\mathbb{P}^{\pi^D}_x$ and $\mathbb{P}^{\pi^D}_{xQ_0}$ respectively and it holds*

$$\mathbb{P}^{\pi^D}_x = \mathbb{P}^{\pi^D}_{xQ_0}$$

on $\mathcal{B}(\mathbb{R}^+) \otimes \mathcal{B}(E_X)$.

Proof. The existence of $\mathbb{P}_x^{\pi^D}$ and $\mathbb{P}_{xQ_0}^{\pi^D}$ have been discussed when introducing the pseudo-embedded process and its derived filtered process. Both processes induce an initial distribution of $\delta_0 \otimes \delta_x$ under the initial observation $x \in E_X$.

Let now $n \geq 0$ and $(s_0, x_0, \ldots, s_n, x_n) \in (\mathbb{R}^+ \times E_X)^{n+1}$. The pseudo-embedded process induces now the following transition law:

$$\begin{aligned}\mathbb{P}_x^{\pi^D}(\tilde{S}_{n+1} \in ds, \tilde{X}_{n+1} \in dx \mid \tilde{S}_0 = s_0, \tilde{X}_0 = x_0, \ldots, \tilde{S}_n = s_n, \tilde{X}_n = x_n) \\ = \sum_{j=1}^{q} \tilde{Q}_{SXY}^{\pi^D,n}(ds \otimes dx \otimes E_Y^0 \mid s_0, x_0, \ldots, s_n, x_n, y^j) \\ \mathbb{P}_x^{\pi^D}(\tilde{Y}_n = y^j \mid H_n = h_n),\end{aligned}$$

where h_n denotes the observed history under π^D and $(s_0, x_0, \ldots, s_n, x_n)$ and we clearly make use of the slight abuse of notation when writing $\mathbb{P}_x^{\pi^D}$ for distributions on $(\mathbb{R}^+ \times E_X)^\infty$ and on $(\mathbb{R}^+ \times E_X \times E_Y)^\infty$.

Making use of the factorization of the filter into $M_n = \mu_n \circ H_n$, the above now can be written as

$$\begin{aligned}&= \sum_{j=1}^{q} \tilde{Q}_{SXY}^{\pi^D,n}(ds \otimes dx \otimes E_Y^0 \mid s_0, x_0, \ldots, s_n, x_n, y^j)\ \mu_n(h_n)(\{y^j\}) \\ &= Q'_{SXM}(ds \otimes dx \otimes \mathbf{P}(E_Y^0) \mid s_n, x_n, \mu_n(h_n), \pi_n^D(h_n)) \\ &= Q'^{\pi^D,n}_{SXM}(ds \otimes dx \otimes \mathbf{P}(E_Y^0) \mid s_0, x_0, \ldots, s_n, x_n, \mu_n(h_n)) \\ &= \mathbb{P}_{xQ_0}^{\pi^D}(\tilde{S}_{n+1} \in ds, \tilde{X}_{n+1} \in dx \mid \tilde{S}_0 = s_0, \tilde{X}_0 = x_0, \ldots, \tilde{S}_n = s_n, \tilde{X}_n = x_n). \qquad \square\end{aligned}$$

Optimal control problem for derived filtered process :
We can now state the optimal control problem for a π^D-controlled derived filtered process.

Definition 2.34. *Let $(\tilde{S}_n, \tilde{X}_n, \tilde{Y}_n)_{n\geq 0}$ a pseudo-embedded process with initial conditional distribution Q_0 and $(\tilde{S}_n, \tilde{X}_n, M_n)_{n\geq 0}$ its derived filtered process. For an initial observation $(0, x, Q_0(x, \cdot))$ of the derived filtered process, we define the cost of a policy $\pi^D \in \Pi^D$ by*

$$J'\left(x, Q_0(x, \cdot), \pi^D\right) := \mathbb{E}_{xQ_0}^{\pi^D}\left[\sum_{k=0}^{\infty} e^{-\beta \sum_{n=0}^{k} \tilde{S}_n} g'\left(\tilde{X}_k, M_k, \pi_k^D(H_k)\right)\right],$$

where g' is as of Definition 2.32.

The value function *of the derived filtered control model gives the minimal cost under an initial observation $(0, x, Q_0(x, \cdot))$ and is defined for all $x \in E_X, Q_0(x, \cdot) \in \mathbf{P}(E_Y^0)$ as*

$$J'\left(x, Q_0(x, \cdot)\right) := \inf_{\pi^D \in \Pi^D} J'\left(x, Q_0(x, \cdot), \pi^D\right). \tag{2.26}$$

The filtered optimization problem *is then to find, for $x \in E_X, Q_0(x, \cdot) \in \mathbf{P}(E_Y^0)$, a policy $\pi^{\star D} \in \Pi^D$ such that we get*

$$J'\left(x, Q_0(x, \cdot)\right) = J'\left(x, Q_0(x, \cdot), \pi^{\star D}\right).$$

2.2.4 Equivalent time-discrete optimization problem under complete observation

Having introduced the optimal control problem for a derived filtered model, we will now prove that the latter optimization problem is equivalent to the optimal control problem for the underlying pseudo-embedded process. With other words, we will show, that, every policy $\pi^{\star D}$ that is optimal for a pseudo-embedded process is optimal for its derived filtered process and vice-versa.

We will further show, that, for derived filtered processes, an optimal control policy can already be found in the class of Markov policies. The latter one can be understood as a subset of Π^D in a sense that we will specify. This is a classical result for MDPs under complete observation.

As a consequence, solving the optimal control problem for the derived filtered process in the class of its Markov policies provides a history dependent relaxed control policy $\pi^{\star D}$ that is optimal for the underlying pseudo-embedded process.

Equivalence of optimization porblems for π^D-controlled processes :
We start with a lemma that will be crucial in the proof for the equivalence of the optimal control problems for pseudo-embedded and corresponding derived processes.

Lemma 2.35. *Let $v : \mathcal{H}_{n-1} \times \mathcal{R} \times \mathbb{R}^+ \times E_X \times E_Y \to \mathbb{R}$ a measurable mapping. Then, in the setting of a derived filtered process, for all $h_{n-1} \in \mathcal{H}_{n-1}$ and all $\pi^D \in \Pi^D$ it holds:*

$$\begin{aligned}
&\sum_{i=1}^{q} \mu_{n-1}(h_{n-1})(\{y^i\}) \int \tilde{Q}_{SXY}^{\pi^D,n-1}\left(ds_n \otimes dx_n \otimes dy_n \mid s_0, x_0, \dots s_{n-1}, x_{n-1}, y^i\right) \\
&\qquad\qquad v\left(h_{n-1}, \pi_{n-1}^D(h_{n-1}), s_n, x_n, y_n\right) \\
&= \sum_{i=1}^{q} \mu_{n-1}(h_{n-1})(\{y^i\}) \int \tilde{Q}_{SXY}^{\pi^D,n-1}\left(ds_n \otimes dx_n \otimes E_Y^0 \mid s_0, x_0, \dots s_{n-1}, x_{n-1}, y^i\right) \\
&\sum_{j=1}^{q} \chi\left(\mu_{n-1}(h_{n-1}), s_n, x_n, \pi_{n-1}^D(h_{n-1})\right)(\{y^j\})\ v\left(h_{n-1}, \pi_{n-1}^D(h_{n-1}), s_n, x_n, y^j\right) \qquad (2.27)
\end{aligned}$$

Proof. To simplify notations, we will write $\mu_n^j(h_n)$ for $\mu_n(h_n)(\{y^j\})$ and $\chi^k(\cdot)$ for $\chi(\cdot)(\{y^k\})$. Based on the transition law $\tilde{q}_{SXY}^{\pi^D,n-1}$, the left hand side of (2.27) transforms into

$$\begin{aligned}
&\sum_{i=1}^{q} \mu_{n-1}^i(h_{n-1}) \int_{\mathbb{R}^+} \exp\left(-\Lambda^{\pi_{n-1}^D(h_{n-1})}(y^i, s_n)\right) \sum_{j=1}^{q} \int_{E_X} f_\epsilon(x_n - \psi(y^j)) \\
&\qquad v(h_{n-1}, \pi_{n-1}^D(h_{n-1}), s_n, x_n, y^j) \int_A \lambda^A\left(\Phi^{\pi_{n-1}^D(h_{n-1})}(y^i, s_n), a\right) \\
&\qquad Q^A\left(\Phi^{\pi_{n-1}^D(h_{n-1})}(y^i, s_n), a; \{y^j\}\right)\ \pi_{n-1}^D(h_{n-1})(s_n)(da)\ \nu(dx_n)\ ds_n. \qquad (2.28)
\end{aligned}$$

For the right hand side of equation (2.27), we first recognize the sum w.r.t. index i together with the integration w.r.t. ds_n, dx_n as an integration w.r.t. ds_n, dx_n given the observed history h_{n-1}. With other words:

$$\begin{aligned}
&\sum_{i=1}^{q} \mu_{n-1}^i(h_{n-1}) \int \mathbb{P}_x^\pi\left(\tilde{S}_n \in ds_n, \tilde{X}_n \in dx_n \mid s_{n-1}, x_{n-1}, y^i, \pi_{n-1}^D(h_{n-1})\right) \\
&\qquad\qquad = \int \mathbb{P}_x^\pi(\tilde{S}_n \in ds_n, \tilde{X}_n \in dx_n \mid \tilde{H}_{n-1} = h_{n-1})
\end{aligned}$$

This observation together with the result of corollary 2.26 lead to a reformulation of the right hand side of equation (2.27) of

$$\int_{E_X} \nu(dx_n) \int_{\mathbb{R}^+} \sum_{j=1}^{q} \left[\gamma_n^{\pi^D}(h_{n-1}, \{y^j\}, ds_n) f_\epsilon(x_n - \psi(y^j))\right] \sum_{k=1}^{q} \chi^k \left(\mu_{n-1}(h_{n-1}), s_n, x_n, \pi_{n-1}^D(h_{n-1})\right) \; v\left(h_{n-1}, \pi_{n-1}^D(h_{n-1}), s_n, x_n, y^k\right) \quad (2.29)$$

Now, applying the result of Lemma 2.28 and re-writing $\gamma_n^{\pi^D}$, we obtain

$$\int_{E_X} \int_{\mathbb{R}^+} \sum_{j=1}^{q} \sum_{i=1}^{q} \mu_{n-1}^i(h_{n-1}) \exp\left(-\Lambda^{\pi_{n-1}^D(h_{n-1})}(y^i, s_n)\right) f_\epsilon(x_n - \psi(y^j)) \sum_{k=1}^{q} \chi^k \left(\mu_{n-1}(h_{n-1}), s_n, x_n, \pi_{n-1}^D(h_{n-1})\right) \; v(h_{n-1}, \pi_{n-1}^D(h_{n-1}), s_n, x_n, y^k) \int_A \lambda^A \left(\Phi^{\pi_{n-1}^D(h_{n-1})}(y^i, s_n), a\right) \; Q^A \left(\Phi^{\pi_{n-1}^D(h_{n-1})}(y^i, s_n), a; \{y^j\}\right) \pi_{n-1}^D(h_{n-1})(s_n)(da) \; ds_n \; \nu(dx_n). \quad (2.30)$$

After inserting the definition of χ^k into equation (2.30), $\overline{\chi}$ cancels out and we obtain

$$\int_{\mathbb{R}^+} \int_{E_X} \sum_{k=1}^{q} \sum_{l=1}^{q} \mu_{n-1}^l(h_{n-1}) \exp\left(-\Lambda^{\pi_{n-1}^D(h_{n-1})}(y^l, s_n)\right) \; f_\epsilon(x_n - \psi(y^k)) \; v(h_{n-1}, \pi_{n-1}(h_{n-1}), s_n, x_n, y^k) \int_A \lambda^A \left(\Phi^{\pi_{n-1}^D(h_{n-1})}(y^l, s_n), a\right) \; Q^A \left(\Phi^{\pi_{n-1}^D(h_{n-1})}(y^l, s_n), a; \{y^k\}\right) \; \pi_{n-1}^D(h_{n-1})(s_n)(da) \; \nu(dx_n) \; ds_n, \quad (2.31)$$

which is equal to the left hand side of (2.27) as developped in equation (2.28). □

Based on this lemma we can now state and prove the main result of this section, the equivalence of the optimal control problem on $\tilde{J}$ to the optimal control problem on J'.

Proposition 2.36. *In the situation of Definition 2.34, let $\tilde{J}(x, \pi^D)$ denote the cost of a policy $\pi^D \in \Pi^D$ under initial observation $x \in E_X$ for the underlying pseudo-embedded process. We then have for all $x \in E_X$ and for all $\pi^D \in \Pi^D$*

$$\tilde{J}(x, \pi^D) = J'(x, Q_0(x, \cdot), \pi^D), \quad (2.32)$$

and for all $x \in E_X$ we have for the value functions

$$\tilde{J}(x) = J'(x, Q_0(x, \cdot)). \quad (2.33)$$

Proof. According to the definitions of $\tilde{J}$ and J', we have to show

$$\mathbb{E}_x^{\pi^D}\left[\sum_{k=0}^{\infty} e^{-\beta \sum_{n=0}^{k} \tilde{S}_n} g(\tilde{X}_k, \tilde{Y}_k, \pi_k^D(\tilde{H}_k))\right] = \mathbb{E}_{xQ_0}^{\pi^D}\left[\sum_{k=0}^{\infty} e^{-\beta \sum_{n=0}^{k} \tilde{S}_n} g'(\tilde{X}_k, M_k, \pi_k^D(H_k)\right]. \quad (2.34)$$

As the cost function c is non-negative, the functions g and g' are non-negative and we can swap integration and summation. So, we need to show

$$\sum_{k=0}^{\infty} \mathbb{E}_x^{\pi^D}\left[e^{-\beta\sum_{n=0}^{k}\tilde{S}_n} g(\tilde{X}_k, \tilde{Y}_k, \pi_k^D(\tilde{H}_k))\right] = \sum_{k=0}^{\infty} \mathbb{E}_{xQ_0}^{\pi^D}\left[e^{-\beta\sum_{n=0}^{k}\tilde{S}_n} g'(\tilde{X}_k, M_k, \pi_k^D(H_k)\right]. \quad (2.35)$$

This is done if we can show that, for $k \geq 0$, we have

$$\mathbb{E}_x^{\pi^D}\left[e^{-\beta\sum_{n=0}^{k}\tilde{S}_n} g(\tilde{X}_k, \tilde{Y}_k, \pi_k^D(\tilde{H}_k))\right] = \mathbb{E}_{xQ_0}^{\pi^D}\left[e^{-\beta\sum_{n=0}^{k}\tilde{S}_n} g'(\tilde{X}_k, M_k, \pi_k^D(H_k)\right]. \quad (2.36)$$

This result can be shown along the same lines as the proof of theorem 5.3.2 in [6] but we give an adapted version of the proof here. It follows actually by induction from Lemma 2.35 for v^{π^D} defined as:

$$v^{\pi^D}(H_{k-1}, \pi_{k-1}^D(H_{k-1}), \tilde{S}_k, \tilde{X}_k, \tilde{Y}_k) := e^{-\beta\sum_{n=0}^{k}\tilde{S}_n} g(\tilde{X}_k, \tilde{Y}_k, \pi_k^D(H_{k-1}, \pi_{k-1}^D(H_{k-1}), \tilde{S}_k, \tilde{X}_k)). \quad (2.37)$$

With this definition and for $k = 0$, we get for the left hand side of equation (2.36):

$$\begin{aligned}\mathbb{E}_x^{\pi^D}\left[e^{-\beta\tilde{S}_0} g(\tilde{X}_0, \tilde{Y}_0, \pi_0^D(\tilde{H}_0)\right] &= \mathbb{E}_x^{\pi^D}\left[1 \cdot g(x, \tilde{Y}_0, \pi_0^D((0,x)))\right] \\ &= \sum_{i=0}^{q} g(x, y^i, \pi_0^D((0,x))) \cdot Q_0(x, \{y^i\}).\end{aligned} \quad (2.38)$$

On the other hand, we find for the right hand side of equation (2.36):

$$\begin{aligned}\mathbb{E}_{xQ_0}^{\pi^D}\left[e^{-\beta\tilde{S}_0} g'(\tilde{X}_0, M_0, \pi_0^D(H_0))\right] &= \mathbb{E}_{xQ_0}^{\pi^D}\left[1 \cdot g'(x, Q_0(x\cdot), \pi_0^D((0,x)))\right] \\ &= \sum_{i=0}^{q} g(x, y^i, \pi_0^D((0,x))) \cdot Q_0(x, \{y^i\}).\end{aligned} \quad (2.39)$$

and equation (2.36) holds for $k = 0$.

Let now $k \geq 1$ and suppose, equation (2.36) holds for $0, \ldots, k-1$. Using the rule for iterated expectations, the left hand side transforms into:

$$\mathbb{E}_x^{\pi^D}\left[e^{-\beta\sum_{n=0}^{k}\tilde{S}_n} g(\tilde{X}_k, \tilde{Y}_k, \pi_k^D(\tilde{H}_k))\right] = \mathbb{E}_x^{\pi^D}\left[E_x^{\pi^D}\left[e^{-\beta\sum_{n=0}^{k}\tilde{S}_n} g(\tilde{X}_k, \tilde{Y}_k, \pi_k^D(\tilde{H}_k)) \mid \tilde{H}_{k-1}\right]\right]. \quad (2.40)$$

We will now prove in two steps that the latter expression transforms into

$$\mathbb{E}_{xQ_0}^{\pi^D}\left[\mathbb{E}_{xQ_0}^{\pi^D}\left[e^{-\beta\sum_{n=0}^{k}\tilde{S}_n} g'(\tilde{X}_k, M_k, \pi_k^D(\tilde{H}_k) \mid \tilde{H}_{k-1}\right]\right], \quad (2.41)$$

which is equal to the the right hand side of (2.36) and the proof will be finished.

Step 1 is now, to explain how to get from $\mathbb{E}_x^{\pi^D}$ to $\mathbb{E}_{xQ_0}^{\pi^D}$ for the outer expectation in the iterated conditional expectation of (2.40).

Under a fixed policy π^D, it is easy to see that $\sigma(\tilde{H}_{k-1}) = \sigma(\tilde{S}_0, \tilde{X}_0, \ldots, \tilde{S}_{k-1}, \tilde{X}_{k-1})$ and we get

$$\mathbb{E}_x^{\pi^D}\left[E_x^{\pi^D}\left[e^{-\beta\sum_{n=0}^{k}\tilde{S}_n} g(\tilde{X}_k, \tilde{Y}_k, \pi_k^D(\tilde{H}_k)) \mid \sigma(\tilde{H}_{k-1})\right]\right]$$
$$= \int_{(\mathbb{R}^+\times E_X)^k} \mathbb{P}_x^{\pi^D}\left(d(s_0, x_0, \ldots, s_{k-1}, x_{k-1})\right)$$
$$E_x^{\pi^D}\left[e^{-\beta\sum_{n=0}^{k}\tilde{S}_n} g(\tilde{X}_k, \tilde{Y}_k, \pi_k^D(\tilde{H}_k)) \mid \tilde{S}_0 = s_0, \tilde{X}_0 = x_0, \ldots \tilde{S}_{k-1} = s_{k-1}, \tilde{X}_{k-1} = x_{k-1}\right].$$

Applying Lemma 2.33, we can replace $\mathbb{P}_x^{\pi^D}$ by $\mathbb{P}_{xQ_0}^{\pi^D}$ in the above integral and writing the resulting expression using the $\mathbb{E}$-operator we find

$$= \mathbb{E}_{xQ_0}^{\pi^D}\left[E_x^{\pi^D}\left[e^{-\beta\sum_{n=0}^{k}\tilde{S}_n} g(\tilde{X}_k, \tilde{Y}_k, \pi_k^D(\tilde{H}_k)) \mid \sigma(\tilde{H}_{k-1})\right]\right].$$

For step 2, remains to show that, for all h_{k-1} observable under π^D, we have

$$E_x^{\pi^D}\left[e^{-\beta\sum_{n=0}^{k}\tilde{S}_n} g(\tilde{X}_k, \tilde{Y}_k, \pi_k^D(\tilde{H}_k)) \mid \tilde{H}_{k-1} = h_{k-1}\right]$$
$$= \mathbb{E}_{xQ_0}^{\pi^D}\left[e^{-\beta\sum_{n=0}^{k}\tilde{S}_n} g'(\tilde{X}_k, M_k, \pi_k^D(\tilde{H}_k) \mid \tilde{H}_{k-1} = h_{k-1}\right]. \quad (2.42)$$

Building the observed history h_{k-1} according to the earlier introduced mechanism as $h_{k-1} := (s_0, x_0, \pi_0^D(h_0), \ldots, s_{k-1}, x_{k-1})$, the conditional expectation from the left hand side can be written as

$$E_x^{\pi^D}\left[e^{-\beta\sum_{n=0}^{k}\tilde{S}_n} g(\tilde{X}_k, \tilde{Y}_k, \pi_k^D(\tilde{H}_k)) \mid \tilde{H}_{k-1} = h_{k-1}\right]$$
$$= \sum_{i=1}^{q} \mathbb{P}_x^{\pi^D}(\tilde{Y}_{k-1} = y^i \mid \tilde{H}_{k-1} = h_{k-1}) \int \tilde{Q}_{SXY}^{\pi^D,k-1}(ds_k \otimes dx_k \otimes dy^k \mid s_0, x_0, \ldots$$
$$\ldots, s_{k-1}, x_{k-1}, y^i)\ e^{-\beta\sum_{n=0}^{k-1} s_n}\ e^{-\beta s_k} g\left(x_k, y^k, \pi_k^D(h_{k-1}, \pi_{k-1}^D(h_{k-1}), s_k, x_k)\right).$$

The latter expression can be written using the factorization μ of the filter as well as using the function v^{π^D} defined at the beginning of this proof and we get:

$$= \sum_{i=1}^{q} \mu_{k-1}^i(h_{k-1}) \int \tilde{Q}_{SXY}^{\pi^D,k-1}(ds_k \otimes dx_k \otimes dy^k \mid s_0, x_0, \ldots, s_{k-1}, x_{k-1}, y^i)$$
$$v^{\pi^D}(h_{k-1}, \pi_{k-1}^D(h_{k-1}), s_k, x_k, y^k).$$

Applying Lemma 2.35 to the function v^{π^D}, this is equal to

$$= \sum_{i=1}^{q} \mu_{k-1}^i(h_{k-1}) \int \tilde{Q}_{SXY}^{\pi^D,k-1}(ds_k \otimes dx_k \otimes E_Y^0 \mid s_0, x_0, \ldots, s_{k-1}, x_{k-1}, y^i)$$
$$\sum_{j=1}^{q} \chi^j\left(\mu_{k-1}(h_{k-1}), s_k, x_k, \pi_{k-1}^D(h_{k-1})\right)\ v^{\pi^D}(h_{k-1}, \pi_{k-1}^D(h_{k-1}), s_k, x_k, y^j).$$

Now applying the definitions of v^{π^D} as well as of g', this becomes

$$= \sum_{i=1}^{q} \mu_{k-1}^{i}(h_{k-1}) \int \tilde{Q}_{SXY}^{\pi^D,k-1}(ds_k \otimes dx_k \otimes E_Y^0 \mid s_0, x_0, \dots, s_{k-1}, x_{k-1}, y^i)$$
$$e^{-\beta \sum_{n=0}^{k-1} s_n}\, e^{-\beta s_k} g'\left(x_k, \chi(\mu_{k-1}(h_{k-1}), s_k, x_k, \pi_{k-1}^D(h_{k-1})), \pi_k^D(h_{k-1}, \pi_{k-1}^D(h_{k-1}), s_k, x_k)\right).$$

Finally, applying the definition of the transition law for the derived filtered process, this transforms into

$$= \int Q'^{\,\pi^D,k-1}_{SXM}(ds_k \otimes dx_k \otimes d\rho_k \mid s_0, x_0, \dots, s_{k-1}, x_{k-1}, \mu_{k-1}(h_{k-1}))$$
$$e^{-\beta \sum_{n=0}^{k-1} s_n}\, e^{-\beta s_k} g'\left(x_k, \rho_k, \pi_k^D(h_{k-1}, \pi_{k-1}^D(h_{k-1}), s_k, x_k)\right)$$
$$= \mathbb{E}_{xQ_0}^{\pi^D}\left[e^{-\beta \sum_{n=0}^{k} \tilde{S}_n} g'(\tilde{X}_k, M_k, \pi_k^D(\tilde{H}_k) \mid \tilde{H}_{k-1} = h_{k-1}\right].$$

This terminates the proof for the first statement of the proposition. The second statement $\tilde{J}(x) = J'(x, Q_0(x, \cdot))$ is an immediate consequence of the first one when minimizing in both cases over the class Π^D. □

Restriction to Markov policies for the derived filtered process :
With Proposition 2.36 we showed that solving the optimal control problem for the time-discrete pseudo-embedded process (thus, for a PO-MDP) is equivalent to solving the control problem for the derived filtered process (thus, for an MDP) if for both, the class Π^D of history dependent relaxed piecewise open loop policies is admissible.

For completely observable MDPs, however, people usually only use Markov policies as admissible policies. Markov policies are not composed of history dependent decision rules but only of decision rules depending on the current state of the process. The reason why people only look at Markov policies is a well-known result of Hinderer. He showed that for fully observable time-discrete optimal control problems, the value function is not improved if optimization is done over the class of history dependent policies instead of Markov policies only.

Hence, we will introduce the class of Markov policies Π^M for the derived filtered process to then explain how to apply Hinderer's result. Combining both, we can then conclude that optimizing over the class of Markov policies Π^M is enough for calculating the value function J' and an optimal policy, if existent, shall be found in Π^M.

Definition 2.37. *For a derived filtered model, we define:*

(i) *A* state dependent decision rule *for the derived filtered model is a measurable mapping*

$$f : \mathbb{R}^+ \times E_X \times \mathbf{P}(E_Y^0) \to \mathcal{R}.$$

We write F for the set of all decision rules.

(ii) *A* Markov policy *for the derived filtered model is a sequence $\pi^M = (f_0, f_1, \dots)$ of state dependent decision rules, where we apply decision rule f_n at stage n of the process. We denote the set of all Markov policies for the filtered model by $\Pi^M := F^\infty = \times_{i=0}^{\infty} F$.*

(iii) *A* stationary policy *for the derived filtered model is a constant sequence $\pi^S = (f, f, \dots)$ with $f \in F$, i.e. we apply the same decision rule at every stage of the process. We denote the set of all stationary policies by Π^S.*

Based on this definition, we can now describe in what sense a Markov policy π^M can be understood as a special case of a history dependent policy and thus, we can understand Π^M as a subset of Π^D.

Lemma 2.38. *Let Q_0 an initial transition kernel from E_X to $\mathbf{P}(E_Y^0)$ for a pseudo-embedded process, Π^D the set of its history dependent relaxed control policies and Π^M the set of Markov policies for the corresponding derived filtered process. Then, the mapping*

$$\Pi^M \ni \pi^M = (f_0, f_1, \dots) \mapsto \tilde{\pi}^M := (\tilde{\pi}_0^M, \tilde{\pi}_1^M, \dots) \in \Pi^D,$$

is well-defined by setting, for all $n \geq 0$,

$$\tilde{\pi}_n^M : \mathcal{H}_n \to \mathcal{R},\ h_n = (s_0, x_0, r_0, \dots, s_n, x_n) \mapsto f_n(s_n, x_n, \mu_n(h_n)),$$

where we use the definition of the factorized filter $\mu_0((s_0, x_0)) := Q_0(x_0, \cdot)$ and for $n \geq 1$: $\mu_n(h_n) := \chi(\mu_{n-1}(h_{n-1}), s_n, x_n, r_{n-1})$.

Proof. We have to show that $\tilde{\pi}_n^M : \mathcal{H}_n \to \mathcal{R}$ is measurbale for all $n \geq 0$. As f_n is measurable from $\mathbb{R}^+ \times E_X \times \mathbb{P}(E_Y^0)$ to $\mathcal{R}$ by definition, it is sufficient to show that the mapping

$$i_{\mu_n} : \mathcal{H}_n \to \mathbb{R}^+ \times E_X \times \mathbb{P}(E_Y^0), h_n = (s_0, x_0, r_0, \dots, s_n, x_n) \mapsto (s_n, x_n, \mu_n(h_n))$$

is measurable for all $n \geq 0$.

For $n = 0$, this follows from $i_{\mu_0}((s_0, x_0)) = (s_0, x_0, Q_0(x_0, \cdot))$ by the measurability of Q_0 as transition kernel. For $n \geq 1$, we get the measurability of

$$i_{\mu_n}(h_n) = (s_n, x_n, \mu_n(h_n)) = (s_n, x_n, \chi(\mu_{n-1}(h_{n-1}), s_n, x_n, r_{n-1}))$$

by induction as χ is measurable. □

The next result shows that the set of Markov policies is large enough to find an optimal policy for a time-discrete fully observable optimization problem as given, e.g. by the derived filtered process.

Lemma 2.39. *In the situation of Definition 2.34, it holds*

$$\inf_{\pi^D \in \Pi^D} J'(x, Q_0(x, \cdot), \pi^D) = \inf_{\pi^M \in \Pi^M} J'(x, Q_0(x, \cdot), \pi^M). \tag{2.43}$$

Proof. Follows from [41], Theorem 18.4. □

With this result in mind, a solution approach for our initial, time-continuous, partially observable optimal control problem for the PO-PDMP $(S_t, X_t, Y_t)_{t \geq 0}$ is the following: we look at its embedded process $(S_n, X_n, Y_n)_{n \in \mathbb{N}_0}$ and construct the corresponding derived filtered process $(S_n, X_n, M_n)_{n \in \mathbb{N}_0}$ based on the filter function χ. We then try to find an optimal policy $\pi^{\star M} \in \Pi^M$ for the derived filtered process and deduct the associated history dependent policy $\pi^{\star D}$ by Lemma 2.38. According to Proposition 2.36, this policy $\pi^{\star D}$ is also optimal for the control problem on the embedded process $(S_n, X_n, Y_n)_{n \in \mathbb{N}_0}$. Finally, by Proposition 2.15, we know that $\pi^{\star D}$ has a corresponding policy $\pi^{\star P} \in \Pi^P$ that is optimal for our initial, time-continuous optimization problem as defined in Definition 1.37.

2.3 Existence of optimal policies for (lower semi-) continuous models

In this Section, we will prove the existence of optimal policies in the finite dimensional case under additional measurability and continuity assumptions. The necessity of these measurability and continuity assumptions, referred to as *(M&C)* in the sequel, lies in the the proof of Lemma 2.45 where we show the existence of one step optimizers. As decision rules $f \in F$ are measurable mappings from E' to $\mathcal{R}$, finding an optimal policy will always rely on some result of measurable selection of optimizers. The theory of stochastic dynamic programming has brought up a diverse set of such kind of measurable selection theorems for optimizers. Classical categories for these results are for exmaple (i) measurable selection for lower semi-continuous functions, (ii) measurable selection for lower semi-analytic functions and (iii) universally measurable selection. Results on all of these three categories can be found in [11], chapter 7. In view of possible applications of the theory we develop here, we focus on measurable selection for lower semi-continuous functions. A more detailed overview on possible applications of these lower semi-continuous models is given in Chapters 5 and 6.

The theory applied here as well as the step by step approach used is largely inspired by [6] as well as, and probably even more, by [34]. However, our model does not fit completely into the settings[6] used in both works, and thus, we will develop the theory here in a version adapted to the setting of our model.

The outline of this section is thus the following: In Subsection 2.3.1 we state the measurability and continuity assumptions *(M&C)* under which we will develop the rest of the theory. We then introduce, in Subsection 2.3.2 the classical operators from stochastic dynamic programming necessary to develop the further solution approach. After showing their most important properties, we will as well proof the existence of one step optimizers. This latter result is at the core of the subsequently following existence proofs for optimal policies for two time horizons: The N-stage model, where we minimize expected discounted cost up to the N-th jump time of the underlying PO-PDMP. Here, N is an upfront defined integer and we discuss this case in Subsection 2.3.3. The second case, $\mathcal{T}_\infty = \infty$, of an infinite time horizon is developed in Subsection 2.3.4. Finally, we will close this Section by discussing sufficient conditions for the model to satisfy the measurability and continuity assumptions *(M&C)* in Subsection 3.

2.3.1 Lower semi-continuity assumptions

In view of possible applications of the theory we develop here, we focus on measurable selection for lower semi-continuous functions. In order to be able to apply these results, we need to take the following assumptions for our model, that we will refer to as *(M&C)*, standing for *"Measurability and Continuity"* assumptions. In Section 3 we will give sufficient conditions for these assumptions to be satisfied.

Assumption 2.40 (Assumptions *(M&C)*). *Let E_X endowed with a topology consistent with the homeomorphism property of ψ (see Definition 1.8), let $\mathcal{R}$ endowed with the Young topology and let $\mathbf{P}(E_Y^0)$ endowed with the weak topology. Further, let $E_X \times \mathbf{P}(E_Y^0) \times \mathcal{R}$ endowed with the corresponding product topology and the corresponding Borel-σ-algebra. We take the following measurability assumptions:*

(M1) We assume $E_X \times \mathbf{P}(E_Y^0) \times \mathcal{R} \ni (x, \rho, r) \mapsto g'(x, \rho, r) \in \overline{\mathbb{R}}$ to be Borel-measurable.

[6] Bäuerle and Rieder work under so-called structure assumptions for general partially observabale Markov decision probblems and Forwick works on stationary problems for fully observable PDMPs

(M2) We assume the transition law q'_{SXM} of the derived filtered process to be a Borel-measurable transition kernel on E' given $E' \times \mathcal{R}$.

We further take the following lower semi-continuity and continuity assumption:

(LSC) We assume $E_X \times \mathbf{P}(E_Y^0) \times \mathcal{R} \ni (x, \rho, r) \mapsto g'(x, \rho, r)$ to be lower semi-continuous and non-negative.

(C) We assume

$$\tilde{q'}(B \mid (s, x, \rho), r) := \frac{\int_{E'} e^{-\beta s'} \mathbb{1}_B(s', x', \rho') \; q'_{SXM}(ds', dx', d\rho' \mid s, x, \rho, r)}{\int_{E'} e^{-\beta s'} \; q'_{SXM}(ds', dx', d\rho' \mid s, x, \rho, r)}$$

to be a continuous transition kernel w.r.t. the weak topology on $\mathbf{P}(E')$. Note that the constant in the denominator is finite, as T_1 is $\mathbb{P}^r_{sx\rho}$-a.s. finite (see desity of T_1).

All these assumptions together, we will refer to as (M&C).

Even though the measurability assumptions would be satisfied if only (LSC) and (C) were assumed, we list them separately here, in order to be able to better isolate in all following proofs whether only measurability or (semi-) continuity are required.

2.3.2 Minimum cost operator and existence of one step optimizer

This section can be understood as introductory to the following two sections. We will introduce the classical operators necessary for the standard approach derived from stochastic dynamic programming that will be used throughout the next two sections. Once given the definitions of these operators, we will list and prove important properties of them that we will use for the existence proofs in both cases, $\mathcal{T}_\infty = T_N$ as well as $\mathcal{T}_\infty = \infty$. Having introduced these operators and their properties, we will prove the existence of one step optimizers under assumptions *(M&C)* in the finite dimensional case.

We start with one important remark concerning the domain of definition of the operators we will need: Even though an initial state of the underlying optimal control problem for our controlled PO-PDMP will always imply an initial state of the derived filtered process of $(0, x, \rho)$, that is, of initial inter-jump time $s_0 = 0$, the theory applied below requires to generalize the model: In what follows, we will extend the derived filtered process to a model where arbitrary initial states of the form (s, x, ρ) are allowed, that is, initial inter-jump times of $s \neq 0$.

This generalization can be done easily, it basically means that the initial distribution of the state $(\tilde{S}_0, \tilde{X}_0, M_0)$ is simply $\delta_s \otimes \delta_x \otimes \delta_\rho$ instead of $\delta_0 \otimes \delta_x \otimes \delta_\rho$. This basically leads to a probability measure $\mathbb{P}^\pi_{sx\rho}$ instead of having a probability measure $\mathbb{P}^\pi_{x\rho}$ for the controlled derived filtered process but one can verify very quickly, that the theory still works the same way.

Important to notice is, that the one-step cost function $g'(x, \rho, r)$ will not be affected by this generalization of the model as by Definition 2.32, we have

$$g'(x, \rho, r) := \sum_{j=1}^{q} g(x, y^j, r)\rho(\{y^j\}),$$

and from Definition 2.1, we get

$$g(x, y, r) := \mathbb{E}^r_{x,y}\left[\int_0^{T_1} e^{-\beta t} \int_A c(x, \Phi^r(y, t), a) \; r_t(da) \; dt\right],$$

and thus, g' is independent of s. Actually, $\mathbb{E}^r_{xy}[T_1] = \mathbb{E}^r_{sxy}[T_1]$ as the density of the first jump-time does not depend on s. In practice, later on, r is dependent on s as $r = f_0(s, x, \rho)$ but this is only due to the way how we select the control r, it is not intrinsic to the definition of g'.

The classical operators from stochastic dynamic programming operate on functions defined on the full state space of the underlying stochastic process. Hence, in our case, functions defined for a state (s, x, ρ). We therefore simply set

$$g'(s, x, \rho, r) := g'(x, \rho, r). \tag{2.44}$$

We further define the cost of a policy $\pi \in \Pi^M$ for a process started in (s, x, ρ) as

$$J'(s, x, \rho, \pi) := \mathbb{E}^\pi_{sx\rho}\left[\sum_{k=0}^{\infty} e^{-\beta \sum_{n=1}^{k} \tilde{S}_n} g'(\tilde{S}_k, \tilde{X}_k, M_k, \pi(\tilde{S}_k, \tilde{X}_k, M_k)\right],$$

as well as the value function for a process started in this point as

$$J'(s, x, \rho) := \inf_{\pi \in \Pi^M} J'(s, x, \rho, \pi).$$

For $s = 0$, which would be the case where we look at an initial state coming from our controlled PO-PDMP, these definitions fit to the earlier introduced definitons (Note that we skip $\tilde{S}_0$ in the discount factor above as there is no sense in having the first periode discounted only because of a process start with $s \neq 0$).

We start with the following definition of two classes of functions that will be of particular importance for the existence proofs:

Definition 2.41. *For the state space $E' = \mathbb{R}^+ \times E_X \times \mathbf{P}(E^0_Y)$, we define the following classes of functions:*

$$\hat{\mathbf{B}}^+(E') := \{w : E' \to [0, \infty] \mid \ w \text{ is measurable}\}$$

$$\hat{\mathbf{C}}^+_{low}(E') := \{w : E' \to [0, \infty] \mid \ w \text{ is lower semi-continuous}\}.$$

The following operators are key for the existence proofs. In the case $T_\infty = \infty$, we will actually show that J' is a fixed point for one of them, namely $\mathcal{T}$ below. In the case $T_\infty = T_N$ for some $N \in \mathbb{N}$, we will show that the value function can be calculated by iterating the operator $\mathcal{T}$ below.

Definition 2.42. *For $(s, x, \rho) \in E'$, $r \in \mathcal{R}$, $f \in F$ and $w \in \hat{\mathbf{B}}^+(E')$ we define:*

(i) $(Hw)((s, x, \rho), r) := g'(s, x, \rho, r) + \int_{E'} e^{-\beta s'} \ w(s', x', \rho') \ q'_{SXM}(ds', dx', d\rho' \mid \rho, r)$

(ii) $(\mathcal{T}_f w)(s, x, \rho) := (Hw)((s, x, \rho), f(s, x, \rho))$

(iii) $(\mathcal{T} w)(s, x, \rho) := \inf_{r \in \mathcal{R}} (Hw)((s, x, \rho), r) = \inf_{f \in F} (\mathcal{T}_f w)(s, x, \rho)$.

Remark 2.43. *Writing $q'_{SXM}(\cdots \mid \rho, r)$ in part (i) of the above definition is valid: A closer look on Definition 2.32 shows that the transition law q'_{SXM} of the derived filtered process is actually only depending on ρ and r but not on s and x as one may recongnize when analyzing the transition law $\tilde{q}_{SXY}$ of the underlying pseudo-embedded process (see Definition 2.6).*

We start by highlighting all properties that are satisfied by the above defined operators and that, at the same time, are crucial for the proofs of the existence of optimal policies.

Lemma 2.44. *The above defined operators $H, \mathcal{T}_f$ and $\mathcal{T}$ satisfy the following properties when applied to functions of $\hat{\mathbf{B}}^+(E')$ or $\hat{\mathbf{C}}^+_{low}(E')$:*

(1) $w \in \hat{\mathbf{B}}^+(E') \implies E' \times \mathcal{R} \ni ((s,x,\rho),r) \mapsto (Hw)((s,x,\rho),r) \in [0,\infty]$ *is measurable.*

$w \in \hat{\mathbf{C}}^+_{low}(E') \implies E' \times \mathcal{R} \ni ((s,x,\rho),r) \mapsto (Hw)((s,x,\rho),r) \in [0,\infty]$ *is lower semi-continuous.*

(2) $w \in \hat{\mathbf{B}}^+(E') \implies \mathcal{T}_f w \in \hat{\mathbf{B}}^+(E')\ \forall f \in F$.

(3) $w \in \hat{\mathbf{C}}^+_{low}(E') \implies \mathcal{T} w \in \hat{\mathbf{C}}^+_{low}(E')$.

(4) $w, w' \in \hat{\mathbf{B}}^+(E')$ *and* $w \leq w' \implies (Hw)((s,x,\rho),r) \leq (Hw')((s,x,\rho),r)\ \forall (s,x,\rho) \in E', r \in \mathcal{R}$.

(5) $w, w' \in \hat{\mathbf{B}}^+(E')$ *and* $w \leq w' \implies \mathcal{T}_{f_0} \cdots \mathcal{T}_{f_k} w \leq \mathcal{T}_{f_0} \cdots \mathcal{T}_{f_k} w'\ \forall f_0, \ldots, f_k \in F, k \in \mathbb{N}$.

(6) $w, w' \in \hat{\mathbf{B}}^+(E')$ *and* $w \leq w' \implies \mathcal{T} w \leq \mathcal{T} w'$.

(7) $w, w' \in \hat{\mathbf{C}}^+_{low}(E')$ *and* $w \leq w' \implies \mathcal{T}^k w \leq \mathcal{T}^k w'\ \forall k \in \mathbb{N}$.

(8) $w \in \hat{\mathbf{C}}^+_{low}(E') \implies \mathcal{T}^k w \leq \mathcal{T}_{f_0} \cdots \mathcal{T}_{f_{k-1}} w\ \forall k \in \mathbb{N}, \forall f_0, \ldots, f_{k-1} \in F$.

(9) $(w_n)_{n\in\mathbb{N}} \subset \hat{\mathbf{B}}^+(E')$ *and* $w_n \uparrow w \implies w \in \hat{\mathbf{B}}^+(E')$ *and* $(Hw_n)((s,x,\rho),r) \uparrow (Hw)((s,x,\rho),r)\ \forall (s,x,\rho) \in E', r \in \mathcal{R}$.

Proof. (1) The first statement follows under Assumption 2.40 directly from [11], Proposition 7.29: As $\mathbb{R}^+, E_X$ and E_Y are complete separable and metrizable spaces, $E' = \mathbb{R}^+ \times E_X \times \mathbf{P}(E_Y^0)$ is a Borel space and $\mathcal{R}$ is a Borel space as well (see Annex A). Further, if $w \in \hat{\mathbf{B}}^+(E')$, then $(s,x,\rho) \mapsto e^{-\beta s} w(s,x,\rho)$ is Borel-measurable as well and thus,

$$((s,x,\rho),r) \mapsto \int_{E'} e^{-\beta s'}\ w(s',x',\rho')\ q'_{SXM}(ds', dx', d\rho' \mid \rho, r)$$

is Borel-measurable according to [11], Proposition 7.29 under Assumption 2.40. Under the latter assumption, we also have g' measurable and thus, Hw is measurable.

For the second statement, notice that the sum of two lower semi-continuous functions is again lower semi-continuous if both functions are bounded from below. By Assumption 2.40 (LSC), we have g' lower semi-continuous and bounded from below. For the integral part in the definition of H, we can apply [11], Proposition 7.31 (a) under Assumption 2.40 (C).

Finally, notice that $(Hw)((s,x,\rho),r) \in [0,\infty]$ as $w \geq 0$ and $g' \geq 0$ according to (LSC).

(2) Follows from first statement of part (1) as $(s,x,\rho) \mapsto (\mathcal{T}_f w)(s,x,\rho)$ is a composition of measurable mappings as f is measurable by definition of F.

(3) Lower semi-continuity follows from second statement of part (1) and [11], Proposition 7.32 as $\mathcal{R}$ is compact. Hw being bounded from below, this follows for $\mathcal{T} w$ as well.

(4) Monotonicity of integral.

(5) From (2), we see that $\mathcal{T}_f w \in \hat{\mathbf{B}}^+(E')$ for all $f \in F$ and all $w \in \hat{\mathbf{B}}^+(E')$. As consequence of (4), we get for $f \in F$ and $w, w' \in \hat{\mathbf{B}}^+(E')$ with $w \leq w' : \mathcal{T}_f w \leq \mathcal{T}_f w'$. Now, the statement follows by induction.

(6) Follows from (4).

(7) Analogously to (5) taking into account (3) and (6).

(8) For $k = 1$, this is a consequence of the definition of the operator $\mathcal{T}$. If the statement holds up to $k-1$, then:

$$\mathcal{T}^k w = \mathcal{T}\left(\mathcal{T}^{k-1} w\right) \leq \mathcal{T}\left(\mathcal{T}_{f_1} \cdots \mathcal{T}_{f_{k-1}} w\right) \leq \mathcal{T}_{f_0} \mathcal{T}_{f_1} \cdots \mathcal{T}_{f_{k-1}} w,$$

where the first inequality holds because of (6) as by induction assumption, $\mathcal{T}^{k-1} w \leq \mathcal{T}_{f_1} \cdots \mathcal{T}_{f_{k-1}} w$ (on both sides we have functions of class $\hat{\mathbf{B}}^+(E')$ by (3) and by (2)). The second inequality holds by definition of $\mathcal{T}$.

(9) Direct consequence of monotone convergence theorem. □

The next result is at the core of the proofs for the existence of optimal policies in both cases, infinite time horizon as well as finite N-stage time horizon. We will now prove the existence of a one step optimizer. It is this result as well, which requires our assumptions *(M&C)* on lower semi-continuity of the model even though, we use them here only indirectly via the semi-continuity property for H as stated in 2.44(1).

Finally, the following Lemma requires a first important result regarding the space $\mathcal{R}$: The space $\mathcal{R}$ is compact under the Young topology. Up to this point in this thesis, it was enough to simply know that there is a special topology we use for the space $\mathcal{R}$ without knowing further details on this topology. Now, to fully understand the following Lemma, we refer to Annex A for an introduction to the Young topology and a proof of the fact that $\mathcal{R}$ is compact and metrizable under the Young topology.

Lemma 2.45 (Existence of one step minimizers). *For all $w \in \hat{\mathbf{C}}^+_{low}(E')$ there exists $f^\star \in F$ such that*

$$\left(\mathcal{T}_{f^\star} w\right)(s, x, \rho) = \left(\mathcal{T} w\right)(s, x, \rho) \qquad \forall (s, x, \rho) \in E'.$$

Proof. This follows from [11], Proposition 7.33. Actually, we have: $E' = \mathbb{R}^+ \times E_X \times \mathbf{P}(E_Y^0)$ is a metrizable space as product of metrizable spaces. The space $\mathcal{R}$ is compact and metrizable according to Proposition A.26 and Lemma A.20. The function $Hw : E' \times \mathcal{R} \to \overline{\mathbb{R}}$ is lower semi-continuous for all $w \in \hat{\mathbf{C}}^+_{low}(E')$ according to 2.44(1). We further have $(\mathcal{T} w)(s, x, \rho) := \inf_{r \in \mathcal{R}} (Hw)((s, x, \rho), r)$. The assumptions of [11], Proposition 7.33, are thus fullfilled and so, there exists a Borel-measurable function $f^\star : E' \to \mathcal{R}$ such that

$$(Hw)((s, x, \rho), f^\star(s, x, \rho)) = (\mathcal{T} w)(s, x, \rho).$$

Now the result follows by definition of $\mathcal{T}_{f^\star}$. □

This closes the necessary pre-work and we can pass to the existence results for optimal policies under assumptions *(M&C)*. We will first investigate the N-stage model in the following section before we pass to the infinite time horizon thereafter.

2.3.3 Existence of optimal policies: Finite N-stage horizon

We will now investigate the N-stage problem, that is, the initial optimization problem from Definition 1.37 with $T_\infty = T_N$ for some $N \in \mathbb{N}$. With other words, we try to minimize the total expected discounted cost up to the N-th jump time of the PO-PDMP, where N is an upfront defined integer.

The approach used here for the existence proof is standard and well known under „reward iteration“ for problems of reward maximization of MDPs. Cost iteration as used below is simply the analogon of this technique. The main result of this section, Theorem 2.51 below, can be understood as an example for the so-called „Structure Theorem“ of the book of Bäuerle and Rieder (Theorem 2.3.8 in [6]). We have chosen to develop the full proof here in a version tailored to the situation of a derived filtered process.

Hence, we start by defining the N-stage problem for a controlled PO-PDMP to then quickly explain why the equivalence to the presented N-stage problem for a derived filtered process follows from all we have shown so far. We then give the definition of the value function at time $n \leq N$ for the remaining stages until N before we prove the cost iteration theorem. The existence result for an optimal policy in the N-stage model then closes this section.

Definition 2.46. *Let $N \in \mathbb{N}$ a fixed number of process jumps in scope. For $\pi \in \times_{n=0}^{N-1} \Pi_n^P$, we define the N-*stage cost of policy *π under initial observation $x \in E_X$ as*

$$J^N(x,\pi) := \mathbb{E}_x^\pi \left[\int_0^{T_N} e^{-\beta t} \int_A c(X_t, Y_t, a)\ \pi_t(da)\ dt \right],$$

where c, β, X_t and Y_t as in Definition 1.37.

The value function *of the N-stage problem gives the minimal cost up to stage N under an initial observation $x \in E_X$ and is defined as*

$$J^N(x) := \inf_{\pi \in \times_{n=0}^{N-1} \Pi_n^P} J^N(x,\pi) \quad \forall\ x \in E_X.$$

The optimization problem *is then to find, for $x \in E_X$, a policy $\pi^\star \in \times_{n=0}^{N-1} \Pi_n^P$ such that we get*

$$J^N(x) = J^N(x, \pi^\star).$$

We take now a slight shortcut and define the N-stage optimization problem for a derived filtered process without passing by the corresponding problem for the pseud-embedded process.

Definition 2.47. *Let $N \in \mathbb{N}$ and $(\tilde{S}_n, \tilde{X}_n, M_n)_{n\geq 0}$ the derived filtered process of a PO-PDMP with initial conditional distribution Q_0. For an initial observation $(0, x, Q_0(x,\cdot))$ of this process, we define the N-stage cost of a policy $\pi \in F^N$ by*

$$J'^N(x, Q_0(x,\cdot), \pi) := \mathbb{E}_{xQ_0}^\pi \left[\sum_{k=0}^{N-1} e^{-\beta \sum_{n=0}^k \tilde{S}_n} g'(\tilde{X}_k, M_k, \pi_k(H_k)) \right],$$

where g' is as of Definition 2.32.

The value function *of the derived filtered N-stage model gives the minimal cost up to stage N under an initial observation $(0, x, Q_0(x,\cdot))$ and is defined for all $x \in E_X, Q_0(x,\cdot) \in \mathbf{P}(E_Y^0)$ as*

$$J'^N(x, Q_0(x,\cdot)) := \inf_{\pi \in F^N} J'^N(x, Q_0(x,\cdot), \pi). \tag{2.45}$$

The filtered optimization problem *is then to find, for* $x \in E_X, Q_0(x,\cdot) \in \mathbf{P}(E_Y^0)$*, a policy* $\pi^\star \in F^N$ *such that we get*

$$J'^N(x, Q_0(x,\cdot)) = J'^N(x, Q_0(x,\cdot), \pi^\star).$$

The next result is a simple consequence of the fact that all proofs done so far for the case $T_\infty = \infty$ are also valid or can be done analogously in the case of a finite N-stage time horizon.

Proposition 2.48. *For all* $N \in \mathbb{N}$*, all possible initial transition kernels* Q_0 *of a PO-PDMP and all* $\pi^{M,N} = (f_0, \dots, f_{N-1}) \in F^N$ *it holds*

$$J^N(x, \pi^{P,N}) = J'^N(x, Q_0(x,\cdot), \pi^{M,N}),$$

where $\pi^{P,N} = (\pi_0^P, \dots, \pi_{N-1}^P)$ *is the corresponding policy (in the sense of the correspondence Theorem 2.11) of* $\pi^{D,M} = (\pi_0^D, \dots, \pi_{N-1}^D)$ *with* $\pi_k^D(h_k) := f_k(s_k, x_k, \mu(h_k))$ *for* $k = 0, \dots, N-1$.

Furthermore, we have

$$J^N(x) = J'^N(x, Q_0(x,\cdot)).$$

Proof. Follows from Proposition 2.36, Proposition 2.15 and Lemma 2.3. All three results can actually be formulated and proven in the corresponding N-stage version. Sums will be finite sums up to $N-1$ instead of up to infinity, underlying state spaces will be N times the Cartesian products instead of infinite Cartesian products and finally, Ionescu-Tulcea in version for finite Cartesian products has to be applied at corresponding places in the proofs when dealing with the probability measures on the state spaces. □

This Proposition in mind, it is enough, to prove existence of an optimal policy for the N-stage problem of a derived filtered process. The classical backward induction approach from stochastic dynamic programming requires the following Definition:

Definition 2.49. *Let* $\pi = (f_0, \dots, f_{N-1}) \in F^N$ *and consider a derived filtered process with initial state* $(0, x_0, Q_0(x_0,\cdot))$*. For* $n \in \mathbb{N}_0$ *and* $N \in \mathbb{N}$ *with* $n \le N$ *we define the cost of policy* π *from stage* n *to stage* N*, discounted to time* T_n*, given that the process state at stage* n *is* (s, x, ρ)*, as:*

$$J'^N_{\pi,n}(s,x,\rho) := \mathbb{E}^{\pi,n}_{sx\rho}\left[\sum_{k=n}^{N-1} e^{-\beta \sum_{l=n+1}^{k} \tilde{S}_l} g'\left(\tilde{S}_k, \tilde{X}_k, M_k, f_k(\tilde{S}_k, \tilde{X}_k, M_k)\right)\right],$$

where we note by $\mathbb{E}^{\pi,n}_{sx\rho}$ *the expectation w.r.t. the probability measure* $\mathbb{P}^{\pi}_{xQ_0(x,\cdot)}(\cdot \mid \tilde{S}_n = s, \tilde{X}_n = x, M_n = \rho)$.

The value function from stage n *to stage* N*, discounted to time* T_n *is defined as*

$$J'^N_n(s,x,\rho) := \inf_{\pi \in F^N} J'^N_{\pi,n}(s,x,\rho).$$

The following result shows that the cost of an N-stage policy can be determined by backward induction with the help of the $\mathcal{T}_f$-operators defined earlier. We will write $\underline{0}$ for a function being constant of value zero.

Proposition 2.50 (Cost iteration). *Let* $N \in \mathbb{N}$ *and* $\pi = (f_0, \dots, f_{N-1}) \in F^N$ *an* N*-stage policy. For* $n = 0, 1, \dots, N-1$ *it holds:*

a) $J'^N_N = J'^N_{\pi,N} = \underline{0}$

b) $J'^N_{\pi,n} = \mathcal{T}_{f_n} J'^N_{\pi,n+1}$

c) $J'^N_{\pi,n} = \mathcal{T}_{f_n} \cdots \mathcal{T}_{f_{N-1}} \underline{0}$

Proof. Part a) is clear by definition of empty sum. Part c) follows from part b) by induction while remembering Lemma 2.44 (2). For part b), we find:

$$\begin{aligned}
&J'^N_{\pi,n}(s,x,\rho) \\
&= \mathbb{E}^{\pi,n}_{sx\rho}\left[\sum_{k=n}^{N-1} e^{-\beta\sum_{l=n+1}^{k}\tilde{S}_l} g'\left(\tilde{S}_k, \tilde{X}_k, M_k, f_k(\tilde{S}_k, \tilde{X}_k, M_k)\right)\right] \\
&= g'\left(s,x,\rho, f_n(s,x,\rho)\right) + \mathbb{E}^{\pi,n}_{sx\rho}\left[\sum_{k=n+1}^{N-1} e^{-\beta\sum_{l=n+1}^{k}\tilde{S}_l} g'\left(\tilde{S}_k, \tilde{X}_k, M_k, f_k(\tilde{S}_k, \tilde{X}_k, M_k)\right)\right]
\end{aligned}$$

The latter expectation can now be written as:

$$\begin{aligned}
&\mathbb{E}^{\pi,n}_{sx\rho}\left[\sum_{k=n+1}^{N-1} e^{-\beta\sum_{l=n+1}^{k}\tilde{S}_l} g'\left(\tilde{S}_k, \tilde{X}_k, M_k, f_k(\tilde{S}_k, \tilde{X}_k, M_k)\right)\right] \\
&\qquad = \mathbb{E}^{\pi,n}_{sx\rho}\left[\mathbb{E}^{\pi,n}_{sx\rho}\left[\sum_{k=n+1}^{N-1} e^{-\beta\sum_{l=n+1}^{k}\tilde{S}_l} g'(\cdots) \mid \sigma(\tilde{S}_{n+1}, \tilde{X}_{n+1}, M_{n+1})\right]\right] \\
&\qquad\qquad = \int Q'_{SXM}\left(ds' \otimes dx' \otimes d\rho' | s,x,\rho, f_n(s,x,\rho)\right) \\
&\qquad\qquad\quad e^{-\beta s'}\ \mathbb{E}^{\pi,n+1}_{s'x'\rho'}\left[\sum_{k=n+1}^{N-1} e^{-\beta\sum_{l=n+2}^{k}\tilde{S}_l} g'(\cdots)\right] \\
&\qquad\qquad = \int Q'_{SXM}\left(ds' \otimes dx' \otimes d\rho' | s,x,\rho, f_n(s,x,\rho)\right) \\
&\qquad\qquad\qquad\qquad e^{-\beta s'} J'^N_{\pi,n+1}(s',x',\rho') \qquad \square
\end{aligned}$$

The next result follows directly from Theorem 2.3.8 of [6] as we have shown already that the structure assumption (SA_N) made in [6] is satisfied by our model: We have no terminal cost, operator $\mathcal{T}$ applied to a function of class $\hat{\mathbf{C}}^+_{low}$ delivers a function of class $\hat{\mathbf{C}}^+_{low}$ and we have shown the existence of one step optimizers for functions of class $\hat{\mathbf{C}}^+_{low}$ in Lemma 2.45. Nonetheless, we will give, for the following main result of this section, a proof that is adapted to our model.

Theorem 2.51. *Let $N \in \mathbb{N}$ and let assumptions (M&C) as well as the finite dimensional case assumptions hold. Then, for the N-stage problem of a derived filtered process, it holds:*

a) $J'^N_N = J^N_{\pi,N} = \underline{0}$ *for all* $\pi \in F^N$.

b) For all $n = 0, 1, \ldots, N-1$ *we have*

(i) There exists a minimizer $f^\star_n \in F$ *such that* $\mathcal{T} J'^N_{n+1} = \mathcal{T}_{f^\star_n} J'^N_{n+1}$

(ii) $J'^N_n = \mathcal{T} J'^N_{n+1} \in \hat{\mathbf{C}}^+_{low}$.

(c) Every sequence of minimizers $(f^\star_0, \ldots, f^\star_{N-1})$ *as ob b) determines an optimal policy for the N-stage problem of the underlying derived filtered model.*

Proof. Part a) is clear by definition of empty sum. Part c) follows from part b): By induction one can show $J_0'^N = \mathcal{T}^N \underline{0} = \mathcal{T}_{f_0^\star} \cdot \mathcal{T}_{f_1^\star} \cdots \mathcal{T}_{f_n^\star} \underline{0}$. By definition, it holds $J_0'^N = J'^N$.

Part b), we prove by induction. For $n = N-1$ part (i) follows from a) and Lemma 2.45, note that $\underline{0} \in \hat{\mathbf{C}}_{low}^+$. Part (ii) then follows from the definitions of $\mathcal{T}$ and $J_{N-1}'^N$ together with Lemma 2.44 part (3).

Suppose now, that for $N-1, N-2, \ldots, n+1$ part b) holds. Then, as $J_{n+1}'^N \in \hat{\mathbf{C}}_{low}^+$, by Lemma 2.45, there exists an optimizer $f_n^\star$ with $\mathcal{T} J_{n+1}'^N = \mathcal{T}_{f_n^\star} J_{n+1}'^N$. By induction hypothesis and cost iteration (see previous proposition), we then find

$$\mathcal{T} J_{n+1}'^N = \mathcal{T}_{f_n^\star} J_{n+1}'^N = \mathcal{T}_{f_n^\star} \cdot \mathcal{T}_{f_{n+1}^\star} \cdots \mathcal{T}_{f_{N-1}^\star} \underline{0} = J_{\pi_n^\star, n}'^N,$$

where we note by $\pi_k^\star$ the policy $\pi_k^\star = (f_k^\star, \ldots, f_{N-1}^\star) \in F^{N-k}$. By definition of $J_n'^N$ it follows

$$\mathcal{T} J_{n+1}'^N \geq J_n'^N \tag{2.46}$$

On the other hand, for arbitrary $\pi_n = (f_n, \ldots, f_{N-1}) \in F^{N-n}$, it follows by cost iteration, by monotonicity of $\mathcal{T}_f$ operator (Lemma 2.44) and by definition of $J_{n+1}'^N$:

$$J_{\pi_n, n}'^N = \mathcal{T}_{f_n} J_{\pi_{n+1}, n+1}'^N \geq \mathcal{T}_{f_n} J_{n+1}'^N \geq \mathcal{T}_{f_n^\star} J_{n+1}'^N = \mathcal{T} J_{n+1}'^N.$$

Taking the infimum over all policies $\pi_n \in F^{N-n}$ leads to $J_n'^N \geq \mathcal{T} J_{n+1}'^N$ and together with (2.46) part (ii) follows remembering Lemma 2.44 part(3). □

We showed the existence of an optimal policy for the N-stage problem of a derived filtered model. By the equivalence shown between the N-stage problems for derived filtered models and for the underlying PO-PDMP model, the main result of this section is a Corollary of the previous result.

Corollary 2.52. *Under assumptions (M&C), an optimal policy $\pi^{\star P} \in \Pi^P$ exists for the N-stage problem of a PO-PDMP in the finite dimensional case. An optimal policy is given by the corresponding policy (in the sense of the correspondance Theorem) to*

$$\pi^{\star D}(h_k) := f_k^\star(s_k, x_k, \mu_k(h_k)),$$

where $(f_0^\star, \ldots, f_{N-1}^\star) \in F^N$ is an optimal policy for the derived filtered model and s_k, x_k are the last two components of h_k.

2.3.4 Existence of optimal policies: Infinite time horizon

We will now investigate the optimization problem for the infinite time horizon, i.e. for the case $T_\infty = \infty$. The goal of this section is twofold: While aiming to prove the existence of optimal policies for the derived filtered model, we will also give a characterization of the value function as fixed point of the earlier defined operator $\mathcal{T}$. This characterization is not only at the core of the existence of optimal *stationary* policies but also the key starting point for further characterizations of the concrete form of an optimal policy as we will see later in Chapter 5.

The approach followed in the sequel is a standard approach for MDPs under complete observation and can be found, for example, in [6]. The presented approach is inspired from the work of Forwick [34] who investigated existence of optimal policies for completely observable PDMPs. Given the differences in the state spaces and especially in the class of admissible policies (we use history dependent policies, in the case of completely observable PDMPs Forwick uses Markov policies for the PDMP already), we develop all intermediate results required to prove our main result.

In what follows, we present now a sequence of important intermediate results that will finally lead to the proof for the existence of optimal policies for the derived filtered model. The key ingredients, however, will also be the properties of the operators $\mathcal{T}$ and $\mathcal{T}_f$ (see Lemma 2.44), the existence of one step optimizers (see Lemma 2.45) and the cost iteration Proposition 2.50. An immediate consequence of the latter is

Corollary 2.53. *For all $(s,x,\rho) \in E'$ and $\pi = (f_0, f_1, \dots) \in \Pi^M$ we have for all $n \in \mathbb{N}$:*

$$(\mathcal{T}_{f_0} \cdots \mathcal{T}_{f_{n-2}} \mathcal{T}_{f_{n-1}} \, \underline{0} \,)\,(s,x,\rho) = \mathbb{E}^{\pi}_{sx\rho}\left[\sum_{k=0}^{n-1} e^{-\beta T_k} g'(\tilde{S}_k, \tilde{X}_k, M_k, f_k(\tilde{S}_k, \tilde{X}_k, M_k)\right],$$

where $T_k = \sum_{l=1}^{k} \tilde{S}_l$. Furthermore, we get

$$J'(s,x,\rho,\pi) = \lim_{n\to\infty} (\mathcal{T}_{f_0} \cdots \mathcal{T}_{f_{n-2}} \mathcal{T}_{f_{n-1}} \, \underline{0} \,)\,(s,x,\rho),$$

and $J'(\cdot,\cdot,\cdot,\pi) \in \hat{\mathbf{B}}^+(E')$.

Proof. Direct consequence of cost iteration Proposition 2.50. As $g' \geq 0$ (because cost function c assumed non-negative), monotone convergence leads to

$$J'(s,x,\rho,\pi) = \lim_{n\to\infty} (\mathcal{T}_{f_0} \mathcal{T}_{f_1} \cdots \mathcal{T}_{f_n} \, \underline{0} \,)\,(s,x,\rho).$$

Furthermore, Lemma 2.44(2) implies that the function $\mathcal{T}_{f_0} \mathcal{T}_{f_1} \cdots \mathcal{T}_{f_n} \underline{0}$ lies in $\hat{\mathbf{B}}^+(E')$ for all $n \in \mathbb{N}$ and thus, the pointwise limit $J'(\cdot,\cdot,\cdot,\pi)$ also lies in $\hat{\mathbf{B}}^+(E')$. □

Lemma 2.54. *For $w \in \hat{\mathbf{C}}^+_{low}(E')$ we have*

a) $\mathcal{T}_f w \leq w \Rightarrow J'(\cdot,\cdot,\cdot,(f,f,\dots)) \leq w \qquad \forall f \in F,$

b) $\mathcal{T} w \leq w \Rightarrow J' \leq w.$

Proof. a) Let $\mathcal{T}_f w \leq w$. Let now $k \geq 2$ and suppose that $\mathcal{T}_f^{k-1} w \leq w$ holds, then:

$$\mathcal{T}_f^k w = \mathcal{T}_f \left(\mathcal{T}_f^{k-1} w\right) \leq \mathcal{T}_f w \leq w,$$

where we use the induction assumption together with Lemma 2.44(2) and (5) for the first inequality and the initial assumption for the second inequality. We conclude that for all $k \in \mathbb{N}$ we have $\mathcal{T}_f^k w \leq w$ and thus, we also have $\liminf_{k\to\infty} \mathcal{T}_f^k w \leq w$.

With Corollary 2.53, we then get

$$J'(\cdot,\cdot,\cdot,(f,f,\dots)) = \lim_{k\to\infty} \mathcal{T}_f^k \underline{0} = \liminf_{k\to\infty} \mathcal{T}_f^k \underline{0} \leq \liminf_{k\to\infty} \mathcal{T}_f^k w \leq w.$$

b) For $w \in \hat{\mathbf{C}}^+_{low}(E')$ let $f^\star$ the one step minimizer as of Lemma 2.45. Then, given the assumption $\mathcal{T} w \leq w$, we have

$$\mathcal{T}_{f^\star} w = \mathcal{T} w \leq w,$$

thus, a) holds for $f^\star \in F$ and we get:

$$J' \leq J'(\cdot,\cdot,\cdot,(f^\star, f^\star, \dots)) \leq w$$

. □

Definition 2.55. *For $(s,x,\rho) \in E'$ we define $J^\infty(s,x,\rho) := \lim_{k\to\infty} \left(\mathcal{T}^k \underline{0}\right)(s,x,\rho)$.*

Remark 2.56. *J^∞ is well defined because $g' \geq 0$ implies $\mathcal{T}\underline{0} \geq 0$ and from this, we conlcude with 2.44(7) that $\mathcal{T}^k\underline{0} \geq T^{k-1}\underline{0}$ $\forall k$. The sequence $\left((\mathcal{T}^k\underline{0})(s,x,\rho)\right)_{k\in\mathbb{N}}$ is thus monotone and the pointwise limit is existing in $[0,\infty]$.*

Lemma 2.57. *The following statements hold:*

(1) $J^\infty \in \hat{\mathbf{C}}^+_{low}(E')$,

(2) $J^\infty \leq J'$,

(3) $\mathcal{T}J^\infty \geq J^\infty$,

(4) $\mathcal{T}J^\infty = J^\infty \Rightarrow J^\infty = J'$.

Proof. (1) As $\underline{0} \in \hat{\mathbf{C}}^+_{low}(E')$, Lemma 2.44(3) implies $\mathcal{T}^k\underline{0} \in \hat{\mathbf{C}}^+_{low}(E')$ for all $k \in \mathbb{N}$ and thus, $J^\infty \in \hat{\mathbf{C}}^+_{low}(E')$ as pointwise supremum of lower semi-continuous functions is lower semi-continuous.

(2) Lemma 2.44(8) implies

$$\begin{aligned}
\mathcal{T}^k\underline{0} &\leq \mathcal{T}_{f_0}\cdots\mathcal{T}_{f_{k-1}}\underline{0} && \forall k \in \mathbb{N}, \forall f_0,\dots,f_{k-1} \in F \\
\Rightarrow \lim_{k\to\infty} \mathcal{T}^k\underline{0} &\leq \lim_{k\to\infty} \mathcal{T}_{f_0}\cdots\mathcal{T}_{f_{k-1}}\underline{0} && \forall \pi = (f_0,f_1,\dots) \in \Pi^M \\
\Rightarrow J^\infty &\leq J'(\cdot,\cdot,\cdot,\pi) && \forall \pi \in \Pi^M \\
\Rightarrow J^\infty &\leq J'.
\end{aligned}$$

(3) By definition $J^\infty \geq \mathcal{T}^k\underline{0}$ $\forall k$ and thus (by monotonicity of $\mathcal{T}$ according to 2.44(6)), $\mathcal{T}J^\infty \geq \mathcal{T}(\mathcal{T}^k\underline{0}) = \mathcal{T}^{k+1}\underline{0}$ $\forall k$ and the statement follows by taking the limit for $k \to \infty$.

(4) From (2) we get $J^\infty \leq J'$, thus we need to show the implication $\mathcal{T}J^\infty = J^\infty \Rightarrow J' \leq J^\infty$. This, however, follows from Lemma 2.54 b). □

Theorem 2.58. *The function J^∞ satisfies the following properties:*

(1) $\mathcal{T}J^\infty = J^\infty$ *and*

(2) $J^\infty = J'$.

Proof. Statement (2) follows from (1) and Lemma 2.57(4). To show (1), it is enough to show $\mathcal{T}J^\infty \leq J^\infty$ (see Lemma 2.57(3)). With $\underline{0} \in \hat{\mathbf{C}}^+_{low}(E')$, we also have $\mathcal{T}^k\underline{0} \in \hat{\mathbf{C}}^+_{low}(E')$ for all $k \in \mathbb{N}$ (see Lemma 2.44(3)). Applying now Lemma 2.45 to $\mathcal{T}^k\underline{0}$, there is, for each $k \in \mathbb{N}$, a $f_k^\star \in F$ such that

$$\left(\mathcal{T}(\mathcal{T}^k\underline{0})\right)(s,x,\rho) = \mathcal{T}_{f_k^\star}(\mathcal{T}^k\underline{0})(s,x,\rho) \qquad \forall (s,x,\rho) \in E'.$$

Fix now $(s,x,\rho) \in E'$ and define $r_k := f_k^\star(s,x,\rho) \in \mathcal{R}$. As $\mathcal{R}$ is compact, the sequence $(r_k)_{k\in\mathbb{N}}$ has a convergent subsequence $(r_{k_l})_{l\in\mathbb{N}}$ with $\lim_{l\to\infty} r_{k_l} = r_\infty \in \mathcal{R}$.

For $l \in \mathbb{N}$ it now holds:

$$J^\infty(s,x,\rho) \geq \left(\mathcal{T}^{k_l+1}\underline{0}\right)(s,x,\rho) = \left(\mathcal{T}(\mathcal{T}^{k_l}\underline{0})\right)(s,x,\rho)$$
$$= \left(\mathcal{T}_{f^\star_{k_l}}(\mathcal{T}^{k_l}\underline{0})\right)(s,x,\rho) = \left(H\mathcal{T}^{k_l}\underline{0}\right)((s,x,\rho),r_{k_l}).$$

By monotonicity of H (see Lemma 2.44(4)), we further get for $k \leq k_l$:

$$J^\infty(s,x,\rho) \geq \left(H\mathcal{T}^{k_l}\underline{0}\right)((s,x,\rho),r_{k_l}) \geq \left(H\mathcal{T}^{k}\underline{0}\right)((s,x,\rho),r_{k_l}).$$

Looking now at the limit for $l \to \infty$ we find:

$$J^\infty(s,x,\rho) \geq \liminf_{l\to\infty}\left(H\mathcal{T}^{k}\underline{0}\right)((s,x,\rho),r_{k_l}) \geq \left(H\mathcal{T}^{k}\underline{0}\right)((s,x,\rho),r_\infty) \qquad \forall k \in \mathbb{N},$$

where the second inequality holds because of the lower semi-continuity of H (see Lemma 2.44(1), and for this classical property of lower semi-continuous functions see, e.g., [11], Lemma 7.13).

Applying now monotone convergence for $k \to \infty$ (see Lemma 2.44(9)), we get:

$$J^\infty(s,x,\rho) \geq (HJ^\infty)((s,x,\rho),r_\infty) \geq (\mathcal{T}J^\infty)(s,x,\rho),$$

and the last inequality is simply the definition of $\mathcal{T}$. □

Theorem 2.59. *The function $J' : E' \to [0,\infty]$ is lower semi-continuous and $\mathcal{T}J' = J'$.*

Proof. According to Theorem 2.58(2), we have $J' = J^\infty$ and from Lemma 2.57(1) we have $J^\infty \in \hat{\mathbf{C}}^+_{low}(E')$. As $J' = J^\infty$ and $\mathcal{T}J^\infty = J^\infty$ according to Theorem 2.58, the second statement follows. □

Having all these intermediate results in mind, we can now state and prove the second main result of this section: The existence of optimal policies for the derived filtered model for $T_\infty = \infty$ and thus, according to the theory developed before, for our initial optimization problem for a PO-PDMP under infinite time horizon.

Theorem 2.60. *The following statements hold:*

(1) Let $f^\star \in F$, then it holds:
$(\mathcal{T}_{f^\star}J')(s,x,\rho) = (\mathcal{T}J')(s,x,\rho)\ \forall (s,x,\rho) \in E' \Longrightarrow \pi^\star := (f^\star, f^\star, \dots)$ *is optimal*

(2) There exists an optimal stationary policy $\pi^\star \in \Pi^M$, i.e.

$$J'(s,x,\rho,\pi^\star) = \inf_{\pi\in\Pi^M} J'(s,x,\rho,\pi) \qquad \forall (s,x,\rho) \in E'$$

and $\pi^\star = (f^\star, f^\star, \dots)$ for a decision rule $f^\star \in F$.

Proof. (1) If for $f^\star \in F$ we have $\mathcal{T}_{f^\star}J' = \mathcal{T}J'$, then Theorem 2.59 implies $\mathcal{T}_{f^\star}J' = J'$.
Now, applying Lemma 2.54(a) to $J' \in \hat{\mathbf{C}}^+_{low}(E')$ and to $f^\star \in F$, we get

$$J'(\cdot,\cdot,\cdot,(f^\star,f^\star,\dots)) \leq J'$$

and thus, $\pi^\star = (f^\star, f^\star, \dots) \in \Pi^M$ is optimal.

(2) According to Theorem 2.59 we have $J' \in \hat{\mathbf{C}}^+_{low}(E')$. Lemma 2.45 then implies that there exists $f^\star \in F$ such that $\mathcal{T}_{f^\star}J' = \mathcal{T}J'$ and now, part (1) implies that $\pi^\star = (f^\star, f^\star, \dots) \in \Pi^M$ is optimal. □

We now got the existence of optimal policies $\pi^\star \in \Pi^M$ for the derived filtered process. We even learned that there is an optimal policy within the set of stationary policies. While proving this existence, we implicitly got some more insights into the structure of such an optimal stationary policy: An optimal stationary policy $\pi^\star$ for the derived filtered model does actually not depend on the inter-jump time s of a state (s, x, ρ) but only on x and ρ.

Theorem 2.61. *An optimal stationary policy $\pi^\star = (f^\star, f^\star, \dots) \in \Pi^M$ does not depend on the observed inter-jump time s, i.e.*

$$f^\star(s, x, \rho) = f^\star(0, x, \rho) \qquad \forall (s, x, \rho) \in E'.$$

Furthermore, $J'(s, x, \rho) = J'(0, x, \rho)$ for all $(s, x, \rho) \in E'$, i.e. J' is not depending on the observed inter-jump time.

Proof. According to Theorem 2.59, the value function J' satisfies $J' = \mathcal{T} J'$. Writing this equality using the definition of $\mathcal{T}$ we get for $(s, x, \rho) \in E'$:

$$\begin{aligned} J'(s, x, \rho) &= \left(\mathcal{T} J'\right)(s, x, \rho) \\ &= \inf_{r \in \mathcal{R}} \left(H J'\right)((s, x, \rho), r) \\ &= \inf_{r \in \mathcal{R}} \left\{ g'(s, x, \rho, r) + \int_{E'} e^{-\beta s'} J'(s', x', \rho')\ q'_{SXM}(ds', dx', d\rho' \mid \rho, r) \right\} \end{aligned}$$

Applying now (2.44) together with the definition of g' as well as the definition of q'_{SXM}, we get

$$\begin{aligned} = \inf_{r \in \mathcal{R}} \Bigg\{ &\sum_{i=1}^{q} g(x, y^i, r) \rho(\{y^i\}) + \sum_{i=1}^{q} \rho(\{y^i\}) \\ &\int_{[0,\infty] \times E_X} J'\left(s', x', \chi(\rho, s', x', r)\right)\ \tilde{Q}_{SXY}(ds' \otimes dx' \otimes E_Y^0 \mid y^i, r) \Bigg\}. \end{aligned}$$

Hence, $J'(s, x, \rho)$ is not depending on s but only on x and ρ. As a consequence, the optimizer $f^\star$ is not depending on s. □

The last result should not be surprising as there is no meaning in observing an initial inter-jump time other than zero, if not, this means, somebody forgot to set the time counter to zero before starting the experience. Actually, the cost function c is only depending on $\tilde{X}_k$ and $\tilde{Y}_k$, not on the inter-jump time $\tilde{S}_k$ and we only observe the inter-jump time $\tilde{S}_t$ to calculate the filter M_k. All information observed is then contained in the filter M_k, especially all information about observed inter-jump times.

In a sense, the last result „brings back to normal“ the „artificial“ extension we made to the problem at the beginning of this section when allowing initial inter-jump times of $s \neq 0$. This extension, however, made the definition of the operators H and $\mathcal{T}$ much more consistent to standard dynamic programming operators and finally, the proof of the cost iteration Proposition was possible. Without this extension, making appear $\mathbb{E}^{\pi, n+1}_{s' x' \rho'}$ would not have been possible in that proof.

We summarize our main result in the following

Corollary 2.62. *Under assumptions (M&C), there exists an optimal policy $\pi^{\star P} \in \Pi^P$ for the optimization problem stated for a PO-PDMP in the finite dimensional case. Such an optimal policy is given by the corresponding policy (according to correspondance Theorem) to the policy $\pi^{\star D} \in \Pi^D$ given by*

$$\pi_k^{\star D}(h_k) := f^\star(x_k, \mu_k(h_k)),$$

where $f^\star \in F$ is a decision rule for the derived filtered model that defines an optimal stationary policy to the latter and where x_k is the last component of $h_k \in \mathcal{H}_k$.

Chapter 3

Sufficient conditions for (lower semi-) continuity of the model

In the last Chapter, we proved the existence of optimal policies in the finite dimensional case under assumptions *(M&C)* on measurability and (lower semi-) continuity of the derived filtered model. The latter assumptions are assumptions made for the derived filtered model, not for the initial PO-PDMP model. The important question we want to investigate in this Chapter is thus: What (additional) assumptions do we need to take on the initial PO-PDMP model in order to find the derived filtered model satisfy assumptions *(M&C)*?

For completely observable PDMPs, one directly obtains a completely observable MDP when following the analogous version of our first reformulation of the problem. A version of assumptions *(M&C)*, adapted to this situation of complete observation, was given by Forwick in [34]. Forwick could prove, under these assumptions, existence of optimal policies for the MDP and thus for the underlying PDMP. In order to satisfy these assumptions, he only needed an additional Lipschitz-continuity assumption on the vector field defining the controlled drift via an ODE in his PDMP model.

Inspired by this result of Forwick, we will investigate the question whether in the partially observable case, an additional condition on the drift is enough as well in order to get existence of optimal policies.

As assumptions *(M&C)* are twofold by containing a lower semi-continuity condition on the one step cost function g' of the derived filtered model and a weak continuity condition on the transition kernel $\tilde{q}'$, we will split our investigation in two separate streams: First, we will investigate what conditions on the initial PO-PDMP are sufficient to get the derived filtered model satisfy assumption (LSC). This is done is Section 3.1. Second, we will provide a sufficient condition on the initial PO-PDMP in order to get the transition kernel $\tilde{q}'$ satisfy assumption (C) of weak continuity. This is done in Section 3.2

It will turn out that we have to restrict our initial optimization problem to problems where only the drift Φ of the unobservable state of the PO-PDMP can be controlled by the agent but not the jump rate λ nor the transition kernel Q. This is due to the filter χ not being continuous in the variable r which finally is linked back to properties of the Young topology on $\mathcal{R}$. We discuss this issue in detail in Section 3.2.2. In Chapter 4 we will then show how to get a „sufficiently continuous“ filter (in a sense to be defined properly) for models where the inter-jump time is not observable and thus, the filter not depending on the inter-jump time.

3.1 Lower semi-continuity of the one step cost function

In this Section, we will show that the one step cost function g' of the derived filtered model is lower semi-continuous whenever the cost function of the PO-PDMP is lower semi-continuous and the drift Φ has a continuous dependence on the relaxed control r. It will turn out, that the latter continuous dependence is assured for ODE defined drifts under some Lipschitz condition on the vector field intervening in the ODE.

So far, we did not use any specific property of the Young topology on $\mathcal{R}$ except the fact that the space $\mathcal{R}$ is compact under the Young topology. The latter fact was used in the proof of the existence of one step optimizers. In order to prove now the lower semi-continuity of g', we will need some characterization of the convergence w.r.t. the Young topology. We provide the necessary theory in Annex A, where Lemma A.21 is of special interest for the main results of this Paragraph. The reader not familiar with convergence w.r.t. the Young topology might want to read this Annex first. Remember as well, that, for some metric space M we denote by $\mathbf{C}(M)$ the set of all continuous (w.r.t. the topology induced by the metrics) and bounded functions from M to $\mathbb{R}$.

We start by summarizing the two assumptions that are sufficient to take on the PO-PDMP in order to get a derived filtered model with lower semi-continuous one step cost function g'.

Assumption 3.1 (Continuous dependence of drift on relaxed control). *We assume the mapping $r \mapsto \Phi^r(y^k, t)$ to be continuous for all $y^k \in E_Y^0$ and $t \geq 0$.*

Assumption 3.2 (Lower semi-continuous cost function). *We assume the cost function $c : E_X \times E_Y \times A \to \mathbb{R}^+$ to be lower semi-continuous w.r.t. the product topology.*

While assuming a lower semi-continuous cost function c seems reasonable at first glance (also in view of possible applications), the continuous dependence of the drift Φ on the relaxed control r might seem as a serious restriction to the model. Many applications, however, are built on ODE defined drifts and in such cases, a Lipschitz continuity condition on the vector field in the ODE is sufficient to guarantee the required continuous dependence of Φ on r. We summarize this in the following Theorem.

Theorem 3.3. *Let $E_Y = \mathbb{R}^d$ for some $d \in \mathbb{N}$ and let $b : \mathbb{R}^d \times A \to \mathbb{R}^d$ continuous. If there is a constant $L > 0$ such that for all $x, y \in \mathbb{R}^d$ and for all $a \in A$, we have*

$$\|b(x,a) - b(y,a)\| \leq L \|x - y\|,$$

then, the initial value problem

$$\frac{d}{dt}\Phi^r(y,t) = \int_A b\left(\Phi^r(y,t), a\right) \; r_t(da) \qquad\qquad \Phi^r(y,0) = y$$

admits a unique solution for all $r \in \mathcal{R}$ and $y \in \mathbb{R}^d$ and this solution $\Phi^r(y,\cdot) : [0,\infty) \to \mathbb{R}^d$ satisfies Assumption 3.1.

Proof. Requires the Gronwall inequality. A full proof of this Theorem as well as of the Gronwall inequality in the version required for the proof of the theorem can be found in [34], Theorem 2.2.6 and Lemma 2.2.7. □

We start now by deriving a representation of the one step cost function g' that will be helpful in the sequel. First, we need the following conditional density for inter jump times.

Lemma 3.4. *Let $(\tilde{S}_k, \tilde{X}_k, \tilde{Y}_k)_{k\in\mathbb{N}}$ a pseudo-embedded process with jump intensity λ and drift Φ. Then, the probability distribution of $\tilde{S}_1$ given $\tilde{Y}_0 = y$ under relaxed control $r \in \mathcal{R}$ has a density $f^r_{\tilde{S}_1}(s \mid y)$ that is given by*

$$f^r_{\tilde{S}_1}(s \mid y) = \exp\left(-\Lambda^r(y,s)\right) \int_A \lambda^A\left(\Phi^r(y,s),a\right)\ r_s(da),$$

where we define

$$\Lambda^r(y,s) := \int_0^s \int_A \lambda^A\left(\Phi^r(y,\tau),a\right)\ r_\tau(da)\ d\tau.$$

Proof. Follows directly from the definition of the transition law $\tilde{Q}_{SXY}$ (see Definition 2.6(ii)) of the pseudo-embedded process as

$$\begin{aligned}\mathbb{P}^r_{xy}(\tilde{S}_1 \le t) &= \tilde{Q}_{SXY}([0,t]\times E_X \times E_Y \mid 0,x,y,[r]) \\ &= \int_0^t \exp\left(-\int_0^s \int_A \lambda^A\left(\Phi^r(y,\tau),a\right)\ r_\tau(da)\ d\tau\right) \int_A \lambda^A\left(\Phi^r(y,s),a\right)\ r_s(da)\ ds.\end{aligned}$$

To simplify notations, we define:

Definition 3.5. *Let β the discount factor of the discounted cost function. We then define*

$$\eta^r(y^k,t) := \exp\left(-\beta t - \Lambda^r(y^k,t)\right),$$

where $r \in \mathcal{R}, y^k \in E^0_Y$ and $t \in \mathbb{R}^+$.

With this notation, we can represent the one step cost function as follows:

Lemma 3.6. *The one step cost function g' can be written as*

$$g'(x,\rho,r) = \sum_{k=1}^q \rho_k \int_0^\infty \eta^r(y^k,t) \int_A c\left(x,\Phi^r(y^k,t),a\right)\ r_t(da)\ dt,$$

where we use the notation $\rho_k := \rho(\{y^k\})$ for $k \in \{1,\dots,q\}$ and $y^k \in E^0_Y$.

Proof. By definition of g' and g we obtain:

$$\begin{aligned}g'(x,\rho,r) &= \sum_{k=1}^q \rho_k\ g(x,y^k,r) \\ &= \sum_{k=1}^q \rho_k\ \mathbb{E}^r_{xy^k}\left[\int_0^{T_1} e^{-\beta t} \int_A c\left(x,\Phi^r(y^k,t),a\right)\ r_t(da)\ dt\right]\end{aligned}$$

As $T_1 = \tilde{S}_1$ and applying the density of $\tilde{S}_1$ we get

$$= \sum_{k=1}^q \rho_k \int_0^\infty f^r_{\tilde{S}_1}(s \mid y^k) \int_0^s e^{-\beta t} \int_A c\left(x,\Phi^r(y^k,t),a\right)\ r_t(da)\ dt\ ds.$$

Now applying Fubini, this becomes

$$= \sum_{k=1}^{q} \rho_k \int_0^\infty \int_t^\infty f^r_{\tilde{S}_1}(s \mid y^k)\ e^{-\beta t} \int_A c\left(x, \Phi^r(y^k,t), a\right)\ r_t(da)\ ds\ dt.$$

Rearranging the terms and using that $\int_t^\infty f^r_{\tilde{S}_1}(s \mid y^k)\ ds = \mathbb{P}^r_{xy^k}(\tilde{S}_1 > t) = 1 - \mathbb{P}^r_{xy^k}(\tilde{S}_1 \leq t) = e^{-\Lambda^r(y^k,t)}$, we get

$$= \sum_{k=1}^{q} \rho_k \int_0^\infty\ e^{-\beta t} \int_A c\left(x, \Phi^r(y^k,t), a\right)\ r_t(da)\ e^{-\Lambda^r(y^k,t)}\ dt.$$

Appying now the definition of $\eta^r(y^k,t)$ the result follows. □

The next lemma will be crucial for the proof of the lower semi-continuity of g'. We actually get a continuous dependence of Λ^r on r whenever this is true for the controlled drift Φ^r.

Lemma 3.7. *Under Assumption 3.1, the mapping $r \mapsto \Lambda^r(y^k,t)$ is continuous for all $y^k \in E_Y^0$ and $t \geq 0$.*

Proof. Let $y^k \in E_Y^0$ and $t \geq 0$. Further, let (r^n) a sequence in $\mathcal{R}$ with $r^n \to r \in \mathcal{R}$ for $n \to \infty$. If now $r \mapsto \Phi^r(y_k,t)$ is continuous, then $\Phi^{r^n}(y_k,t) \to \Phi^r(y_k,t)$ for $n \to \infty$.

By definition of Λ^r, we then get:

$$\begin{aligned}&\left|\Lambda^{r^n}(y^k,t) - \Lambda^r(y^k,t)\right| \\ &\qquad = \left|\int_0^t \int_A \lambda^A\left(\Phi^{r^n}(y^k,s),a\right)\ r^n_s(da)\ ds - \int_0^t \int_A \lambda^A\left(\Phi^r(y^k,s),a\right)\ r_s(da)\ ds\right|.\end{aligned}$$

By adding zero and re-grouping of the terms, this expression satisfies

$$\begin{aligned}&\leq \left|\int_0^t \int_A \left\{\lambda^A\left(\Phi^{r^n}(y^k,s),a\right) - \lambda^A\left(\Phi^r(y^k,s),a\right)\right\} r^n_s(da)\ ds\right| \\ &\qquad + \left|\int_0^t \int_A \lambda^A\left(\Phi^r(y^k,s),a\right)\ r^n_s(da)\ ds - \int_0^t \int_A \lambda^A\left(\Phi^r(y^k,s),a\right)\ r_s(da)\ ds\right|. \quad (3.1)\end{aligned}$$

Looking now at the first summand of the above sum, we find

$$\begin{aligned}&\left|\int_0^t \int_A \left\{\lambda^A\left(\Phi^{r^n}(y^k,s),a\right) - \lambda^A\left(\Phi^r(y^k,s),a\right)\right\} r^n_s(da)\ ds\right| \\ &\qquad \leq \int_0^t \left\|\lambda^A(\Phi^{r^n}(y^k,s),\cdot) - \lambda^A(\Phi^r(y^k,s),\cdot)\right\|_\infty\ ds \overset{(n\to\infty)}{\longrightarrow} 0.\end{aligned}$$

The $\|\cdot\|_\infty$ norm is well defined because of the boundedness of λ^A (see Definition 1.32) and thus the integral exists. As λ^A is continuous on $E_Y \times A$ (see Definition 1.32), as E_Y is separable and metrizable and as A is a compact metric space, Lemma B.5 applies on λ^A. Further, as $r \mapsto \Phi^r(y^k,s)$ is continuous we have $\Phi^{r^n}(y^k,s) \to \Phi^r(y^k,s)$ and thus,

$$\left\|\lambda^A(\Phi^{r^n}(y^k,s),\cdot) - \lambda^A(\Phi^r(y^k,s),\cdot)\right\|_\infty \overset{(n\to\infty)}{\longrightarrow} 0.$$

By the boundedness of λ^A, dominated convergence leads to the convergence of the integral towards zero.

Now, looking at the second summand in (3.1), we find

$$\int_0^t \int_A \lambda^A \left(\Phi^r(y^k,s),a\right)\ r_s^n(da)\ ds - \int_0^t \int_A \lambda^A \left(\Phi^r(y^k,s),a\right)\ r_s(da)\ ds \overset{(n\to\infty)}{\longrightarrow} 0. \quad (3.2)$$

Actually, by continuity and boundedness of λ^A, we have

$$\mathbb{R}^+ \ni s \mapsto \mathbb{1}_{[0,t]}(s)\ \lambda^A(\Phi^r(y^k,s),\cdot) \in \mathbf{C}(A).$$

Further, by boundedness of λ^A, we also have

$$\int_0^\infty \left\| \mathbb{1}_{[0,t]}(s)\ \lambda^A(\Phi^r(y^k,s),\cdot)\right\|_\infty\ ds \le \int_0^t \left\|\lambda^A\right\|_\infty\ ds < \infty.$$

We therefore have $s \mapsto \mathbb{1}_{[0,t]}(s)\ \lambda^A(\Phi^r(y^k,s),\cdot) \in \mathbb{X} = L^1(\mathbb{R}^+,\mathbf{C}(A))$ and by Lemma A.21, the convergence (3.2) follows. □

We now turn to the main result of this Section:

Theorem 3.8. *Under Assumption 3.1 and Assumption 3.2, the one-step cost function $(x,\rho,r) \mapsto g'(x,\rho,r)$ of the derived filtered model is lower semi-continuous.*

Proof. Assume $r \mapsto \Phi^r(y^k,t)$ is continuous for all $y^k \in E_Y^0$ and all $t \ge 0$.

We first show, that, if $c \in \mathbf{C}(E_X \times E_Y \times A)$, then g' is continuous. We then show that if c is lower semi-continuous g' is lower semi-continuous because in this case, c can be approximated from below by a sequence $(c_k) \subset \mathbf{C}(E_X \times E_Y \times A)$ and we will see that this then implies that g' can be approximated from below by a sequence of continuous functions (g'_k) as well and thus, g' is lower semi-continuous.

Let $c \in \mathbf{C}(E_X \times E_Y \times A)$ and $((x^n,\rho^n,r^n))_{n\in\mathbb{N}}$ a sequence in $E_X \times \mathbf{P}(E_Y^0) \times \mathcal{R}$ with $(x^n,\rho^n,r^n) \to (x,\rho,r)$ for $n \to \infty$ w.r.t. the product topoloy. Based on the representation of g' of Lemma 3.6 we then get

$$\begin{aligned} |g'(x^n,\rho^n,r^n) - g'(x,\rho,r)| \ \le\ & \sum_{k=1}^q \Big| \rho_k^n \int_0^\infty \eta^{r^n}(y^k,t) \int_A c(x^n,\Phi^{r^n}(y_k,t),a)\ r_t^n(da)\ dt \\ & - \rho_k \int_0^\infty \eta^r(y^k,t) \int_A c(x,\Phi^r(y^k,t),a)\ r_t(da)\ dt \Big|. \end{aligned}$$

We will show that for $k = 1,\dots,q$ we have $\rho_k^n \to \rho_k$ and

$$\int_0^\infty \eta^{r^n}(y^k,t) \int_A c(x^n,\Phi^{r^n}(y^k,t),a)\ r_t^n(da)\ dt \overset{n\to\infty}{\longrightarrow} \int_0^\infty \eta^r(y^k,t) \int_A c(x,\Phi^r(y^k,t),a)\ r_t(da)\ dt.$$

The continuity of g' then follows.

To show the first part, let $k \in \{1,\dots,q\}$ and note, that, E_Y^0 is a finite discrete space, thus, the function $E_Y^0 \ni y^i \mapsto f(y^i) = \mathbb{1}_{i=k}$ is of type $\mathbf{C}(E_Y^0)$. Now, $(x^n,\rho^n,r^n) \to (x,\rho,r)$ implies $\rho^n \to \rho$ and this implies $\rho_k^n \to \rho_k$ according to Lemma A.4 applied for the above defined f.

To show now the second part, let $k \in \{1,\dots,q\}$ and consider

$$\begin{aligned} & \Big| \int_0^\infty \eta^{r^n}(y^k,t) \int_A c(x^n,\Phi^{r^n}(y^k,t),a)\ r_t^n(da)\ dt \\ & \qquad - \int_0^\infty \eta^r(y^k,t) \int_A c(x,\Phi^r(y^k,t),a)\ r_t(da)\ dt \Big| \\ & \le \Big| \int_0^\infty \left(\eta^{r_n}(y^k,t) - \eta^r(y^k,t)\right) \int_A c(x^n,\Phi^{r^n}(y^k,t),a)\ r_t^n(da)\ dt \Big| \\ & + \Big| \int_0^\infty \eta^r(y^k,t) \cdot \Big\{ \int_A c(x^n,\Phi^{r^n}(y^k,t),a)\ r_t^n(da) - \int_A c(x,\Phi^r(y^k,t),a)\ r_t(da) \Big\}\ dt \Big|. \quad (3.3) \end{aligned}$$

Now, as c is bounded (remember, in this first step we assume $c \in \mathbf{C}(E_X \times E_Y \times A)$), the first summand satisfies

$$\left| \int_0^\infty \left(\eta^{r_n}(y^k,t) - \eta^r(y^k,t) \right) \int_A c(x^n, \Phi^{r^n}(y^k,t),a)\ r_t^n(da)\ dt \right| \\ \leq \|c\|_\infty \int_0^\infty \left| \eta^{r_n}(y^k,t) - \eta^r(y^k,t) \right|\ dt \overset{n\to\infty}{\longrightarrow} 0.$$

The convergence follows from dominated convergence where $\eta^{r_n}(y^k,t) - \eta^r(y^k,t)$ is dominated by $2 \cdot e^{-\beta t}$ and

$$\eta^{r_n}(y^k,t) = \exp(-\beta t - \Lambda^{r^n}(y^k,t)) \overset{n\to\infty}{\longrightarrow} \exp(-\beta t - \Lambda^r(y^k,t)) = \eta^r(y^k,t)$$

because of Lemma 3.7.

The second summand of (3.3) can be dominated by adding zero to obtain

$$\left| \int_0^\infty \eta^r(y^k,t) \cdot \left\{ \int_A c(x^n, \Phi^{r^n}(y^k,t),a)\ r_t^n(da) - \int_A c(x, \Phi^r(y^k,t),a)\ r_t(da) \right\}\ dt \right| \leq A + B$$

where we define

$$A := \int_0^\infty \eta^r(y^k,t) \int_A \left| c(x^n, \Phi^{r^n}(y^k,t),a) - c(x, \Phi^r(y^k,t),a) \right|\ r_t^n(da)\ dt$$

and

$$B := \left| \int_0^\infty \eta^r(y^k,t) \int_A c(x, \Phi^r(y^k,t),a)\ r_t^n(da)\ dt \right. \\ \left. - \int_0^\infty \eta^r(y^k,t) \int_A c(x, \Phi^r(y^k,t),a)\ r_t(da)\ dt \right|.$$

We will show, that both, A and B converge to zero. First, as c is continuous and bounded, as E_X and E_Y are separable and metrizable, as A is compact and metric and as $\Phi^{r^n}(y_k,t) \to \Phi^r(y_k,t)$ by Assumption 3.1, Lemma B.5 implies

$$\left\| c(x^n, \Phi^{r^n}(y^k,t),\cdot) - c(x, \Phi^r(y^k,t),\cdot) \right\|_\infty \to 0\ (n \to \infty).$$

We thus get

$$A \leq \int_0^\infty \eta^r(y^k,t) \left\| c(x^n, \Phi^{r^n}(y^k,t),\cdot) - c(x, \Phi^r(y^k,t),\cdot) \right\|_\infty\ dt \to 0\ (n \to \infty),$$

by dominated convergence applied for dominating function $t \mapsto 2 \cdot \|c\|_\infty\, \eta^r(y^k,t)$.

For B we get convergence to zero from Lemma A.21 as

$$t \mapsto \eta^r(y^k,t) \cdot c(x, \Phi^r(y^k,t),\cdot)$$

lies in $\mathbb{X} = L^1(\mathbb{R}^+, \mathbf{C}(A))$ because c is continuous and bounded and because of

$$\int_0^\infty \left\| \eta^r(y^k,t) \cdot c(x, \Phi^r(y^k,t),\cdot) \right\|_\infty\ dt \leq \|c\|_\infty \int_0^\infty \eta^r(y^k,t)\ dt \leq \|c\|_\infty \int_0^\infty e^{-\beta t}\ dt < \infty.$$

We showed that g' is continuous if c is continuous and bounded and we also get that g' is bounded in this case as

$$\begin{aligned} |g'(x,\rho,r)| &\leq \sum_{k=1}^q \left| \rho_k \int_0^\infty \eta^r(y^k,t) \int_A c(x, \Phi^r(y^k,t),a)\ r_t(da)\ dt \right| \\ &\leq \sum_{k=1}^q \|c\|_\infty \int_0^\infty \left| \eta^r(y^k,t) \right|\ dt \\ &\leq q \cdot \|c\|_\infty \cdot \int_0^\infty e^{-\beta t} dt \\ &\leq \frac{q}{\beta} \cdot \|c\|_\infty < \infty. \end{aligned}$$

Now, let c lower semi-continuous (and non-negative, what we always assume for c), then, there is a sequence $(c_m) \subset \mathbf{C}(E_X \times \mathbb{P}(E_Y^0) \times A)$ with $c_m \uparrow c$ for $m \to \infty$ (see [11], Lemma 7.14).

We then define

$$g'_m(x, \rho, r) := \sum_{k=1}^{q} \rho_k \int_0^\infty \eta^r(y^k, t) \int_A c_m(x, \Phi^r(y^k, t), a)\ r_t(da)\ dt.$$

We have shown before, that this is a continuous and bounded function. Furthermore, we find

$$g'_m(x, \rho, r) \leq g'(x, \rho, r)$$

as $c_m(x, \Phi^r(y^k, t), a) \leq c(x, \Phi^r(y^k, t), a)$ and by monotonicity of the integral. Finally, monotone convergence gives $g'_m \uparrow g'$ for $m \to \infty$. Thus, g' can be approximated from below by a sequence of continuous and bounded functions and by [11], Lemma 7.14, g' is lower semi-continuous. □

3.2 Weak continuity of the transition kernel

3.2.1 Models with controlled drift but uncontrolled jump rate and state transition

Having formulated two assumptions for the PO-PDMP sufficient for the lower semi-continuity of the one step cost function of the derived filtered model to hold, we will now turn to the investigation of the weak continuity of the transition kernel $\tilde{q}'$ of Assumption 2.40 (C). Again, the task is to find sufficient conditions for the PO-PDMP in order to find this latter assumption satisfied by the derived filtered model.

We will show that for models with controlled drift but uncontrolled jump intensity λ and uncontrolled transition kernel Q for the unobservable state, Assumption 2.40 (C) is satisfied. We start by formalizing this assumption on the PO-PDMP:

Assumption 3.9 (Controlled drift only). *We assume the jump intensity as well as the transition kernel for the unobservable state of the PO-PDMP to be uncontrolled, i.e. there exists $a_0 \in A$ such that for all $a \in A, y^j \in E_Y^0, y \in E_Y$:*

$$\lambda^A(y, a) = \lambda^A(y, a_0) \qquad \text{and} \qquad Q^A(y, a; \{y^j\}) = Q^A(y, a_0; \{y^j\}).$$

Remark 3.10. *Note that instead of assuming the existence of $a_0 \in A$ with $\lambda^A(y, a) = \lambda^A(y, a_0)$ for all $a \in A, y \in E_Y$ one could also assume $\lambda^A(y, a) = \lambda(y)$ for all $a \in A, y \in E_Y$, thus referring to the jump intensity of the uncontrolled PO-PDMP. As λ was only assumed to be a measurable mapping, we then need to add the assumption of λ being continuous and bounded as we assume for λ^A. Analogously, we could formulate the assumption for Q and Q^A.*

Remember that the transition law Q'_{SXM} of the derived filtered process does not depend on the last observed inter-jump time s, neither on the last observed noisy measurement x, i.e., for suitable borel sets B_X and B_M

$$Q'_{SXM}([0, t] \times B_X \times B_M \mid s, x, \rho, r) = Q'_{SXM}([0, t] \times B_X \times B_M \mid \rho, r).$$

This property shall be clear when looking at Definition 2.32 (dependence of q'_{SXM} on s and x is only via $\tilde{q}_{SXY}$) and at Definition 2.6 ($\tilde{q}_{SXY}$ does not depend on s and x).

In order to prove now the weak continuity of the transition kernel

$$\tilde{q'}(B \mid \rho, r) := \frac{\int_{E_{SXM}} e^{-\beta s'} \mathbb{1}_B(s', x', \rho')\ q'_{SXM}(ds', dx', d\rho' \mid \rho, r)}{\int_{E_{SXM}} e^{-\beta s'}\ q'_{SXM}(ds', dx', d\rho' \mid \rho, r)}, \tag{3.4}$$

we show that for arbitrary $f \in \mathbf{C}(\mathbb{R}^+ \times E_X \times \mathbf{P}(E_Y^0))$ and for $(\rho_n, r_n) \overset{(n\to\infty)}{\longrightarrow} (\rho_\infty, r_\infty) \in \mathbf{P}(E_Y^0) \times \mathcal{R}$, the following convergence of integrals holds:

$$\int f(s,x,\rho) e^{-\beta s}\, Q'_{SXM}(ds \otimes dx \otimes d\rho \mid \rho_n, r_n) \overset{(n\to\infty)}{\to} \int f(s,x,\rho) e^{-\beta s}\, Q'_{SXM}(ds \otimes dx \otimes d\rho \mid \rho_\infty, r_\infty).$$

Once we proved this convergence of integrals, the convergence of $\int f(s,x,\rho)\ \tilde{q}'(ds, dx, d\rho \mid \rho_n, r_n)$ follows for $f \in \mathbf{C}(\mathbb{R}^+ \times E_X \times \mathbf{P}(E_Y^0))$ as $\mathbb{1}_{E_{SXM}}(s,x,\rho)$ is of type $\mathbf{C}(\mathbb{R}^+ \times E_X \times \mathbf{P}(E_Y^0))$ and thus, also the denominator of (3.4) converges. By Proposition A.7, the weak continuity of $\tilde{q}'$ then follows.

The main result of this section is thus the following:

Proposition 3.11. *Let $f \in \mathbf{C}(\mathbb{R}^+ \times E_X \times \mathbf{P}(E_Y^0))$ and $(\rho_n, r_n) \overset{(n\to\infty)}{\longrightarrow} (\rho_\infty, r_\infty) \in \mathbf{P}(E_Y^0) \times \mathcal{R}$, then the following convergence of integrals holds under Assumption 3.1 and Assumption 3.9:*

$$\int f(s,x,\rho) e^{-\beta s}\, Q'_{SXM}(ds \otimes dx \otimes d\rho \mid \rho_n, r_n) \overset{(n\to\infty)}{\to} \int f(s,x,\rho) e^{-\beta s}\, Q'_{SXM}(ds \otimes dx \otimes d\rho \mid \rho_\infty, r_\infty).$$

Proof. We start by applying the definition of q'_{SXM} (see Definition 2.32) as well as the definition of $\tilde{q}_{SXY}$ (see Definition 2.6) and obtain:

$$\begin{aligned}
&\int f(s,x,\rho)\ e^{-\beta s}\ Q'_{SXM}(ds \otimes dx \otimes d\rho \mid \rho_n, r_n) \\
&\overset{\text{def.}q'}{=} \sum_{i=1}^{q} \rho_n^i \int_{[0,\infty)\times E_X} f(s,x,\chi(\rho_n,s,x,r_n))\ e^{-\beta s} \cdots \\
&\quad \cdots\ \tilde{Q}_{SXY}(ds \otimes dx \otimes E_Y^0 \mid y^i, r_n) \\
&\overset{\text{def.}\tilde{q}}{=} \sum_{i=1}^{q} \rho_n^i \int_0^\infty e^{-\beta s}\ \exp(-\Lambda^{r_n}(y^i,s)) \int_A \sum_{j=1}^{q} Q^A(\Phi^{r_n}(y^i,s),a;\{y^j\})\ \lambda^A(\Phi^{r_n}(y^i,s),a) \cdots \\
&\quad \cdots \int_{E_X} f(s,x,\chi(\rho_n,s,x,r_n)) f_\epsilon(x-\psi(y^j))\ \nu(dx)\ r_n(s;da)\ ds \\
&\overset{\text{Ass.3.9}}{=} \sum_{i=1}^{q} \rho_n^i \int_0^\infty e^{-\beta s}\ \exp(-\Lambda^{r_n}(y^i,s))\ \lambda^A(\Phi^{r_n}(y^i,s),a_0)\ \sum_{j=1}^{q} Q^A(\Phi^{r_n}(y^i,s),a_0;\{y^j\}) \cdots \\
&\quad \cdots \int_{E_X} f(s,x,\chi(\rho_n,s,x,r_n)) f_\epsilon(x-\psi(y^j))\ \nu(dx)\ ds
\end{aligned}$$

Now, as by assumption $\rho_n \to \rho_\infty$, we have $\rho_n^i \to \rho_\infty^i$. Remains to show that for all $i = 1, \ldots, q$, the integral converges to

$$\begin{aligned}
\int_0^\infty e^{-\beta s}\ \exp(-\Lambda^{r_\infty}(y^i,s))\ \lambda^A(\Phi^{r_\infty}(y^i,s),a_0)\ \sum_{j=1}^{q} Q^A(\Phi^{r_\infty}(y^i,s),a_0;\{y^j\}) \cdots \\
\cdots \int_{E_X} f(s,x,\chi(\rho_\infty,s,x,r_\infty)) f_\epsilon(x-\psi(y^j))\ \nu(dx)\ ds.
\end{aligned}$$

This convergence of integrals follows by dominated convergence as $\exp, \lambda^A, Q^A$ and f are bounded by definition and thus, $e^{-\beta s} \cdot K$ is an integrable upper bound with K selected accordingly. Further, we have convergence of $\Lambda^{r_n} \to \Lambda^{r\infty}$ by Lemma 3.7, convergence of $\Phi^{r_n}(y^i, s) \to \Phi^{r\infty}(y^i, s)$ by Assumption 3.1 and λ^A as well as Q^A are (weakly) continuous in the argument y (see definitions). Remains to show that $\chi(\rho_n, s, x, r_n) \to \chi(\rho_\infty, s, x, r_\infty)$, then with $f \in \mathbf{C}(\mathbb{R}^+ \times E_X \times \mathbf{P}(E_Y^0))$ the result follows with DOM.

Looking at Definition 2.30 of the filter equation and applying Assumption 3.9, we have

$$\chi_i^j(\rho_n, s, x, r_n) := \rho_n^i \exp\left(-\Lambda^{r_n}(y^i, s)\right) \lambda^A\left(\Phi^{r_n}(y^i, s), a_0\right) Q^A\left(\Phi^{r_n}(y^i, s), a_0; \{y^j\}\right) f_\epsilon(x - \psi(y^j))$$

and with the same continuity arguments as above, this converges to

$$\rho_\infty^i \exp\left(-\Lambda^{r\infty}(y^i, s)\right) \lambda^A\left(\Phi^{r\infty}(y^i, s), a_0\right) Q^A\left(\Phi^{r\infty}(y^i, s), a_0; \{y^j\}\right) f_\epsilon(x - \psi(y^j)) = \chi_i^j(\rho_\infty, s, x, r_\infty) \quad \square$$

To conclude this Section, we like to remark the following: While in the case of completely observable PDMPs, Assumptions 3.1 and 3.2 are enough to guarantee the existence of optimal policies (see [34]), here, in the case of a PO-PDMP, we need additional assumptions to guarantee existence of optimal policies. So far, with Assumption 3.9, we presented one sufficient condition for existence of optimal policies. However, the latter Assumption is a serious restriction of the set of admissible models. For a huge variety of applications, however, this assumptions is satisfied, as very often, the jump intensity λ and the state transition kernel Q of the unobservable state are external factors that the responsible agent cannot control. Think of queues where you can control the service speed of the server but not the intensity of arrivals and not the amount of arrivals in bulk arrival models. See also Chapter 6 for more examples where Assumption 3.9 applies.

The reason why, for instance, we have to assume λ and Q to be uncontrolled lies in a missing continuity property of the filter equation χ as we will now outline in the next Section.

3.2.2 The continuity issue of the filter

One could ask why the restriction to models with „controlled drift only“ was necessary in the last Section. The answer shall be briefly illustrated here in this Section. It basically comes back to the fact that the filter χ is not continuous in the variable r in general. With χ not continuous in the variable r, Proposition 3.11 cannot be shown in the general case where λ and Q can be controlled as well. Actually, the proof of Proposition 3.11 fails when one cannot show that $f(s, x, \chi(\rho_n, s, x, r_n)) \to f(s, x, \chi(\rho_\infty, s, x, r_\infty))$ for $n \to \infty$.

Remember the definition of the filter equation for χ as of Definition 2.30:

$$\chi_i^j(\mu, s, x, r) := \mu^i \exp\left(-\Lambda^r(y^i, s)\right) \int_A \lambda^A\left(\Phi^r(y^i, s), a\right) Q^A\left(\Phi^r(y^i, s), a; \{y^j\}\right) r_s(da) f_\epsilon(x - \psi(y^j)).$$

The problem is to prove continuity of the mapping

$$r \mapsto \int_A \lambda^A\left(\Phi^r(y^i, s), a\right) Q^A\left(\Phi^r(y^i, s), a; \{y^j\}\right) r_s(da).$$

A standard approach would probably be: Let $r^n \to r$ $(n \to \infty)$ in $\mathcal{R}$, then:

$$\left| \int_A \lambda^A \left(\Phi^{r^n}(y^i, s), a \right) Q^A \left(\Phi^{r^n}(y^i, s), a; \{y^j\} \right) \; r_s^n(da) \right.$$
$$\left. - \int_A \lambda^A \left(\Phi^{r}(y^i, s), a \right) Q^A \left(\Phi^{r}(y^i, s), a; \{y^j\} \right) \; r_s(da) \right| \leq (I_A) + (I_B),$$

where we define

$$(I_A) := \left| \int_A \lambda^A \left(\Phi^{r^n}(y^i, s), a \right) Q^A \left(\Phi^{r^n}(y^i, s), a; \{y^j\} \right) \; r_s^n(da) \right.$$
$$\left. - \int_A \lambda^A \left(\Phi^{r}(y^i, s), a \right) Q^A \left(\Phi^{r}(y^i, s), a; \{y^j\} \right) \; r_s^n(da) \right|$$

and

$$(I_B) := \left| \int_A \lambda^A \left(\Phi^{r}(y^i, s), a \right) Q^A \left(\Phi^{r}(y^i, s), a; \{y^j\} \right) \; r_s^n(da) \right.$$
$$\left. - \int_A \lambda^A \left(\Phi^{r}(y^i, s), a \right) Q^A \left(\Phi^{r}(y^i, s), a; \{y^j\} \right) \; r_s(da) \right|$$

As both, λ^A and Q^A are bounded, we get

$$(I_A) \leq \left\| \lambda^A \left(\Phi^{r^n}(y^i, s), \cdot \right) Q^A \left(\Phi^{r^n}(y^i, s), \cdot; \{y^j\} \right) - \lambda^A \left(\Phi^{r}(y^i, s), \cdot \right) Q^A \left(\Phi^{r}(y^i, s), \cdot; \{y^j\} \right) \right\|_\infty$$

and as the function

$$\kappa : \mathcal{R} \times A \to \mathbb{R}, (r, a) \mapsto \kappa(r, a) := \lambda^A \left(\Phi^{r}(y^i, s), a \right) Q^A \left(\Phi^{r}(y^i, s), a; \{y^j\} \right)$$

is continuous in (r, a), as the space $\mathcal{R}$ is separable and metrizable and as A is a compact metric space, we can apply Lemma B.5 to find that

$$(I_A) \overset{(n \to \infty)}{\longrightarrow} 0.$$

Expression (I_B) unfortunately, does not converge to zero for $n \to \infty$ in general. All we know by the characterization of convergence in $\mathcal{R}$ (see Lemma A.21) is that, e.g., for all intervals of type $[\alpha, \beta] \subset [0, \infty)$, we get

$$\int_\alpha^\beta \int_A \lambda^A \left(\Phi^{r}(y^i, s), a \right) Q^A \left(\Phi^{r}(y^i, s), a; \{y^j\} \right) \; r_s^n(da) \; ds$$
$$\overset{(n \to \infty)}{\longrightarrow} \int_\alpha^\beta \int_A \lambda^A \left(\Phi^{r}(y^i, s), a \right) Q^A \left(\Phi^{r}(y^i, s), a; \{y^j\} \right) \; r_s(da) \; ds.$$

This, unfortunately, is not enough to deduct the convergence (for almost all s) of

$$\int_A \lambda^A \left(\Phi^{r}(y^i, s), a \right) Q^A \left(\Phi^{r}(y^i, s), a; \{y^j\} \right) \; r_s^n(da)$$
$$\overset{(n \to \infty)}{\longrightarrow} \int_A \lambda^A \left(\Phi^{r}(y^i, s), a \right) Q^A \left(\Phi^{r}(y^i, s), a; \{y^j\} \right) \; r_s(da).$$

The following example shows that for a sequence f_n of measurable functions $f_n : \mathbb{R}^+ \to \mathbb{R}^+$ with $\int_\alpha^\beta f_n(s) \; ds \overset{(n \to \infty)}{\longrightarrow} \int_\alpha^\beta f(s) \; ds$ for all $[\alpha, \beta] \subset [0, \infty)$ one cannot derive the convergence $f_n(s) \overset{(n \to \infty)}{\longrightarrow} f(s)$ pointwise for almost all $s \in \mathbb{R}^+$.

Take for example

$$f_1(s) := \sum_{k=0}^{\infty} 2 \cdot \mathbb{1}_{[k;k+\frac{1}{2})}(s) + \mathbb{1}_{[k+\frac{1}{2};k+1)}(s),$$

that is the function which is constant 2 on the first half of every integer interval and that is 1 on the second half of every integer interval. By defining $k_{1,0} := 0$ and $k_{1,l} := k_{1,l-1} + 1$ for every $l \geq 1$, we simply can write f_1 as

$$f_1(s) := \sum_{l=0}^{\infty} 2 \cdot \mathbb{1}_{[k_{1,l};k_{1,l}+\frac{1}{2})}(s) + \mathbb{1}_{[k_{1,l}+\frac{1}{2};k_{1,l+1})}(s).$$

We can now generalize this procedure for $n \geq 2$: We split into two halfs every interval where f_{n-1} is constant and again setting the first half of such a newly defined interval on constant function value 2 and the second half on constant function value 1. Precisely, this means, we define the interval bounds for $n \geq 1$ by $k_{n,0} := 0$ and $k_{n,l} := k_{n,l-1} + 2 \cdot \left(\frac{1}{2}\right)^n$ for $l \geq 1$. With these new interval bounds we can define, for all $n \geq 1$ (and this is consistent to the above definition of f_1):

$$f_n(s) := \sum_{l=0}^{\infty} 2 \cdot \mathbb{1}_{[k_{n,l};k_{n,l}+(\frac{1}{2})^n)}(s) + \mathbb{1}_{[k_{n,l}+(\frac{1}{2})^n;k_{n,l+1})}(s).$$

For this sequence of functions it holds:

- f_n is measurable for all $n \geq 1$
- $\int_\alpha^{\alpha+1} f_n(s)\ ds = \frac{3}{2}$ for all $\alpha \in \mathbb{N}$ (function value 2 and 1 is equally distributed on every integer interval of the form $[\alpha, \alpha + 1)$.
- $\int_\alpha^\beta f_n(s)\ ds \overset{(n\to\infty)}{\longrightarrow} \frac{3}{2} \cdot (\beta - \alpha)$ for all $[\alpha, \beta) \subset \mathbb{R}^+$

But now consider the sequence g_n of functions defined analogously to f_n with only difference that the two possible function values of g_n are not 2 and 1 as for f_n but are 3 and 0. Then, still it holds:

- g_n is measurable for all $n \geq 1$
- $\int_\alpha^{\alpha+1} g_n(s)\ ds = \frac{3}{2}$ for all $\alpha \in \mathbb{N}$ (function value 3 and 0 is equally distributed on every integer interval of the form $[\alpha, \alpha + 1)$.
- $\int_\alpha^\beta g_n(s)\ ds \overset{(n\to\infty)}{\longrightarrow} \frac{3}{2}(\beta - \alpha)$ for all $[\alpha, \beta) \subset \mathbb{R}^+$

Obviously, the two sequences of functions f_n and g_n have the same limit property for integrals taken over arbitrary subintervals of $\mathbb{R}^+$ but in no way one can say that for almost all $s \in \mathbb{R}^+$, $f_n(s)$ and $g_n(s)$ have the same pointwise limit for $(n \to \infty)$.

For further examples of sequences of relaxed controls that do converge in the Young topology but not in „any of the usual senses“, see also [47], pages 87ff and equation (6.5) with corresponding remark.

The discontinuity of χ in the variable r is a serious problem for the existence of optimal policies for the cost optimal control of a PO-PDMP model as presented in Chapter 1. This is an example for a PO-MDP where (weak) continuity of the filter cannot be shown

and thus, weak continuity of the transition kernel for the states of the derived filtered completely observable MDP cannot be derived.

However, having reformulated our initial optimization problem into an equivalent optimization problem for a PO-MDP in Section 2.1, we can apply recent results from the domain of research on PO-MDPs in order to further investigate the question of sufficient conditions for the existence of optimal policies. This is what we will do in the next Chapter.

Chapter 4

Models with unobservable inter-jump time

At the end of the last Chapter, we pointed out that the filter χ is not continuous in its argument $r \in \mathcal{R}$. As a consequence, we had to restrict our investigations to models with uncontrolled intensity λ and uncontrolled transition law Q. For these models, we could show that assumption (C) holds even though the filter χ was not continuous.

In this Chapter, we will present another class of models where we will obtain existence of optimal policies while allowing to control the intensity λ and the transition law Q. The principal idea is to apply a different filter that is sufficiently continuous in order to get assumption (C) satisfied. In view of Lemma A.21 (characterization of convergence in Young topology), this filter should not depend on the inter-jump time as we will explain in detail in Section 4.1 below. Two classes of models could bring up such a filter: (i) models with noisy measurement of the inter-jump time and (ii) models with unobservable inter-jump time. In this Chapter, we will focus on the latter class of models and leverage a recent result of Feinberg [33] from 2016. Feinberg's result can be seen as an important contribution to the general theory of partially observable Markov Decision Processes (PO-MDP):

In principle, the approach for solving a total discounted cost problem for a PO-MDP is clear. One needs to reduce the PO-MDP to a CO-MDP by the help of a filter as we did in Section 2.2. A large variety of literature exists on reducing PO-MDPs to CO-MDPs. Sawarigi and Yoshikawa [56] worked on filtering PO-MDPs with countable state spaces. Bertsekas and Shreve [11] as well as Yushkevich [63], Rhenius [54] and Hérnandez-Lerma [38] worked on the same problem for Borel state spaces. The topic of existence of optimal policies for CO-MDPs with total (discounted) expected cost problems was investigated by, e.g., Sondik [59] for finite state and action spaces or, e.g., by Hérnandez-Lerma [38] and Hérnandez-Lerma and Romera [39] for situations where the filter is weakly continuous.

All these works, however, do not deal with the problem we faced in the last Section, namely, having a filter that is not (weakly) continuous. However, this situation can be faced in concrete applications as we saw. Feinberg's result now helps to „weaken“ this problem: While Hérnandez-Lerma still needs weakly continuous filters in [38] to prove existence of optimal policies for PO-MDPs, Feinberg et al. achieve to show existence of optimal policies for PO-MDPs in situations where the filter is not weakly continuous.

Actually, Feinberg achieves to show weak continuity of the transition kernel of the CO-MDP already if the filter only converges (weakly) for a suitable subsequence of a converging sequence of arguments of the filter. With other words, and in the notations of this thesis, Feinberg does not need $\chi(\cdots, r_n)$ to converge to $\chi(\cdots, r_\infty)$ for $r_n \to r_\infty$ but convergence of $\chi(\cdots, r_{n_k})$ for a suitable subsequence (r_{n_k}) of (r_n) is enough.

Feinberg et al. also provide a sufficient condition on the „transition kernel for the unobservable state“ of the PO-MDP and on the „observation kernel[1]“ of the observable state of the PO-MDP. If this condition is satisfied, a filter satisfying their convergence condition exists and thus, existence of optimal policies can be shown.

However, the results of Feinberg cannot be applied directly to our optimization problem for the pseudo-embedded process: The observation kernel for the pseudo-embedded process does not satisfy the sufficient condition of Feinberg. This is mainly due to the perfect observation of the inter-jump time. For models with unobservable inter-jump time, however, Feinberg's results and ideas can be applied and transferred to the corresponding pseudo-embedded process.

The outline of this Chapter is thus the following: We first give a motivation why models with unobservable or not perfectly observed inter-jump time are good candidates for getting filters with sufficient continuity properties (Section 4.1). As Feinberg only considers models where the cost function is only depending on the unobservable component of a state, we discuss the restriction to models with cost function c not depending on the noisy measurement x of a post-jump state y in Section 4.2. The consequences arising from the fact that the inter-jump time cannot be observed in this new model are discussed in Section 4.3. Based on Feinberg's approach, we then show existence of a filter with suitable properties for us in Section 4.4. Finally, we can show that the model with unobservable inter-jump time and running cost that does not depend on the observable component of a state of the process satisfies assumptions (LSC) and (C). Hence, existence of one-step optimizers can be shown and thus existence of optimal policies follows. This is discussed in Section 4.5.

4.1 Motivation

In Section 3.2.2, we highlighted the issue of the filter χ not being continuous in its argument $r \in \mathcal{R}$. Remember Lemma A.21, where convergence w.r.t. the Young topology is characterized by

$$r^n \overset{n\to\infty}{\longrightarrow} r \quad \Longleftrightarrow \quad \int_0^\infty \int_A \psi(t,a)\; r_t^n(da)\; dt \overset{n\to\infty}{\longrightarrow} \int_0^\infty \int_A \psi(t,a)\; r_t(da)\; dt \quad \forall \psi \in \mathbb{X},$$

thus by convergence of a double integral where the integral w.r.t. $r_t(da)$ appears within an integral w.r.t. dt. Comparing this characterization to the definition of the filter (see Definition 2.30), where

$$\chi_i^j(\mu, s, x, r) := \\ \mu^i \exp\left(-\Lambda^r(y^i, s)\right) \int_A \lambda^A\left(\Phi^r(y^i,s), a\right) Q^A\left(\Phi^r(y^i,s), a; \{y^j\}\right)\; r_s(da) f_\epsilon(x - \psi(y^j)),$$

we find an integral w.r.t. $r_s(da)$ that is not contained in an integral w.r.t. ds. This is the principal reason for χ not being continuous in the argument $r \in \mathcal{R}$.

In order to get a filter that is continuous in its argument r, one should make appear, in a suitable way, an integral w.r.t. ds in the definition of the filter. This, however, would mean that the inter-jump time s cannot be observed correctly if the filter is „averaging“ over this inter-jump time by he help of some integral w.r.t. ds. Situations where this could apply are, e.g.:

[1] conditional distribution of the observable state given the unobservable state and the last executed control action

(i) Noisy measurements of the inter-jump time

(ii) Unobservable inter-jump times.

In the first situation, a noisy measurement of the inter-jump time could be modeled analogously to the noisy measurement of the post-jump state as we did in our model: One could introduce a denisty $f_{\tilde{S}}$ of measurement noise for the inter-jump time. This would make appear an integral of type

$$\int_0^\infty f_{\tilde{S}}(s' - s)ds'$$

in the filter. Our model with perfect observation of the inter-jump time could then be approximated by models with noisy measurement of the inter-jump time where the support of the noise density is more and more concentrated around zero.

In the second situation, where the inter-jump time is unobservable, the filter for the post-jump state shall contain an integral $\int_0^\infty \dots ds$. Actually, given the observation x and control r the conditional probability of having a post-jump state y^j is the joint conditional distribution of $(Y = y^j, s \in \mathbb{R}^+)$ given x and r.

In view of possible applications for models of these two situations, situation (i) seems a very realistic setting as one could always assume a whatever „small“, i.e. concentrated around zero, measurement noise for the inter-jump time observation. One has to measure time and measurements are always noisy in real life. In situation (ii), only the inter-jump time, i.e. the time since the last jump up to the current jump is assumed to be unobservable. This does not mean that the point in time when a jump occurs is not observable. This might be - and also shall be - the case as an agent has to know when to apply the n-th decision rule of a policy. The agent simply has no watch to measure time elapsed between two points in time.

For the rest of this Chapter, we will focus on situation (ii) of models with unobservable inter-jump time.

4.2 Restriction to running cost only depending on unobservable state

For the rest of this Chapter we will take the following assumption on the running cost function c:

Assumption 4.1. *We assume the running cost function c to be a lower semi-continuous (w.r.t the product topology) function $c : E_Y \times A \to \mathbb{R}^+$.*

In view of Assumption 3.2, this is a restriction to running cost functions that are not depending on the observable component $x \in E_X$ of a state of the PO-PDMP. As we have developed the theory for running cost functions $c : E_X \times E_Y \times A \to \mathbb{R}^+$, the full theory still holds for this restricted set of admissible running cost functions.

In particular, we like to emphasize that the definition of the one period cost function g (compare Definition 2.1) then becomes:

$$g(y, r) := \mathbb{E}_y^r \left[\int_0^{T_1} e^{-\beta t} \int_A c(\Phi^r(y, t), a)\ r_t(da)\ dt \right]. \tag{4.1}$$

Note that we can write $\mathbb{E}_y^r$ instead of $\mathbb{E}_{x,y}^r$ as of Definition 2.1. Actually, as c does not depend on x and as the density of the distribution of T_1 does not depend on x but only on y (see Lemma 3.4) the probability measure to apply only depends on y and on r.

The one-step cost function g' for the derived filtered model (compare equation (2.24)) then becomes, for $\rho \in \mathbf{P}(E_Y^0)$:

$$g'(\rho, r) := \sum_{i=1}^{q} g\left(y^i, r\right) \rho^i.$$

The equivalence of the initial optimization problem and the optimization problem for the derived pseudo-embedded process $(\tilde{S}_k, \tilde{X}_k, \tilde{Y}_k)_{k\geq 0}$ still holds (see Proposition 2.15) as we proved it for functions c and g depending on x and y. Thus, under restriction to functions c and g no longer depending on x these results are still valid. Note that the value function for the optimization problem for the pseudo-embedded process will still depend on the initial observation $x \in E_X$, i.e.

$$\tilde{J}(x) = \min_{\pi \in \Pi^D} \tilde{J}(x, \pi^D),$$

as we will still have the cost of a policy $\pi^D \in \Pi^D$ depend on x, i.e.

$$\tilde{J}(x, \pi^D) := \mathbb{E}_x^{\pi^D}\left[\sum_{k=0}^{\infty} e^{-\beta \sum_{n=0}^{k} \tilde{S}_n} g(\tilde{Y}_k, \pi_k^D(\tilde{H}_k))\right]$$

(see Definition 2.14). This is because the initial distribution of $\tilde{Y}_0$ depends on the initial observation $x \in E_X$ as this initial distribution is given by $Q_0(x, \cdot)$.

As a summary to this Section we conclude that the whole theory developed in this thesis is also valid for running cost functions c that only depend on the unobservable component of a state instead of being a function of both, observable and unobservable component of a state of the PO-PDMP.

4.3 Consequences of unobservable inter-jump time

Assuming the inter-jump time to be unobservable has consequences on the space of observable histories. This is what we will discuss in this Section.

First, a short comment on what we mean by „unobservable“ inter-jump time. We basically want to develop a filtered model where the filter does not depend on the inter-jump time. Thus, one could say we consider models where the inter-jump time is observable but we do not use this available information as input for the filter. In view of Lemma 2.38, where we explained how a Markov policy for the filtered model can be understood as a history dependent policy, we have to consider the inter-jump time as unobservable as long as the filter does not use this information as input. Actually, in Lemma 2.38, we interpret a component ρ of a state of the filtered process as $\rho = \mu_n(h_n)$, where h_n stands for the observed history up to the n-th jump time. We then argue that $f_n(\cdots, \mu_n(h_n))$ is thus depending on the full observed history up to the n-th jump time. If now, however, the filter is not depending on the inter-jump time, then $\mu_n(h_n)$ is not depending on $(s_0, s_1, \ldots, s_n)$, the history of observed inter-jump times. As long as we do not restrict the observable histories to observations of $(\tilde{X}_0, \tilde{X}_1, \ldots)$ only, we thus cannot reason any longer that a Markov policy for the filtered model can be understood as history dependent policy.

For a model with unobservable inter-jump time we thus define the space of observable histories by recursion as: $\hat{\mathcal{H}}_0 := E_X$ and for $n \geq 1$, we define $\hat{\mathcal{H}}_n := \hat{\mathcal{H}}_{n-1} \times \mathcal{R} \times E_X$.

A history dependent decision rule at stage n is then defined accordingly as a measurable mapping $\hat{\pi}_n^D : \hat{\mathcal{H}}_n \to \mathcal{R}$. With this slight modification, the theory developed in this thesis still holds for the resulting model with this adapted definition of observable histories. Actually, the principal properties of the underlying spaces are not affected and thus, even the correspondence theorem still holds.

4.4 A filter with suitable properties

In this Section, we will show the existence of a filter with properties that will turn out to be sufficient in order to prove existence of optimal policies for the derived filtered problem. The reasoning in this Section is based on Feinberg's publication [33] of 2016. As Feinberg is considering a PO-MDP with two components, one observable and one unobservable, we suggest to start from the pseudo-embedded process $(\tilde{S}_k, \tilde{X}_k, \tilde{Y}_k)_{k\geq 0}$ with transition law $\tilde{q}_{SXY}$ and only consider the process $(\tilde{X}_k, \tilde{Y}_k)_{k\geq 0}$. This is a process with one observable and one unobservable component.

We thus ignore the component $\tilde{S}_k$ of the pseudo-embedded process. As the running cost function c as well as the one period cost function g do not take the inter-jump time as argument of the respective function, ignoring this component of the pseudo-embedded process is no issue for the following theory. The only moment where the concrete realization of the inter-jump time becomes important is for the discounting. However, we can separate, in our reasoning, the discounting from the one step cost function as such.

Throughout this Chapter we will keep Assumption 2.16 on finitely many possible post-jump states, i.e. $\tilde{Y}_k \in E_Y^0$ almost surely for all $k \geq 0$. The next result follows directly from the definition of a pseudo-embedded process and its transition law.

Corollary 4.2. *Under Assumption 2.16, the transition law of the process $(\tilde{X}_k, \tilde{Y}_k)_{k\geq 0}$ is given by*

$$\tilde{Q}_{XY}\left(B_X \times \{y^j\} \mid x, y^i, r\right) = \tilde{Q}_{SXY}\left(\mathbb{R}^+ \times B_X \times \{y^j\} \mid y^i, r\right), \tag{4.2}$$

where $B_X \in \mathcal{B}(E_X), r \in \mathcal{R}$ and $y^i, y^j \in E_Y^0$.

In view of Feinberg's approach, we „separate“ the transition law $\tilde{q}_{XY}$ into

- a transition kernel for the unobservable state, denoted by $P\left(\tilde{Y}_k = y^j \mid \tilde{Y}_{k-1} = y^i, r\right)$ and
- an observation kernel, i.e. a transition kernel for the observable component of a state given the unobservable component and the last executed control action, denoted by $Q^{obs}\left(B_X \mid y^j, r\right)$ for $B_X \in \mathcal{B}(E_X)$.

In our conrete model, the transition kernel for the unobservable state and the observation kernel have the following form:

Lemma 4.3. *For $y^i, y^j \in E_Y^0$ and $r \in \mathcal{R}$, the transition kernel for the unobservable state is given by*

$$P\left(\tilde{Y}_k = y^j \mid \tilde{Y}_{k-1} = y^i, r\right) = \int_0^\infty \exp\left(-\Lambda^r(y^i, s)\right) \int_A \lambda^A\left(\Phi^r(y^i, s), a\right) Q^A\left(\Phi^r(y^i, s), a; \{y^j\}\right) r_s(da)\, ds. \tag{4.3}$$

For $B_X \in \mathcal{B}(E_X), y^j \in E_Y^0$ and $r \in \mathcal{R}$, the observation kernel is given by

$$Q^{obs}\left(B_X \mid y^j, r\right) = \int_{B_X} f_\epsilon\left(x - \psi(y^j)\right)\nu(dx). \tag{4.4}$$

Proof. Follows directly from the definition of $\tilde{q}_{SXY}$, see Definition 2.6. Actually, under Assumption 2.16 (finite dimensional case), the transition law $\tilde{q}_{SXY}$ is given by

$$\tilde{Q}_{SXY}([0,t] \times B_X \times \{y^j\} \mid \tilde{Y}_n = y^i, r) = \int_0^t \exp\left(-\Lambda^r(y^i, s')\right) \int_A \lambda^A\left(\Phi^r(y^i, s'), a\right)\; Q^A\left(\Phi^r(y^i, s'), a; \{y^j\}\right)\; r_{s'}(da)\; ds' \int_{B_X} f_\epsilon\left(x - \psi(y^j)\right)\nu(dx),$$

where $B_X \in \mathcal{B}(E_X)$ and $y^i, y^j \in E_Y^0$. Now, the results follow by setting $t = \infty$ in the above, i.e. $\tilde{S}_{n+1} \leq \infty$. □

In what follows, we strictly follow the approach of Feinberg in [33], pages 7ff. We also try to stick to Feinberg's notation whenever this is possible given the notation we used so far in this thesis. But note that Feinberg denotes the observable state with y while we use this notation for the unobservable state. In terms of the meaning of x and y will will stick to our notation of this thesis, for the rest of Feinberg's notation we try to stay as close to [33] as possible in order to make as transparent as possible how Feinberg's model fits to our concrete model here.

Given a posterior distribution ρ of $\tilde{Y}_k$ at stage k, denote by $R(B_X \times \{y^j\} \mid \rho, r)$ the joint probability that at stage $k+1$, the observable state belongs to $B_X \in \mathcal{B}(E_X)$ and the unobservable state is $y^j \in E_Y^0$. We then get

$$R(B_X \times \{y^j\} \mid \rho, r) = \sum_{i=1}^{q} Q^{obs}\left(B_X \mid y^j, r\right) P\left(\{y^j\} \mid y^i, r\right) \rho^i.$$

R is a stochastic kernel on $E_X \times E_Y^0$ given $\mathbf{P}(E_Y^0) \times \mathcal{R}$, see [11], Section 10.3.

The probability that observation $x \in E_X$ at stage $k+1$ belongs to $B_X \in \mathcal{B}(E_X)$, given that at stage k the posterior state probability for $\tilde{Y}_k$ is ρ and relaxed control $r \in \mathcal{R}$ is executed is then

$$R'\left(B_X \mid \rho, r\right) = \sum_{i=1}^{q}\sum_{j=1}^{q} Q^{obs}\left(B_X \mid y^j, r\right) P\left(\{y^j\} \mid y^i, r\right) \rho^i. \tag{4.5}$$

Observe that R' is a stochastic kernel on E_X given $\mathbf{P}(E_Y^0) \times \mathcal{R}$. By [11], Proposition 7.27, there exists a stochastic kernel $\hat{\chi}$ on E_Y^0 given $\mathbf{P}(E_Y^0) \times \mathcal{R} \times E_X$ such that

$$R(B_X \times \{y^j\} \mid \rho, r) = \int_{B_X} \hat{\chi}\left(\{y^j\} \mid \rho, r, x\right) R'\left(dx \mid \rho, r\right). \tag{4.6}$$

The stochastic kernel $\hat{\chi}(\cdot \mid \rho, r, x)$ defines a measurable mapping $\hat{\chi} : \mathbf{P}(E_Y^0) \times \mathcal{R} \times E_X \to \mathbf{P}(E_Y^0)$, where $\hat{\chi}(\rho, r, x)(\cdot) = \hat{\chi}(\cdot \mid \rho, r, x)$. For each pair $(\rho, r) \in \mathbf{P}(E_Y^0) \times \mathcal{R}$, the mapping $\hat{\chi}(\rho, r, \cdot) : E_X \to \mathbf{P}(E_Y^0)$ is defined $R'(\cdot \mid \rho, r)$-almost surely uniquely in $x \in E_X$ (see [11], Corollary 7.27.1).

For a posterior distribution $\rho_k \in \mathbf{P}(E_Y^0)$, a relaxed control $r \in \mathcal{R}$ and an observation x_{k+1}, the posterior distribution $\rho_{k+1} \in \mathbf{P}(E_Y^0)$ is then

$$\rho_{k+1} = \hat{\chi}(\rho_k, r, x_{k+1}).$$

However, when passing from the PO-MDP model $(\tilde{X}_k, \tilde{Y}_k)_{k\geq 0}$ to the CO-MDP model $(\rho_k)_{k\geq 0}$ where we only have the (observable) posterior distributions, the observation x_{k+1} is not available, and therefore, x_{k+1} is a random variable with the distribution $R'(\cdot \mid \rho_k, r)$.

Thus, ρ_{k+1} is a random variable with values in $\mathbf{P}(E_Y^0)$ whose distribution is defined uniquely by the stochastic kernel

$$\hat{q}(D \mid \rho, r) := \int_{E_X} \mathbb{1}_{\{\hat{\chi}(\rho,r,x)\in D\}} R'(dx \mid \rho, r), \tag{4.7}$$

where $D \in \mathcal{B}\left(\mathbf{P}(E_Y^0)\right), \rho \in \mathbf{P}(E_Y^0), r \in \mathcal{R}$ (see [38], page 87).

From Feinberg's publication [33], we now get the following results:

Theorem 4.4 (Feinberg, Theorem 3.7 in [33]). *If the stochastic kernel $P(dy \mid y, r)$ is weakly continuous and if the observation kernel $Q^{obs}(dx \mid y, r)$ is continuous in total variation, then R' is setwise continuous and Assumption (H) below holds.*

Assumption 4.5 (Feinberg, Assumption (H) in [33]). *There exists a stochastic kernel $\hat{\chi}$ on E_Y^0 given $\mathbf{P}(E_Y^0) \times \mathcal{R}$ satisfying (4.6) such that: if a sequence $(\rho_n) \subset \mathbf{P}(E_Y^0)$ converges weakly to $\rho \in \mathbf{P}(E_Y^0)$ and a sequence $(r^n) \subset \mathcal{R}$ converges (in the Young topology) to $r \in \mathcal{R}$ as $n \to \infty$, then there exists a subsequence (ρ_{n_k}, r^{n_k}) of (ρ_n, r^n) and a measurable set $B_X \in \mathcal{B}(E_X)$ with $R'(B_X \mid \rho, r) = 1$ and for all $x \in B_X$ we have the weak convergence:*

$$\hat{\chi}(\rho_{n_k}, r^{n_k}, x) \overset{n\to\infty}{\longrightarrow} \hat{\chi}(\rho, r, x).$$

We will now show that under the following Assumption on „boundedness away from zero" for λ^A, the requirements of Feinberg's Theorem 3.7, which is Theorem 4.4 above, are satisfied by the process $(\tilde{X}_k, \tilde{Y}_k)$.

Assumption 4.6. *We assume that there exists $\lambda_0 > 0$ such that $\lambda^A(y, a) \geq \lambda_0$ for all $y \in E_Y, a \in A$.*

Lemma 4.7. *The stochastic kernel $P(\cdot \mid y, r)$ is weakly continuous if Assumption 4.6 holds.*

Proof. As E_Y^0 is of finite cardinality, we have to show that for each $y^j \in E_Y^0 = \{y^1, \ldots, y^q\}$, we have convergence of $P(\{y^j\} \mid y_n, r^n) \to P(\{y^j\} \mid y^i, r)$ as $n \to \infty$ for an arbitrary sequence (y_n, r^n) with $(y_n, r^n) \to (y^i, r)$ as $n \to \infty$. As E_Y^0 is of finite cardinality, convergence of $y_n \to y^i$ means that there exists $N \in \mathbb{N}$ such that for all $n \geq N : y_n = y^i$. Hence, we have to show convergence of $P(\{y^j\} \mid y^i, r^n) \to P(\{y^j\} \mid y^i, r)$ as $N \leq n \to \infty$. Thus, consider:

$$\begin{aligned}
&\left|P(\{y^j\} \mid y^i, r^n) - P(\{y^j\} \mid y^i, r)\right| \\
&\quad = \left| \int_0^\infty \exp\left(-\Lambda^{r^n}(y^i, s)\right) \int_A \lambda^A\left(\Phi^{r^n}(y^i, s), a\right) Q^A\left(\Phi^{r^n}(y^i, s), a; \{y^j\}\right) r_s^n(da)\, ds \right. \\
&\qquad \left. - \int_0^\infty \exp\left(-\Lambda^{r}(y^i, s)\right) \int_A \lambda^A\left(\Phi^{r}(y^i, s), a\right) Q^A\left(\Phi^{r}(y^i, s), a; \{y^j\}\right) r_s(da)\, ds \right|
\end{aligned}$$

Now, the latter expression converges to zero as $n \to \infty$. The proof for this convergence is analogous to the proof of Theorem 3.8. One only needs to remember that Q^A is bounded as it is a probability measure. Further, we defined Q^A to be a weakly continuous transition kernel in Definition 1.33. Hence, as E_Y^0 is of finite cardinality, $Q^A(\cdot, \cdot; \{y^j\})$ is continuous. Further, we defined λ^A as a bounded and continuous function (see Definition 1.32). Remembering these facts about Q^A and λ^A, the product of both has the same

properties as those used of function c in the proof of Theorem 3.8. The same reasoning thus applies with the exception that dominated convergence here requires the intensity λ^A to be bounded away from zero what we assume in Assumption 4.6. Note that in Theorem 3.8 we did not need this assumption as there, we not only had $\exp\left(-\Lambda^r(y^i,s)\right)$ but $\eta^r(y^i,s) = \exp\left(-\Lambda^r(y^i,s) - \beta s\right)$ and we got an integrable upper bound by $\exp(-\beta s)$. □

Lemma 4.8. *The observation kernel $Q^{obs}(\cdot \mid y, r)$ is continuous in total variation.*

Proof. Trivial as $Q^{obs}(dx \mid y, r) = f_\epsilon(x - \psi(y))\nu(dx)$ does not depend on r and convergence in E_Y^0 is convergence in discrete topology. Hence, if $y_n \to y$ in E_Y^0, then there exists $N \in \mathbb{N}$ such that for all $n \geq N : y_n = y$ and thus $Q^{obs}(dx \mid y_n, r) = Q^{obs}(dx \mid y, r)$ for $n \geq N$. □

The last two results combined with Theorem 4.4 now showed that R' is setwise continuous and that there exists a filter $\hat{\chi}$ satisfying Assumption (H). In preparation of the next section we point out the following two results.

Corollary 4.9. *The transition kernel R' is given by*

$$R'(dx \mid \rho, r) = \sum_{i=1}^{q}\sum_{j=1}^{q} \rho^i f_\epsilon\left(x - \psi(y^j)\right)\nu(dx) \int_0^\infty \exp\left(-\Lambda^r(y^i,s)\right) \int_A \lambda^A\left(\Phi^r(y^i,s),a\right) Q^A\left(\Phi^r(y^i,s),a;\{y^j\}\right) r_s(da)\, ds. \quad (4.8)$$

Proof. Follows directly from (4.3), (4.4) and (4.5). □

The next result will be important for the proof of the main result of this Chapter. When combining one-step cost and discounting, a factor $e^{-\beta s}$ will appear. Hence, in preparation for the following we state the next result.

Lemma 4.10. *Let β the discount factor of the initial optimization problem. Adding the factor $e^{-\beta s}$ to R' in order to define*

$$\hat{R}'(dx \mid \rho, r) := \sum_{i=1}^{q}\sum_{j=1}^{q} \rho^i f_\epsilon\left(x - \psi(y^j)\right)\nu(dx) \int_0^\infty \exp(-\beta s)\exp\left(-\Lambda^r(y^i,s)\right) \int_A \lambda^A\left(\Phi^r(y^i,s),a\right) Q^A\left(\Phi^r(y^i,s),a;\{y^j\}\right) r_s(da)\, ds$$

leads to $\hat{R}'$ being a setwise continuous measure that can be normalized, for each pair (ρ, r), to a probability measure. Furthermore, $R'(dx \mid \rho, r)$ and $\hat{R}'(dx \mid \rho, r)$ have the same null sets.

Proof. Inserting the factor $\exp(-\beta s)$ does not change the value of the integral $\int_0^\infty \cdots ds$ to zero where this integral was not zero without this factor and vice versa. Hence, the null sets are not changed by inserting this factor. In order to prove setwise convergence, one would investigate separately every summand for $i, j \in \{1, \ldots, q\}$. Here, convergence of $\int_0^\infty \cdots ds$ still holds by the same reasoning as in Lemma 4.7. □

4.5 Existence of optimal policies

Based on the filter $\hat{\chi}$ satisfying Assumption (H) as developed in the previous Section, we can now define a filtered model derived from the pseudo-embedded process $(\tilde{S}_k, \tilde{X}_k, \tilde{Y}_k)$. Analogously to Definition 2.32 while simply replacing χ by $\hat{\chi}$, we obtain a transition law for this derived filtered process $(\tilde{S}_k, \tilde{X}_k, M_k)$ defined for $B_X \in \mathcal{B}(E_X)$ and $C \in \mathcal{B}(\mathbf{P}(E_Y^0))$ by

$$\hat{Q}'_{SXM}([0,t] \times B_X \times C \mid \rho_{k-1}, r) := \sum_{i=1}^{q} \int_0^t \int_{B_X} \mathbb{1}_{\{\hat{\chi}(\rho_{k-1},r,x) \in C\}} \, \tilde{Q}_{SXY}\left(ds \otimes dx \otimes E_Y^0 \mid \tilde{Y}_{k-1} = y^i, r\right) \rho_{k-1}^i. \quad (4.9)$$

Now, as we restricted our model to running cost functions that do not depend on the observable state $x \in E_X$, we got one-setp cost functions g and g' that do not depend on x, see Section 4.2. This finally leads to a value function $\hat{J}'$ not depending on the initial observation $x \in E_X$ but only on the initial conditional distribution of $\tilde{Y}_0$, which in our model is given by $Q_0(x;\cdot)$ where x is the initial observation of $\tilde{X}_0$. Hence, implicitely, $\hat{J}'$ still depends on x but this is only due to the way how we determine an initial distribution of $\tilde{Y}_0$ in our model.

With this in mind and in view of Definition 2.41, the adequate classes of functions $\hat{\mathbf{B}}^+$ and $\hat{\mathbf{C}}^+_{low}$ for the model where c is not depending on the observable state are:

Definition 4.11. *For the state space $\mathbf{P}(E_Y^0)$, we define the following classes of functions:*

$$\hat{\mathbf{B}}^+(\mathbf{P}(E_Y^0)) := \{w : \mathbf{P}(E_Y^0) \to [0,\infty] \mid \ w \text{ is measurable}\}$$

$$\hat{\mathbf{C}}^+_{low}(\mathbf{P}(E_Y^0)) := \{w : \mathbf{P}(E_Y^0) \to [0,\infty] \mid \ w \text{ is lower semi-continuous}\}.$$

As the inter-jump time is not observable in this model, we have to define decision rules that do not depend on the inter-jump time, see also Section 4.3. We thus define (compare Definition 2.37):

Definition 4.12. *For a derived filtered model with unobservable inter-jump time and cost not depending on the observable state, we define:*

(i) *A* state dependent decision rule *for the derived filtered model with unobservable inter-jump time is a measurable mapping*

$$\hat{f} : \mathbf{P}(E_Y^0) \to \mathcal{R}.$$

We write $\hat{F}$ for the set of all decision rules.

(ii) *A* Markov policy *for the derived filtered model with unobservable inter-jump time is a sequence $\hat{\pi}^M = (\hat{f}_0, \hat{f}_1, \dots)$ of state dependent decision rules, where we apply decision rule $\hat{f}_n$ at stage n of the process. We denote the set of all Markov policies for the filtered model by $\hat{\Pi}^M := \hat{F}^\infty = \times_{i=0}^\infty \hat{F}$.*

Based on the above introduced classes of functions and decision rules, the operators H, $\mathcal{T}_{\hat{f}}$ and $\mathcal{T}$ take the following form (compare Definition 2.42):

Definition 4.13. *For $\rho \in \mathbf{P}(E_Y^0)$, $r \in \mathcal{R}$, $\hat{f} \in \hat{F}$ and $w \in \hat{\mathbf{B}}^+(\mathbf{P}(E_Y^0))$ we define:*

(i) $(Hw)(\rho, r) := g'(\rho, r) + \int_{\mathbf{P}(E_Y^0)} w(\rho') \int_0^\infty e^{-\beta s'} \int_{E_X} \hat{Q}'_{SXM}(ds' \otimes dx' \otimes d\rho' \mid \rho, r)$

(ii) $(\mathcal{T}_{\hat{f}}w)(\rho) := (Hw)(\rho, \hat{f}(\rho))$

(iii) $(\mathcal{T}w)(\rho) := \inf_{r\in\mathcal{R}}(Hw)(\rho), r) = \inf_{\hat{f}\in\hat{F}}(\mathcal{T}_f w)(\rho)$.

A closer look on the definition of operator H above brings up:

Lemma 4.14. *For the integral part of the definition of operator H, it holds:*

$$\int_{\mathbf{P}(E_Y^0)} w(\rho') \int_0^\infty e^{-\beta s'} \int_{E_X} \hat{Q}'_{SXM}(ds' \otimes dx' \otimes d\rho' \mid \rho, r) = \int_{\mathbf{P}(E_Y^0)} w(\rho')\ \hat{Q}'(d\rho' \mid \rho, r),$$

where for $D \in \mathcal{B}(\mathbf{P}(E_Y^0))$, we define (compare also (4.7)):

$$\hat{Q}'(D \mid \rho, r) := \int_{E_X} \mathbb{1}_{\{\hat{\chi}(\rho,r,x)\in D\}} \hat{R}'(dx \mid \rho, r). \tag{4.10}$$

Proof. Based on (4.9) we obtain:

$$\int_{\mathbf{P}(E_Y^0)} w(\rho') \int_0^\infty e^{-\beta s'} \int_{E_X} \hat{Q}'_{SXM}(ds' \otimes dx' \otimes d\rho' \mid \rho, r)$$
$$= \int_{E_X} w\left(\hat{\chi}(\rho, r, x')\right) \int_0^\infty e^{-\beta s'} \sum_{i=1}^q \rho^i\ \tilde{Q}_{SXY}\left(ds' \otimes dx' \otimes E_Y^0 \mid y^i, r\right).$$

Now applying the definition of $\tilde{q}_{SXY}$, see Definition 2.6, the latter expression becomes:

$$= \sum_{i=1}^q \sum_{j=1}^q \rho^i \int_{E_X} w\left(\hat{\chi}(\rho, r, x')\right) f_\epsilon\left(x' - \psi(y^j)\right) \nu(dx') \int_0^\infty \exp(-\beta s') \exp\left(-\Lambda^r(y^i, s')\right)$$
$$\int_A \lambda^A\left(\Phi^r(y^i, s\prime), a\right) Q^A\left(\Phi^r(y^i, s'), a; \{y^j\}\right) r_{s'}(da)\ ds',$$

and by definition of $\hat{R}'(dx \mid \rho, r)$ (see Lemma 4.10) the result follows. □

As $g'(\rho, r)$ is lower semi-continuous in (ρ, r) by Theorem 3.8 (that was shown for g' even depending on x) as long as Assumption 3.1 and Assumption 3.2 hold, our model with running cost not depending on the observable state satisfies Assumption (LSC). In order to prove existence of one-step optimizers (see Lemma 2.45), we need Hw to be lower semi-continuous for $w \in \hat{\mathbf{C}}^+_{low}(\mathbf{P}(E_Y^0))$. As we showed in Lemma 2.44 (i), this property is satisfied if we can show that $\hat{q}'(\cdot \mid \rho, r)$ is a weakly continuous measure. In that sense, the next result is the main result of this Chapter as with the next result, existence of one-step optimizers follows. With the existence of one-step optimizers, all results of Sections 2.3.3 and 2.3.4 hold in adapted version.

Theorem 4.15. *The transition kernel $\hat{Q}'(\cdot \mid \rho, r)$ is weakly continuous.*

The proof we present here for the above Theorem follows the ideas of Feinberg's proof for his Theorem 3.5 in [33]. However, we added the idea of replacing Feinberg's R' by our $\hat{R}'$ which still makes the proof possible because of Lemma 4.10. We need the following result on a generalized version of Fatou's Lemma:

Lemma 4.16. *Let S an arbitrary metric space, $(\mu_n)_{n\geq 0}$ a sequence in $\mathbf{P}(S)$ and $(f_n)_{n\geq 0}$ a sequence of measurable nonnegative $\overline{\mathbb{R}}$-valued functions on S. If μ_n converges setwise to $\mu \in \mathbf{P}(S)$ for $n \to \infty$, then*

$$\int_S \liminf_{n\to\infty} f_n(s)\mu(ds) \leq \liminf_{n\to\infty} \int_S f_n(s)\mu_n(ds). \tag{4.11}$$

Proof. For a proof, see [55], page 231. □

With this result we can now give the proof of Theorem 4.15:

Proof (of Theorem 4.15). According to Parthasarathy ([51], Theorem 6.1, p.40), Billingsley ([12], Theorem 2.1), the stochastic kernel $\hat{Q}'(d\rho' \mid \rho, r)$ from $\mathbf{P}(E_Y^0) \times \mathcal{R}$ to $\mathbf{P}(E_Y^0)$ is weakly continuous if and only if $\hat{Q}'(D \mid \rho, r)$ is lower semi-continuous in $(\rho, r) \in \mathbf{P}(E_Y^0) \times \mathcal{R}$ for every open set $D \subset \mathbf{P}(E_Y^0)$, that is,

$$\liminf_{n\to\infty} \hat{Q}'(D \mid \rho_n, r^n) \geq \hat{Q}'(D \mid \rho, r), \tag{4.12}$$

for all sequences $(\rho_n, r^n)_{n\geq 0} \in \mathbf{P}(E_Y^0) \times \mathcal{R}$ with $\rho_n \to \rho \in \mathbf{P}(E_Y^0)$ weakly and $r^n \to r \in \mathcal{R}$ in Young topology as $n \to \infty$. We show (4.12) by contradiction, thus, suppose that for some sequence $(\rho_n, r^n)_{n\geq 0} \in \mathbf{P}(E_Y^0) \times \mathcal{R}$ with the required convergence properties we have

$$\liminf_{n\to\infty} \hat{Q}'(D \mid \rho_n, r^n) < \hat{Q}'(D \mid \rho, r).$$

Then, there exists $\epsilon > 0$ and a subsequence $(\rho_{n_k}, r^{n_k})_{k\geq 0}$ of $(\rho_n, r^n)_{n\geq 0}$ such that

$$\hat{Q}'(D \mid \rho_{n_k}, r^{n_k}) \leq \hat{Q}'(D \mid \rho, r) - \epsilon, \qquad k = 1, 2, \ldots. \tag{4.13}$$

Now by Theorem 4.4 together with Lemma 4.7 and Lemma 4.8, there exists a subsequence $(\rho_{n_{k_l}}, r^{n_{k_l}})_{l\geq 0}$ of (ρ_{n_k}, r^{n_k}) such that $\hat{\chi}\left(\rho_{n_{k_l}}, r^{n_{k_l}}, x\right) \to \hat{\chi}(\rho, r, x)$ weakly as $l \to \infty$, $R'(\cdot \mid \rho, r)$-almost surely in $x \in E_X$. By Lemma 4.10, the latter convergence is as well $\hat{R}'(\cdot \mid \rho, r)$-almost surely in $x \in E_X$.

Since D is open in $\mathbf{P}(E_Y^0)$, we thus get

$$\liminf_{l\to\infty} \mathbb{1}_{\left\{\hat{\chi}\left(\rho_{n_{k_l}}, r^{n_{k_l}}, x\right) \in D\right\}} \geq \mathbb{1}_{\{\hat{\chi}(\rho,r,x)\in D\}}, \quad \hat{R}'(\cdot \mid \rho, r)\text{-almost surely in } x \in E_X. \tag{4.14}$$

Now apply Lemma 4.16 to the setwise converging sequence (see Lemma 4.10) $\hat{R}'(\cdot \mid \rho_{n_{k_l}}, r^{n_{k_l}})$ to obtain

$$\begin{aligned}\liminf_{l\to\infty} \int_{E_X} \mathbb{1}_{\left\{\hat{\chi}\left(\rho_{n_{k_l}}, r^{n_{k_l}}, x\right) \in D\right\}} \hat{R}'(dx \mid \rho_{n_{k_l}}, r^{n_{k_l}}) \\ \geq \int_{E_X} \liminf_{l\to\infty} \mathbb{1}_{\left\{\hat{\chi}\left(\rho_{n_{k_l}}, r^{n_{k_l}}, x\right) \in D\right\}} \hat{R}'(dx \mid \rho, r).\end{aligned} \tag{4.15}$$

By (4.10), we recognize the lefthand side of (4.15) as $\liminf_{l\to\infty} \hat{q}'\left(D \mid \rho_{n_{k_l}}, r^{n_{k_l}}\right)$. Applying (4.14), monotonicity of integrals and (4.10) to the righthand side of (4.15) we finally get

$$\liminf_{l\to\infty} \hat{Q}'\left(D \mid \rho_{n_{k_l}}, r^{n_{k_l}}\right) \geq \hat{Q}'\left(D \mid \rho, r\right),$$

which contradicts (4.13). Hence, (4.12) holds. □

Chapter 5

Application: Models with convex cost function

Throughout the last Chapters, we developed a PO-PDMP control model and showed existence of optimal policies for the resulting optimization problem of minimizing total discounted cost over lifetime. However, we only showed that an optimal policy exists. We did neither address the question of uniqueness of an optimal policy, nor characterize an optimal policy further.

The goal of this Chapter is thus twofold: We will provide a concrete example for the PO-PDMP control theory developed. We restrict our investigations to one concrete model here, further domains of application of the theory will be discussed in Chapter 6. For this concrete model, we will then address the question of uniqueness of an optimal policy. It will turn out, that for the running cost function being strictly convex, in this model, there will exist a unique optimal policy. We will even be able to characterize this optimal policy as being deterministic and of „bang-bang“ type, i.e. the agent will always try to achieve as quickly as possible a certain system state.

As characterization of optimal policies is not the topic of many publications, even for completely observable PDMP control problems, we will analyze the concrete application example in both situations: under complete as well as under partial observation. We will also link both versions of the example by showing that the completely observable version is contained as a special case in our partially observable model developed.

The outline of this chapter is thus the following: In Section 5.1, we will motivate very briefly the concrete application model we will study in detail in this chapter. The mathematical model with constant jump intensity and uncontrolled jump transition kernel is then introduced in Section 5.2. Once the mathematical model introduced, we will show how the completely observable version of the example can be understood as a special case of the PO-PDMP model, see Section 5.3. We then proceed to the characterization of an optimal policy for the completely observable model in Section 5.4 and for the partially observable model in Section 5.5.

5.1 Motivation: Optimal control of production lines

In the next Section, we will present a concrete mathematical example of an optimal control problem for PO-PDMPs with convex (and uncontrolled) running cost function c, constant intensity λ and uncontrolled transition kernel Q. We will further assume that the running

cost will only depend on the unobservable component of the PO-PDMP. The goal of this present Section is to motivate the mathematical model by some examples of optimal control of production processes.

Imagine a production process where one requires a tank to be filled up to, e.g. half of its capacity throughout the whole production process. One could think of plastics-processing industries, where a tank for plastics granulate is feeding the production line. One could also think of a production process in chemicals industries where a fluid (e.g., a solution of some crystal) has to be of a given concentration in a tank feeding the production process. Or, as a third example, think of a production process where air temperature (or even, analogously, humidity in the air) has to be of a given temperature (concentration). Assume further, that, in all these examples, final products of perfect quality are produced as long as fill levels of tanks, concentration of solutions in tanks, temperatures etc. are kept at the required levels.

Imagine now, in the case of chemicals production (others analogously), that the following happens: At random points in time, goverened by intensity λ, the concentration of the chemicals solution in the tank jumps to higher or lower values, goverend by the jump transition kernel Q. This is happening due to technical issues, machine break downs, etc., that cannot be influenced by the agents. When running the production process with other concentrations than required in the tank, however, quality issues arise in the final products of the production process. Customers will recognize these quality issues and pay less for the products. Thus, penalty cost in terms of foregone profit occure in that case. In many cases, quality issues arise in both cases: too high and too low concentration in the tank. Very often, arising penalty cost is thus symmetric for too high and too low concentration in the tank. Also, penalty cost grow faster the more the optimal concentration in the tank is missed. Hence, one can assume convex (and very often even symmetric) cost functions to model these penalty cost.

An agent can control the concentration in the tank via two mechanisms: (Noisy) Measurements of the concentration right after a jump of concentration and by adding, say water (to lower concentration) or more of the crystal (to increase concentration). The latter is done by pumping in water or crystals, while there is an upper bound for quantity pumped in by minute. The optimization problem for the agent is thus to find, at every point in time, the right pumping rate, in order to keep the concentration in the tank on a level that minimizes discounted penalty cost over lifetime.

The resulting optimization problem is not trivial, as the policy to always keep the concentration as close as possible to the required concentration for perfect product quality is not necessarily the optimal policy. All actually depends heavily on the nature of the jump transition kernel. If this kernel is such that, a jump would almost surely lead to the highest concentration possible if the concentration right before the jump was at this required concentration for perfect product quality, then the following policy might be better: Try to keep concentration on a level close to the required concentration but such that occurring jumps do not lead to concentrations very different from the required concentration. We will formalize this properly during the following Sections.

5.2 Mathematical Model

We will now present the detailed mathematical model we want to investigate throughout the rest of this Chapter. Assume a controlled PO-PDMP as defined in Section 1.3, where we specify the following:

Definition 5.1. *We consider a controlled PO-PDMP as in Definition 1.34, where the underlying spaces are defined as follows:*

(i) The state spaces are $E_Y := \mathbb{R}$ and $E_X := \mathbb{R}$ and in view of Assumption 2.16, let $q \in 2\mathbb{N}$ and $y^1, \dots, y^q \in \mathbb{R}$ with $y^1 < y^2 < \cdots < y^q$ points on the real line forming $E_Y^0 := \{y^1, \dots, y^q\}$ the space of possible post jump states.

(ii) The action space is defined as $A := [-a_{dec}, a_{inc}]$ for $a_{dec}, a_{inc} > 0$.

The noisy observation of the post-jump state is modeled as follows:

(ii) The observation process is defined as in Definitions 1.8, 1.9 and 1.11, where we define here: $\psi : E_Y = \mathbb{R} \to \mathbb{R} = E_X, y \mapsto \psi(y) = y$. We further keep Assumption 1.13, i.e. we assume to have a bounded density f_ϵ of measurement noise, where f_ϵ is the Lebesgue density of the measurement noise.

The characteristics of the underlying PDMP (Y_t) are defined as follows:

(iii) Let $\lambda > 0$ a constant jump intensity. We further assume λ to be uncontrolled, i.e. $\lambda^A(y, a) := \lambda$ for all $y \in E_Y, a \in A$.

(iv) Let $Q : E_Y \to \mathbf{P}(E_Y^0)$ a weakly continuous transition kernel. We assume Q to be uncontrolled, i.e. $Q^A(y, a; \cdot) = Q(y; \cdot)$ for all $y \in E_Y$ and $a \in A$.

(v) Let the controlled drift Φ^r defined by the following initial value problem:

$$\frac{d}{dt}\Phi^r(y,t) = \int_A a \; r_t(da), \qquad \Phi^r(y,0) = y.$$

Remark 5.2. *The above definition of the controlled drift satisfies the requirements of Theorem 3.3 with $b(y, a) := a$. Hence, Assumption 3.1 of continuous dependence of the drift on the relaxed control $r \in \mathcal{R}$ is satisfied. Further, one can reduce the set of admissible controls to the set of deterministic controls $\mathcal{U}$ of measurable mappings $u : \mathbb{R}^+ \to A$. Actually, $u_t := \int_A a \; r_t(da)$ is of class $\mathcal{U}$ for $r \in \mathcal{R}$.*

The above defined controlled PO-PDMP can be understood as a mathematical model for the chemicals production example of the previous Section. The interpretation is the following:

A state $y \in E_Y$ is the current, unobservable concentration of the crystal solution in the tank. Note that for a problem of concentrations[1], one could pass to the state space $E_Y = [0, 1]$ instead of $E_Y = \mathbb{R}$. However, as there is a homeomorphism between both spaces, we decided to keep $E_Y = \mathbb{R}$ here in order to make references to earlier results easier.

The agent gets a noisy measurement $x \in E_X$ of the form $x = \epsilon + y$ if the post-jump state is y. Here, ϵ is the measurement noise with density f_ϵ. The agent can thus observe the inter-jump time s as well as the noisy measurement x for all jump times T_n of the process (Y_t). A (deterministic) decision rule at stage n is a measurable mapping $\pi_n^A : \mathcal{H}_n \times \mathbb{R}^+ \to A$. We further define the set of deterministic policies by $\Pi^A := \times_{n=0}^{\infty} \Pi_n^A$, where Π_n^A is the set of measurable mappings $\pi_n^A : \mathcal{H}_n \times \mathbb{R}^+ \to A$. Note that according to the remark above, we can restrict our investigations to the set of deterministic controls. In fact, the theory will later deliver existence of an optimal relaxed control policy. The

[1] which are typically percentage values between 0% and 100%

previous remark, however, shows that at each point in time t, the only property of interest of this optimal relaxed control r is its expectation (as distribution on A) at time t.

The control an agent can execute is thus changing the concentration in the tank by pumping in water (leading to decreasing concentration, thus modeled by values of $u_t \in [-a_{dec}, 0)$) or by pumping in more of the crystal (leading to increasing concentration, thus modeled by values of $u_t \in (0, a_{inc}]$).

We assume, that the optimal concentration (where no quality issues arise for the final product produced) in the tank is given by the arithmetic mean of $y^{\frac{q}{2}}$ and $y^{\frac{q}{2}+1}$, hence, by $\frac{y^{\frac{q}{2}} + y^{\frac{q}{2}+1}}{2}$. Penalty cost shall be minimal for this concentration. We thus define the running cost function as follows:

Definition 5.3. *The running cost function $c : E_Y \to \mathbb{R}^+$ is assumed to be strictly convex with global minimum at $y^\star := \frac{y^{\frac{q}{2}} + y^{\frac{q}{2}+1}}{2}$.*

Remark 5.4. *Note that a convex function on $\mathbb{R}$ is continuous, hence c is continuous.*

This running cost function is thus uncontrolled and only dependent on the unobservable state of the PO-PDMP. As typically, for this kind of problems in production processes, penalty cost grows (proportionally) the more the concentration in the tank deviates from the optimal concentration, we assume strict convexity of this cost function.

The optimization problem is thus, to minimize over lifetime, the total penalty cost arising from quality issues due to concentration levels in the tank different from $y^\star$, discounted at a constant discount factor $\beta > 0$. This problem is formally defined as:

Definition 5.5. *Let $\beta > 0$ a discount factor, $x \in E_X$ and Π^P the set of history dependent relaxed piecewise open loop policies as defined in Definition 1.27. We define the cost of a policy $\pi \in \Pi^P$ as*

$$J(x, \pi) := \mathbb{E}_x^\pi \left[\int_0^\infty e^{-\beta t} c(Y_t)\ dt \right].$$

The value function of our problem is defined for all $x \in E_X$ as

$$J(x) := \inf_{\pi \in \Pi^P} J(x, \pi),$$

and the optimization problem to solve is: For all $x \in E_X$, find $\pi^\star \in \Pi^P$ such that

$$J(x, \pi^\star) = J(x). \tag{5.1}$$

When comparing the above optimization problem to the general optimization problem in Definition 1.37, one will note that the principal difference between both is the fact that here, the running cost function is not depending on (X_t) and not on $a \in A$. However, the general theory developed in Chapters 1 to 3 still holds under this restriction of the running cost function. We summarize this by the following

Theorem 5.6. *For each $x \in E_X$, there exists $\pi^\star \in \Pi^P$ such that* (5.1) *holds.*

Proof. Follows from Theorem 2.60 (existence of optimal policies for the infinite time horizon problem), as all assumptions taken for the proof of Theorem 2.60 are satisfied by the model defined here in this section. The assumptions to check are summarized in Annex C. With the notations of Annex C, we find the following assumptions clearly satisfied by the here discussed model: (SP), (SF), (OS), (ON), (A), (IC), (IL), (QC).

Some comments on the following assumptions: (D) is satisfied, as here, we have an ODE defined drift, see Definition 5.1 (v) above. Further, (DC) is satisfied as explained in Remark 5.2. The lower semi-continuity (CC) is satisfied as c is not depending on x and a but assumed to be strictly convex in y, see also Remark 5.4. Finally, (IQ) is satisfied as we assume here uncontrolled jump intensity and uncontrolled jump transition kernel. □

As one can understand A as a subset of $\mathbf{P}(A)$ by $a \mapsto \delta_a$, we immediately get the following

Corollary 5.7. *For each $x \in E_X$, there exists an optimal deterministic policy $\pi^{\star A} \in \Pi^A$ such that $J(x, \pi^{\star A}) = J(x)$.*

Proof. Clear from Theorem 5.6, the above mentioned inclusion of A into $\mathbf{P}(A)$ and from Remark 5.2 together with the Correpondence Theorem 2.11. □

In preparation for the following Sections, we end this Section by developing a representation of the value function J' of the derived filtered model of the present concrete PO-PDMP model. This representation will be key for all further investigations.

Actually, as developed in Chapter 2, existence of optimal policies is shown by a reformulation of the initial problem into an equivalent derived filtered problem. For the latter, we showed existence of optimal Markov Policies and explained how these can be understood as history dependent policies for the pseudo-embedded problem. The Correspondence Theorem finally explained the correspondence between a history dependent policy for the pseudo-embedded problem and a history dependent policy for the initial PO-PDMP control problem.

Hence, trying to understand the nature or characteristics of an optimal policy for the initial control problem for the PO-PDMP passes by a study of the optimal policies of its derived filtered control problem. For the latter, the value function J' is a fixed point of the $\mathcal{T}$ operator as shown in Theorem 2.59. Based on this property of the value function, we obtain

Lemma 5.8. *For the PO-PDMP model defined in Definition 5.1 and the associated optimization problem of Definition 5.5, the value function J' of the derived filtered problem is given by*

$$J'(s,x,\rho) = J'(\rho) = \inf_{r\in\mathcal{R}} \left\{ \sum_{i=1}^{q} \rho^i \int_0^\infty e^{-(\beta+\lambda)s'} \left[c\left(\Phi^r(y^i,s')\right) + \lambda \sum_{j=1}^{q} Q\left(\Phi^r(y^i,s');\{y^j\}\right) \right.\right. \\ \left.\left. \int_{\mathbb{R}} f_\epsilon\left(x' - \psi(y^j)\right) J'\left(\chi(\rho,x',s',r)\right)\ dx' \right]\ ds' \right\}. \quad (5.2)$$

In Particular, J' is only a function of the initial conditional distribution of the unobservable component, it is not depending on the initial noisy measurement x of Y_0, neither on the initial inter-jump time.

Proof. By Theorem 2.59 and by the definition of operators T and H in Definition 2.42, we have:

$$\begin{aligned} J'(s,x,\rho) &= \left(TJ'\right)(s,x,\rho) \\ &= \inf_{r\in\mathcal{R}} \left\{ \left(HJ'\right)((s,x,\rho),r) \right\} \\ &= \inf_{r\in\mathcal{R}} \left\{ g'(s,x,\rho,r) + \int_{E'} e^{-\beta s'} J'(s',x',\rho')\ q'_{SXM}\left(ds',dx',d\rho' \mid \rho,r\right) \right\}. \end{aligned}$$

Now, by (2.44) and Lemma 3.6 (remember that λ is constant in this application here), we obtain:

$$g'(s,x,\rho,r) = g'(x,\rho,r) = \sum_{i=1}^{q} \rho^i \int_0^\infty e^{-(\beta+\lambda)s'} c\left(\Phi^r(y^i,s')\right)\ ds'.$$

By Definition 2.32 of q'_{SXM} we get:

$$\int_{E'} e^{-\beta s'} J'(s',x',\rho')\ q'_{SXM}\left(ds',dx',d\rho' \mid \rho,r\right)$$
$$= \sum_{i=1}^{q} \rho^i \int_0^\infty \int_{E_X} e^{-\beta s'} J'\left(s',x',\chi(\rho,x',s',r)\right)\ \tilde{q}_{SXY}\left(ds',dx',\tilde{Y}\in E_Y^0 \mid y^i,r\right).$$

Applying Definition 2.32 of $\tilde{q}_{SXY}$ to this expression leads to:

$$= \sum_{i=1}^{q} \rho^i \int_0^\infty e^{-(\beta+\lambda)s'} \lambda \sum_{j=1}^{q} Q\left(\Phi^r(y^i,s');\{y^j\}\right)$$
$$\int_{E_X} f_\epsilon\left(x'-\psi(y^j)\right) J'\left(s',x',\chi(\rho,x',s',r)\right)\ \nu(dx')\ ds'$$

Now, the result follows by linearity of integrals and applying that $E_X = \mathbb{R}$ endowed with the Lebesgue measure. □

5.3 Completely observable model as special case of partially observable model

The goal of this Section is to show that the PO-PDMP model presented includes the case of a completely observable PDMP model. Perfect observation means no measurement noise, which can be modeled by $f_\epsilon = \delta_0$, i.e. almost surely the measurement noise is zero. Perfect observation also means, that the initial state of the process is known, which can be modeled by $Q_0(y;\cdot) = \delta_y$, where y is the initial observation under $f_\epsilon = \delta_0$.

We will show, that under these two assumptions for the measurement noise and the initial conditional distribution, the value function J' of the derived filtered model is actually the same as the value function of the corresponding completely observable PDMP control problem. We will show this property for the application example of this Chapter but the result also holds for the general model with analogous proofs.

In [34], Forwick studied the question of existence of optimal policies for completely observable PDMP control problems. We cannot develop the full theory for completely observable PDMP control problems here, but based on the first two Chapters of this thesis, the general approach shall be clear: Where we had to do two reformulations, only one reformulation of the initial problem is required in the completely observable case. Passing from the initial problem to the problem for the pseudo-embedded process leads to a completely observable MDP. Thus, Forwick does this first reformulation and introduces thereafter operators $H, \mathcal{T}_f$ and $\mathcal{T}$ for the resulting optimization problem of the pseudo-embedded process. As we already introduced operators with these names, we will call the operators for the completely observable PDMP problem of Forwick $K, \mathcal{L}_f$ and $\mathcal{L}$ respectively, compare Definition 2.3.1 in [34]. Further, we will denote the value function of the completely observable PDMP control problem by v. Analogously to our results,

Forwick shows that v is a fixed point of the operator $\mathcal{L}$, i.e. $v(y) = (\mathcal{L}v)(y)$. Based on Forwick's results and under the assumption of only having finitely many post jump states, having a constant intensity λ and constant discount factor β as well as an uncontrolled jump transition kernel, the value function for the completely observable PDMP control problem satisfies for $y^i \in E_Y^0$:

$$v(y^i) = \inf_{r\in\mathcal{R}} \left\{ \int_0^\infty e^{-(\beta+\lambda)t} \left[c\left(\Phi^r(y^i,t)\right) + \lambda \sum_{j=1}^q Q\left(\Phi^r(y^i,t);\{y^j\}\right) v(y^j) \right] \, dt \right\}. \tag{5.3}$$

The analogy of (5.3) to (5.2) is clear, but still, a proper proof is required to show that $J'(\delta_k) = v(y^k)$ for $k = 1,\dots,q$ if $f_\epsilon = \delta_0$. This will be the main result of this Section, but we first need the following

Lemma 5.9. *If $f_\epsilon = \delta_0$ in the model of Definition 5.1, then for $i,j \in \{1,\dots,q\}$ it holds:*

$$\chi(\delta_i, y^j, t, r) = \delta_j \qquad \forall t \geq 0 \ \forall r \in \mathcal{R}.$$

Proof. First, note that under the assumption $f_\epsilon = \delta_0$, an observation x almost surely has the form $x = \psi(y^j) + 0$ and with ψ the identity in the application example of this Chapter, an observation has thus the form $x = y^j$ for some $j \in 1,\dots,q$.

Now, based on the definition of the filter equation in Definition 2.30, we obtain for the model of Definition 5.1, for $i,j,k,l \in \{1,\dots,q\}$:

$$\chi_k^l(\delta_i, y^j, t, r) = \delta_{ik} e^{-\lambda t} \lambda Q\left(\Phi^r(y^k,t);\{y^l\}\right) f_\epsilon(y^j - y^l) = \delta_{ik}\delta_{jl} e^{-\lambda t} \lambda Q\left(\Phi^r(y^k,t);\{y^l\}\right).$$

Hence, for $l \neq j$, we have $\chi^l(\delta_i, y^j, t, r) = \frac{1}{\overline{\chi}} \sum_{k=1}^q \chi_k^l(\delta_i, y^j, t, r) = 0$.

For $l = j$ we obtain $\chi^j(\delta_i, y^j, t, r) = \frac{1}{\overline{\chi}} \sum_{k=1}^q \chi_k^j(\delta_i, y^j, t, r) = \frac{1}{\overline{\chi}} \chi_i^j(\delta_i, y^j, t, r)$.

Finally, with $\overline{\chi}(\delta_i, y^j, t, r) = \sum_{k=1}^q \sum_{l=1}^q \chi_k^l(\delta_i, y^j, t, r) = \chi_i^j(\delta_i, y^j, t, r)$ we obtain

$$\chi^j(\delta_i, y^j, t, r) = 1$$

. □

With this result in mind, that starting from a precise observation and having no measurement noise, the filter will deliver a new conditional distribution that is concentrated in a point, we can formulate the main result of this Section:

Theorem 5.10. *Let $f_\epsilon = \delta_0$ in the model defined in Definition 5.1, then for $i \in \{1,\dots,q\}$, it holds:*

$$J'(\delta_i) = v(y^i).$$

Proof. Let $i \in \{1,\dots,q\}$ and $f_\epsilon = \delta_0$. We proceed by induction to show $(\mathcal{T}^m \underline{0})(\delta_i) = (\mathcal{L}^m \underline{0})(y^i)$ for all $m \in \mathbb{N}$.

By definition of $\mathcal{T}$ and g', we obtain

$$\begin{aligned}
(\mathcal{T}\underline{0})(\delta_i) &= \inf_{r\in\mathcal{R}} \left\{ \int_0^\infty e^{-(\beta+\lambda)t} \sum_{k=1}^q \delta_{ik} c\left(\Phi^r(y^k,t)\right) \, dt \right\} \\
&= \inf_{r\in\mathcal{R}} \left\{ \int_0^\infty e^{-(\beta+\lambda)t} c\left(\Phi^r(y^i,t)\right) \, dt \right\} \\
&= (\mathcal{L}\underline{0})(y^i),
\end{aligned}$$

where the last equation holds by definition of $\mathcal{L}$, see Forwick [34].

Let now for fix $m \geq 0$ and all $k \in \{1, \ldots, q\} : (\mathcal{T}^m \underline{0})(\delta_k) = (\mathcal{L}^m \underline{0})(y^k)$. Then:

$$
\begin{aligned}
&\left(\mathcal{T}^{m+1}\underline{0}\right)(\delta_k) \\
&= \mathcal{T}\left(\mathcal{T}^m \underline{0}\right)(\delta_k) \\
&= \inf_{r\in\mathcal{R}} \left\{ \int_0^\infty e^{-(\beta+\lambda)t} \left[c\left(\Phi^r(y^k,t)\right) + \lambda \sum_{j=1}^{q} Q\left(\Phi^r(y^k,t);\{y^j\}\right)\left(\mathcal{T}^m\underline{0}\right)\left(\chi(\delta_k, y^j, t, r)\right) \right] dt \right\} \\
&= \inf_{r\in\mathcal{R}} \left\{ \int_0^\infty e^{-(\beta+\lambda)t} \left[c\left(\Phi^r(y^k,t)\right) + \lambda \sum_{j=1}^{q} Q\left(\Phi^r(y^k,t);\{y^j\}\right)\left(\mathcal{T}^m\underline{0}\right)\left(\delta_j\right) \right] dt \right\} \\
&= \inf_{r\in\mathcal{R}} \left\{ \int_0^\infty e^{-(\beta+\lambda)t} \left[c\left(\Phi^r(y^k,t)\right) + \lambda \sum_{j=1}^{q} Q\left(\Phi^r(y^k,t);\{y^j\}\right)\left(\mathcal{L}^m\underline{0}\right)\left(y^j\right) \right] dt \right\} \\
&= \mathcal{L}\left(\mathcal{L}^m \underline{0}\right)(y^k) \\
&= \left(\mathcal{L}^{m+1}\underline{0}\right)(y^k)
\end{aligned}
$$

Thus, we have for all $m \geq 0$ and all $k \in \{1, \ldots, q\} : (\mathcal{T}^m \underline{0})(\delta_k) = (\mathcal{L}^m \underline{0})(y^k)$. With Theorem 2.58 and the corresponding result for the operator $\mathcal{L}$, see [34] Theorem 2.3.9, we finally obtain:

$$J'(\delta_k) = \lim_{m\to\infty} \left(\mathcal{T}^m \underline{0}\right)(\delta_k) = \lim_{m\to\infty} \left(\mathcal{L}^m \underline{0}\right)(y^k) = v(y^k) \qquad \forall k \in \{1, \ldots, q\}.$$

This completes the proof. □

5.4 Optimal policies in completely observable case

The goal of this Section is to characterize optimal policies for the completely observable version of the application example we introduced in the first Section of this Chapter. The guiding question here will be the question of existence of optimal policies of so-called „bang-bang“ type.

The reason why we start by studying the completely observable case first is twofold: First, the completely observable model is easier to handle than the partially observable model. Hence, ideas and approaches can be motivated and introduced, before we then try to extend them to the case of a partially observable version of the model. Second, while a broad range of publications regarding existence of optimal policies for completely observable PDMP control problems exist, there is no general result on the characteristics or properties of an optimal policy. To the best of our knowledge, we do not know about any result delivering a criteria for the existence of , e.g., „bang-bang“ type optimal policies in general. In well known publications like Davis [23], Yushkevich [64] and [62], Dempster [26] and [27] or Forwick et. al [35] - to only cite a few of this broad range of literature - the authors focused essentially on necessary and sufficient conditions for the existence of optimal policies and the question if optimal deterministic, i.e. non-relaxed, controls exist. They did not try to characterize these optimal policies in general. Some authors, like Davis, did such characterizations for very concrete application examples, e.g., for the capacity expansion problem where he also could characterize the optimal policy to be of „bang-bang“ type, see [23], equation (42.19), page 143 or even [24].

The approach we present now in order to characterize an optimal policy is inspired by some techniques coming from the theory of Stochastic Fluid Programs (SFP). We will not

need the full theory of SFPs as one can find it, e.g., in the works of Bäuerle [4] or [5]. Some ideas of these works however, have recently been used in a thesis of Chernysh [17] where an optimality equation appeared that is very similar to (5.3). Hence, we will present, in the following, an approach inspired from techniques of SFPs together with some ideas of Chernysh. As (5.3) does not fully fit into the setting of Chernysh, we will develop all ideas and intermediate results step by step. Finally, we can even weaken the original convexity requirement of Chernysh to a condition of monotonicity together with existence of a global minimum.

The major difference between a typical optimality equation coming from an SFP and (5.3) lies in the fact that in (5.3), the jump transition kernel Q depends on the current state of the process via $\Phi^r(y^i, t)$. In a typical optimality equation for an SFP, there would not be such a dependence on the current process state at this place of the equation. Hence, we split our investigations into two separate streams: First, we study the case of having Q not depending on $\Phi^r(y^i, t)$. Second, we allow Q to depend on $\Phi^r(y^i, t)$ and still try to extract sufficient conditions for an optimal policy to be of „bang-bang“ type.

Finally, one remark on admissible policies for the completely observable PDMP: By the same reasoning as for the filtered model (which is a completely observable MDP) of the PO-PDMP, optimal policies can already be found in the class of Markov policies for the pseudo-embedded process of a completely observable PDMP. Hence, we do not need history dependent policies for the completely observable PDMP and combined with the earlier discussed fact that for the present application example, deterministic controls are sufficient, we get the following class of admissible policies:

Definition 5.11. *The class of admissible policies for the PDMP control problem arising of Definition 5.1 with $f_\epsilon = \delta_0$ and $Q_0(y;\cdot) = \delta_y \ \forall y \in E_Y$ is the class Π^0 of measurable mappings $\pi^0 : E_Y^0 \times \mathbb{R}^+ \to A$.*

Remark 5.12. *Note that an adapted version of the Correspondence Theorem holds (see, e.g., Forwick [34], Theorem 2.2.14) and thus, there exists a correspondence between policies π^0 for the completely observable PDMP and policies for its pseudo-embedded MDP, denoted by $\pi^{0D} : E_Y^0 \to \mathcal{U} := \{u : [0,\infty) \to A \mid u \text{ measurable}\}$ such that*

$$\pi^0(y,\cdot) = \pi^{0D}(y)(\cdot) \qquad \lambda^1 - a.e. \text{ on } \mathbb{R}^+ \ \forall y \in E_Y.$$

5.4.1 Models with state-independent jump transition kernel

As developed earlier, the optimality equation for the value function of our application example under complete observation is given by (5.3). A closer look on this equation repeated below makes appear two summands (I) and (II) with the following interpretation:

$$v(y^i) = \inf_{r\in\mathcal{R}} \left\{ \int_0^\infty e^{-(\beta+\lambda)t} \left[\underbrace{c\left(\Phi^r(y^i,t)\right)}_{(I)} + \underbrace{\lambda \sum_{j=1}^{q} Q\left(\Phi^r(y^i,t);\{y^j\}\right) v(y^j)}_{(II)} \right] dt \right\}. \tag{5.4}$$

Optimizing (I) means optimizing the expected discounted cost up to the next jump of the PDMP whereas optimizing (II) means optimizing the total expected discounted cost from right after the next jump of the PDMP until the end of the project life time (precisely until infinity). Unfortunately, there is a link between (I) and (II) which is the dependence of Q on the pre-jump state $\Phi^r(y^i, t)$. Concretely, this means that controlling the process such that the process runs cost optimal up to the next jump, hence optimal for (I), is not

necessarily optimal for the total expected discounted cost over life time. In our example, given the nature of c (see Definition 5.3), optimizing (I) might mean staying always as close to $y^\star$ as possible. In case $Q(y^\star;\{y^q\}) = 1$ this would lead, however, to the highest cost possible right after the next jump.

The nature and properties of Q are thus mainly influencing, besides the shape of c, the characteristics of an optimal policy. In this Section, we will therefore start by assuming the following:

Assumption 5.13. *We assume that there exist $Q^1, \dots, Q^q \geq 0$ with $\sum_{j=1}^q Q^j = 1$ and $Q(y;\{y^j\}) = Q^j$ for all $y \in E_Y$ and $j = 1, \dots, q$.*

The latter assumption thus models a case where the jump transition does not depend on the current pre-jump state of the process. In many applications this is a suitable assumption. In case of the chemicals production example introduced earlier, this might be a suitable assumption whenever jumps in the concentration of the crystal solution in the tank are due to some machine break downs that do not depend on (meaning occure in function of) the current concentration of the crystal solution in the tank.

Under this assumption, we can write (5.3) as

$$v(y^i) = \inf_{r\in\mathcal{R}} \left\{ \int_0^\infty e^{-(\beta+\lambda)t} \underbrace{c\left(\Phi^r(y^i,t)\right)}_{(I)}\, dt \right\} + \int_0^\infty e^{-(\beta+\lambda)t} \underbrace{\lambda \sum_{j=1}^q Q^j v(y^j)}_{(II)}\, dt. \tag{5.5}$$

We recognize that minimization is only done over the integral containing (I). The integral containing (II) can not be influenced by a control decision of the agent. Clearly, in this situation, an optimal policy only tries to minimize one-period cost up to the next jump of the PDMP as the jump transition does not depend on the current pre-jump state of the process. Whatever the agent does, he cannot influence the total expected discounted cost from the next jump onwards. He only can influence the current running cost up to the next jump. Hence, if c has a global minimum point, the policy trying to steer the process state towards this global minimum point as quickly as possible seems to be an optimal policy. It turns out that this intuition is true:

Proposition 5.14. *For the PO-PDMP model of Definition 5.1 with $f_\epsilon = \delta_0$ and $Q_0(y;\cdot) = \delta_y$ $\forall y \in E_Y$ and under Assumption 5.13 it holds:*

a) For cost functions c as of Definition 5.3, there exists a unique optimal policy of „bang-bang“ type.

b) If the cost function c is not strictly convex but having a global minimum in $y^\star$ and c growing on $(y^\star, \infty)$ and c decreasing on $(-\infty, y^\star)$ (both not necessarily strictly) then there exists an optimal policy (not necessarily unique) of „bang-bang“ type.

In both cases, the optimal policy mentioned is deterministic and given by

$$\pi^{0\star}(y,t) := \mathbb{1}_{\{y\leq y^\star\}} \cdot \mathbb{1}_{\{t\leq \frac{y^\star - y}{a_{inc}}\}} \cdot a_{inc} - \mathbb{1}_{\{y>y^\star\}} \cdot \mathbb{1}_{\{t\leq \frac{y-y^\star}{a_{dec}}\}} \cdot a_{dec}.$$

Remark 5.15. *The optimal policy mentioned in the above Proposition is thus a policy, where the agent is acting the following way:*

- *If the post-jump state y is greater than $y^\star$, hence if the concentration in the tank after a jump of the process is greater than the required level of $y^\star$, then, the agent is pumping in water in order to decrease the concentration. The agent selects to pump in water at highest pumping rate possible a_{dec} until the required concentration $y^\star$ is achieved again. Once $y^\star$ achieved, the agent stops all pumping of water or crystal. If a new jump of the process occurs before $y^\star$ is achieved, the agent re-evaluates how to act depending on whether $y \geq y^\star$ or not after the jump.*
- *If the post-jump state y is less than $y^\star$, by the analogous reasoning, the agent is pumping in more crystal at highest possible pump rate a_{inc} in order to increase the concentration until $y^\star$ is achieved. If a new jump of the process occurs before $y^\star$ is achieved, the agent re-evaluates how to act depending on whether $y \geq y^\star$ or not after the jump.*

In order to prove Proposition 5.14, we need the following result which is inspired by Lemma 3.3.2 in [17]. In our version below, however, we skip the convexity requirement for the function f and weaken this to the monotonicity conditions mentioned below. We start with the following Definition:

Definition 5.16. *Let $B, U > 0$ two positive constants. We then define the class $\mathcal{X}_{BU}$ of measurable functions $X : \mathbb{R}^+ \to \mathbb{R}$ by:*

$$X \in \mathcal{X}_{BU} \Leftrightarrow \forall \delta > 0 \ \forall t \geq 0 : -B\delta \leq X(t+\delta) - X(t) \leq U\delta.$$

Lemma 5.17. *Let $f : \mathbb{R} \to \mathbb{R}$ lower semi-continuous and satisfy*

(i) $\exists\ y_0 \in \mathbb{R} : f(y) \geq f(y_0)\ \forall y \neq y_0$

(ii) *f is increasing on (y_0, ∞)*

(iii) *f is decreasing on $(-\infty, y_0)$.*

Further, let $\mathbb{P}$ a probability measure on $\mathbb{R}^+$, absolutely continuous w.r.t. the Lebesgue measure on $\mathbb{R}^+$ and $B, U > 0$. Then, for $x_0 \in \mathbb{R}$, the optimization problem

$$L(X(t); x_0) := \int_0^\infty f\left(X(t)\right) \mathbb{P}(dt) \longrightarrow \min_{X \in \mathcal{X}_{BU}, X(0) = x_0} \tag{5.6}$$

has a solution $X^\star(t) := k(x_0, y_0, t)$, where we define

$$k(x_0, y_0, t) := \mathbb{1}_{\{x_0 \leq y_0\}} \min(x_0 + Ut; y_0) + \mathbb{1}_{\{x_0 > y_0\}} \max(x_0 - Bt; y_0). \tag{5.7}$$

In case f is strictly monotone in (ii) and (iii), this solution is unique.

Proof. W.l.o.g. let $x_0 < y_0$ (use oppsite monotonicity properties of f otherwise). Let $X^\star(t) := k(x_0, y_0, t) = \min(x_0 + Ut; y_0)$. We show (a) a path $X(t)$ reaching a point $y > y_0$ is not better than $X^\star$ and (b) a path $X(t)$ tending slower towards y_0 than $X^\star(t)$ does (and „slower“ includes also tending away from y_0) is not better than $X^\star$. With (a) and (b) $X^\star$ turns out to be optimal.

For (a) let $X \in \mathcal{X}_{BU}$ with $X(0) = x_0$ and $X(t) > y_0$ for some $t > 0$. Define then $X_1(t) := \min(X(t); y_0)$. As f is lower semi-continuous and by definition of $\mathcal{X}_{BU}$, we find the following subset of $\mathbb{R}^+$ with positive measure:

$$\{t \in \mathbb{R}^+ \mid f\left(X_1(t)\right) \leq f\left(X(t)\right)\}. \tag{5.8}$$

Hence, (a) follows as now

$$L(X_1(t), x_0) \leq L(X(t); x_0). \tag{5.9}$$

For (b) let $X_2, X_3 \in \mathcal{X}_{BU}$ with $X_2(0) = X_3(0) = x_0$ and $X_2(t) < X_3(t) \leq y_0$ for all $t \geq 0$, then by monotonicity of integrals

$$L(X_3(t); x_0) \leq L(X_2(t); x_0). \tag{5.10}$$

In case f is strictly monotone in (ii) and (iii), we obtain strict inequalities in (5.8), (5.10), (5.10) and we obtain a unique optimal solution to (5.6). □

Proof (of Proposition 5.14). Under the assumptions of the Proposition, the value function v takes the form (5.5). As the action space is $A = [-a_{dec}, a_{inc}]$ and by definition of the controlled Drift Φ^r, we have $\Phi^r(y^i, t) \in \mathcal{X}_{a_{dec}a_{inc}}$ with $\Phi^r(y^i, 0) = y^i$. Now, c satisfies the requirements on f of the previous lemma and as $\lambda, \beta > 0$, the measure $e^{-(\beta+\lambda)t}dt$ on $\mathbb{R}^+$ can be normalized to a probability measure that is then absolutely continuous to the Lebesgue measure on $\mathbb{R}^+$.

From the previous Lemma we thus get an optimal Drift of $\Phi^r(y^i, t) = k(y^i, y^\star, t)$ and by taking the derivative w.r.t. the time parameter, we obtain the result. □

5.4.2 Models with state-dependent jump transition kernel

The goal of this Section is to provide a sufficient condition for the existence of optimal „bang-bang“ type policies in the case of state-dependent jump transition kernels. As outlined before, the connection between (I) and (II) in (5.4) is the current pre-jump position. We saw, that if Q is such that, very close to the optimum $y^\star$ of c, jumps will lead to states far away from $y^\star$ but if, on the other hand, jumps will lead to positions near $y^\star$ for pre-jump positions not far away from $y^\star$, trying to be always as close to $y^\star$ as possible is not necessarily the best policy.

The idea of this Section is thus to provide a sufficient condition on Q such that still, the optimal policy is to always try to stay as close to $y^\star$ as possible. In a way, the following assumption can be understood as the following property of Q: The further away from $y^\star$ the pre-jump state of the process lies, the higher is the probability of getting post-jump states even further away from $y^\star$. Meaning: controlling the process „away from $y^\star$“ is not only bad for the one period cost portion (I) in (5.4) but also for (II). The hope is thus, that it turns out to be optimal to try to get as close to $y^\star$ as possible under this assumption.

Assumption 5.18. *We assume Q to be a weakly continuous transition kernel from E_Y to $\mathbf{P}(E_Y^0)$ satisfying the following property: For each function $w : E_Y^0 \to \mathbb{R}^+$ with w increasing on $\{y^{\frac{q}{2}+1}, \ldots, y^q\}$ and w decreasing on $\{y^1, \ldots, y^{\frac{q}{2}}\}$, the function*

$$F_{Qw} : E_Y \to \mathbb{R}^+, y \mapsto \sum_{j=1}^{q} Q(y; \{y^j\})w(y^j)$$

has the following properties:

(i) F_{Qw} is increasing on $(y^\star, \infty)$ and

(ii) F_{Qw} is decreasing on $(-\infty, y^\star)$.

Remark 5.19. *Examples of such transition kernels Q are, e.g., kernels of the following form: Take $Q(y^\star, \cdot)$ the uniform distribution on E_Y^0, hence $Q(y^\star; \{y^j\}) = \frac{1}{q}$ for all $j = 1, \dots, q$. Now, as y increases from $y^\star$, keep the distribution on $\{y^1, \dots, y^{\frac{q}{2}}\}$ constant at probability mass $\frac{1}{q}$ on each of those points. For the distribution on $\{y^{\frac{q}{2}+1}, \dots, y^q\}$ we transfer probability mass more and more on y^q as y increases. Meaning, for $y^\star < y \leq y^{\frac{q}{2}+1}$ we get*

$$Q\left(y; \{y^j\}\right) = \begin{cases} \frac{1}{q}, & j \in \{1, \dots, \frac{q}{2}\} \cup \{\frac{q}{2}+2, \dots, q-1\} \\ \frac{1}{q} \cdot \frac{y^{\frac{q}{2}+1} - y}{y^{\frac{q}{2}+1} - y^\star}, & j = \frac{q}{2}+1 \\ \frac{1}{q} + \frac{1}{q} \cdot \frac{y - y^\star}{y^{\frac{q}{2}+1} - y^\star}, & j = q. \end{cases}$$

This means, we take more and more probability mass away from $y^{\frac{q}{2}+1}$ and put it on top of the probability mass of y^q. One can now extend this definition of Q analogously to intervals $(y^k, y^{k+1}]$ for $k = \frac{q}{2}+1, \dots, q-1$ and by the analogous reasoning of then shifting mass to y^1 one can define Q for $y \leq y^\star$.

Now what we presented can be generalized to transition kernels where one does not simply shift probability mass to the extrem points y^1 and y^q but in principal, shifting to any point further away from $y^\star$ than y is adequate.

Such types of transition kernels can appear in applications very often. Imagine that in our chemicals production example, machine break downs causing an increase of the concentration in the tank get more and more likely as the concentration in the tank increases.

Under the above assumption, we thus get the following result:

Proposition 5.20. *For the PO-PDMP model of Definition 5.1 with $f_\epsilon = \delta_0$ and $Q_0(y; \cdot) = \delta_y$ $\forall y \in E_Y$ and under Assumption 5.18 it holds:*

a) For cost functions c as of Definition 5.3, there exists a unique optimal policy of „bang-bang“ type.

b) If the cost function c is not strictly convex but having a global minimum in $y^\star$ and c growing on $(y^\star, \infty)$ and c decreasing on $(-\infty, y^\star)$ (both not necessarily strictly) then there exists an optimal policy (not necessarily unique) of „bang-bang“ type.

In both cases, the optimal policy mentioned is deterministic and given by

$$\pi^{0\star}(y,t) := \mathbb{1}_{\{y \leq y^\star\}} \cdot \mathbb{1}_{\{t \leq \frac{y^\star - y}{a_{inc}}\}} \cdot a_{inc} - \mathbb{1}_{\{y > y^\star\}} \cdot \mathbb{1}_{\{t \leq \frac{y - y^\star}{a_{dec}}\}} \cdot a_{dec}.$$

Proof. We will prove this result following a five steps approach. Remember that we denote the analogous operators to $\mathcal{T}, \mathcal{T}_f$ and H by $\mathcal{L}, \mathcal{L}_f$ and K in the completely observable model, see introduction to Section 5.3. The five steps then are:

(i) We show that $\mathcal{L}\underline{0}$ is decreasing on $\{y^1, \dots, y^{\frac{q}{2}}\}$ and increasing on $\{y^{\frac{q}{2}+1}, \dots, y^q\}$.

(ii) We show that if w as of Assumption 5.18, then $\mathcal{L}w$ is decreasing on $\{y^1, \dots, y^{\frac{q}{2}}\}$ and increasing on $\{y^{\frac{q}{2}+1}, \dots, y^q\}$.

(iii) By induction, we then get $\forall k \geq 1 : \mathcal{L}^k\underline{0}$ is decreasing on $\{y^1, \dots, y^{\frac{q}{2}}\}$ and increasing on $\{y^{\frac{q}{2}+1}, \dots, y^q\}$.

(iv) By pointwise convergence of $\mathcal{L}^k\underline{0}$ (see Theorem 2.58 and for adapted version to completely observable PDMP problems, see Forwick [34], Theorem 2.3.9), we get

$$v = \lim_{k\to\infty} \mathcal{L}^k\underline{0}$$

is decreasing on $\{y^1, \dots, y^{\frac{q}{2}}\}$ and increasing on $\{y^{\frac{q}{2}+1}, \dots, y^q\}$.

(v) Finally, from (5.4) with its parts (I) and (II) we get the result by applying Lemma 5.17: Actually, the function

$$y \mapsto c\,(y) + \lambda \sum_{j=1}^{q} Q\left(y; \{y^j\}\right) v(y^j)$$

is decreasing on $(-\infty; y^\star)$ and increasing on $(y^\star; \infty)$ as this property is true for c as assumed in the statement of the proposition and as the same property is true for $y \mapsto \lambda \sum_{j=1}^{q} Q\,(y; \{y^j\})\, v(y^j)$ because v satisfies the requirements of Assumption 5.18 because of (iv).

The result now follows by the same reasoning as at the end of proof of Proposition 5.14. Remains to show (i) and (ii).

For (i), we get by definition of $\mathcal{L}$ for $y^i \in E_Y^0$:

$$(\mathcal{L}\underline{0})\,(y^i) = \inf_{r\in\mathcal{R}} \left\{ \int_0^\infty e^{-(\beta+\lambda)t} c\left(\Phi^r(y^i, t)\right)\, dt \right\}.$$

Now by Lemma 5.17, we get an optimal Drift of $\Phi^{r^\star}(y^i, t) = k(y^i, y^\star, t)$. Now let $y^n, y^{n+1} \in \{y^{\frac{q}{2}+1}, \dots, y^q\}$ (the case $y^n, y^{n+1} \in \{y^1, \dots, y^{\frac{q}{2}}\}$ analogously). By definition, $k(y^n, y^\star, t)$ is tending towards $y^\star$ with maximum allowed „speed“ a_{dec}, same is true for $k(y^{n+1}, y^\star, t)$. But as y^{n+1} is further away from y^n, it will always hold $k(y^{n+1}, y^\star, t) - y^\star \geq k(y^n, y^\star, t) - y^\star$ and hence

$$c\left(k(y^{n+1}, y^\star, t)\right) \geq c\,(k(y^n, y^\star, t))\ \ \forall t \geq 0.$$

By monotonicity of integrals, we thus get (i).

For (ii) we apply the same reasoning as in (v) (here, w plays the role of v in (v)) to derive from Lemma 5.17 the existence of an optimal Drift. Now apply the analogous reasoning as for (i) to the integrand $y \mapsto c\,(y) + \lambda \sum_{j=1}^{q} Q\,(y; \{y^j\})\, w(y^j)$. □

5.5 Optimal policies in partially observable case

In this Section we come back to the partially observable version of the application example introduced at the beginning of this Chapter. We thus assume a noise density $f\epsilon$ that is not concentrated to one point and we also assume an initial conditional distribution of Y_0 that is not necessarily a point mass. Hence, we get a value function of the derived filter model as developed in (5.2) and again, we recognize two summands (I) and (II):

$$J'(\rho) = \inf_{r\in\mathcal{R}} \{(I) + (II)\}, \tag{5.11}$$

where

$$(I) = \int_0^\infty e^{-(\beta+\lambda)s'} \sum_{i=1}^q \rho^i c\left(\Phi^r(y^i, s')\right)\, ds',$$

$$(II) = \lambda \int_0^\infty e^{-(\beta+\lambda)s'} \sum_{j=1}^q \sum_{i=1}^q \rho^i Q\left(\Phi^r(y^i, s'); \{y^j\}\right) \int_{\mathbb{R}} f_\epsilon\left(x' - \psi(y^j)\right) J'\left(\chi(\rho, x', s', r)\right) dx' ds'.$$

The difference to the completely observable case is obvious: We get a sum over the current conditional distribution ρ, more precisely over its components ρ^i. For (I) this is not so much of a difference compared to the completely observable case as we will see in the next Section. For (II), besides the sum over ρ^i, we also get a dependence of χ on r which makes the treatment of (II) more complex as in the completely observable case where r only appeared in the argument of Q via Φ^r.

The plan for this Section is thus to first analyze partially observable models with state-independent jump transition kernel in 5.5.1. Thereafter, we turn to the general case of models with state-dependent jump transition kernel in 5.5.2.

5.5.1 Models with state-independent jump transition kernel

We start by analyzing models with state-independent jump transition kernels, hence we take again Assumption 5.13. The first observation is that the filter χ does no longer depend on the relaxed control r under Assumption 5.13.

Corollary 5.21. *Under Assumption 5.13 we have*

$$\chi_i^j(\rho, x, s, r) = \chi_i^j(\rho, x, s) = \rho^i e^{-\lambda s} \lambda Q^j f_\epsilon(x - y^j),$$
$$\chi^j(\rho, x, s, r) = \chi^j(x) = \frac{f_\epsilon(x - y^j) Q^j}{\sum_{k=1}^q Q^k f\epsilon(x - y^k)},$$

meaning that χ_i^j does not depend on $r \in \mathcal{R}$ and that χ^j, hence χ, does only depend on the noisy measurement x.

Proof. The statement about χ_i^j follows directly from Definition 2.30 as we assume λ constant in this application. For χ^j, remember the definition of $\overline{\chi}$ as

$$\overline{\chi}(\rho, x, s, r) := \sum_{k=1}^q \sum_{l=1}^q \chi_k^l(\rho, x, s, r) = \sum_{k=1}^q \sum_{l=1}^q \rho^k e^{-\lambda s} \lambda Q^l f_\epsilon(x - y^l) = e^{-\lambda s} \lambda \sum_{l=1}^q Q^l f_\epsilon(x - y^l),$$

where we used that $\sum_{k=1}^q \rho^k = 1$.

By definition of χ^j, we then obtain:

$$\chi^j(\rho, x, s, r) = \frac{1}{\overline{\chi}(\rho, x, s, r)} \sum_{i=1}^q \chi_i^j(\rho, x, s, r) = \frac{e^{-\lambda s} \lambda Q^j f_\epsilon(x - y^j) \sum_{i=1}^q \rho^i}{e^{-\lambda s} \lambda \sum_{k=1}^q Q^k f_\epsilon(x - y^l)},$$

and the result follows as, again, $\sum_{i=1}^q \rho^i = 1$. □

Based on the previous Corollary, one recognizes immediately that (II) in (5.11) does not depend on $r \in \mathcal{R}$ under Assumption 5.13. The optimality equation for J' under Assump-

tion 5.13 thus becomes:

$$J'(\rho) = \inf_{r\in\mathcal{R}} \left\{ \int_0^\infty e^{-(\beta+\lambda)s'} \sum_{i=1}^q \rho^i c\left(\Phi^r(y^i, s')\right)\, ds' \right\} + \lambda \int_0^\infty e^{-(\beta+\lambda)s'} ds' \sum_{j=1}^q Q^j \int_{\mathbb{R}} f_\epsilon\left(x' - \psi(y^j)\right) J'\left(\chi(x')\right) dx'.$$

We find again a situation where the agent can only try to minimize one period cost up to the next jump. This time, however, we get a convex combination of convex functions as the sum over ρ appears in the integrand. In preparation for the final result of this Section, we need the following:

Lemma 5.22. *For cost functions c as of Definition 5.3 it holds: Let $\rho \in \mathbf{P}(E_Y^0)$ and $y^i \in E_Y^0$ for $i = 1, \dots, q$, then the function*

$$\mathbb{R} \ni y \mapsto \sum_{i=1}^q \rho^i c(y^i + y) \in \mathbb{R}^+$$

is strictly convex and has a unique global minimum $y^{\star\star} \in (y^{\frac{q}{2}} - y^q, y^{\frac{q}{2}+1} - y^1)$.

Proof. As c is strictly convex and $\rho^i \geq 0$ for all $i = 1, \dots, q$ with $\sum_{i=1}^q \rho^i = 1$, the function $y \mapsto \sum_{i=1}^q \rho^i c(y^i + y)$ is strictly convex. Further, as the strictly convex function c has its global minimum in $y^\star := \frac{y^{\frac{q}{2}} + y^{\frac{q}{2}+1}}{2}$, we know that c is decreasing on $(-\infty, y^\star) \supset (-\infty, y^{\frac{q}{2}})$ and increasing on $(y^\star, \infty) \supset (y^{\frac{q}{2}+1}, \infty)$. Hence, $y \mapsto c(y^i + y)$ is decreasing on $(-\infty, y^{\frac{q}{2}} - y^q)$ and increasing on $(y^{\frac{q}{2}+1} - y^1, \infty)$ for all $i = 1, \dots, q$. Now, by strict convexity of $y \mapsto \sum_{i=1}^q \rho^i c(y^i + y)$, the existence of a unique global minimum in the stated interval follows. □

Corollary 5.23. *If c is only convex but not strictly convex and has a global minimum in $y^\star := \frac{y^{\frac{q}{2}} + y^{\frac{q}{2}+1}}{2}$, the statement of the previous lemma still holds but $y^{\star\star}$ is not unique then.*

With this results, we can now apply the analogous reasoning as for the completely observable case to prove the main result of this Section:

Proposition 5.24. *For the PO-PDMP model of Definition 5.1 under Assumption 5.13 it holds:*

a) For cost functions c as of Definition 5.3, there exists a unique optimal policy of „bang-bang“ type.

b) If the cost function c is not strictly convex but having a global minimum in $y^\star$ and c growing on $(y^\star, \infty)$ and c decreasing on $(-\infty, y^\star)$ (both not necessarily strictly) then there exists an optimal policy (not necessarily unique) of „bang-bang“ type.

In both cases, the optimal policy mentioned is deterministic and given by

$$\pi_n^\star(h_n, t) := \mathbb{1}_{\{0 \leq y^{\star\star}(h_n)\}} \cdot \mathbb{1}_{\{t \leq \frac{y^{\star\star}(h_n)}{a_{inc}}\}} \cdot a_{inc} - \mathbb{1}_{\{0 > y^{\star\star}(h_n)\}} \cdot \mathbb{1}_{\{t \leq \frac{-y^{\star\star}(h_n)}{a_{dec}}\}} \cdot a_{dec},$$

where $y^{\star\star}(h_n)$ is the global minimum mentioned in Lemma 5.22 (for a)) resp. in Corollary 5.23 (for b)) for $\rho = \mu_n(h_n)$, i.e. ρ the conditional distribution calculated by iterating χ based on the observed history h_n.

Proof. By the definition of Φ^r in Definition 5.1(v) it holds: $\Phi^r(y^i, t) = y^i + \Phi^r(0, t)$ for all $t \geq 0$. Thus, we get

$$\sum_{i=1}^{q} \rho^i c\left(\Phi^r(y^i, t)\right) = \sum_{i=1}^{q} \rho^i c\left(y^i + \Phi^r(0, t)\right).$$

Now, based on Lemma 5.22 and Corollary 5.23, we can apply Lemma 5.17 and get an optimal drift of

$$\Phi^r(0, t) = k(0, y^{\star\star}(\rho), t).$$

Taking the time derivative of this optimal drift leads to the result. Strictly speaking, this leads first to an optimal decision rule for the derived filtered problem of

$$f^\star(\rho)(t) := \mathbb{1}_{\{0 \leq y^{\star\star}(\rho)\}} \cdot \mathbb{1}_{\{t \leq \frac{y^{\star\star}(\rho)}{a_{inc}}\}} \cdot a_{inc} - \mathbb{1}_{\{0 > y^{\star\star}(\rho)\}} \cdot \mathbb{1}_{\{t \leq \frac{-y^{\star\star}(\rho)}{a_{dec}}\}} \cdot a_{dec},$$

then by Lemma 2.38 we get

$$\pi_n^{\star D}(h_n)(t) = f^\star(\mu_n(h_n))(t),$$

as optimal policy for the pseudo-embedded process and, finally by the Correspondence Theorem 2.11 , we finally get

$$\pi_n^\star(h_n, t) = \pi_n^{\star D}(h_n)(t)$$

as optimal policy for the PO-PDMP. □

With the latter result, we also get existence of an optimal policy of „bang-bang“ type in the partially observable case whenever Q is not state-dependent. However, note that the optimal policy is depending on the conditional distribution ρ as $y^{\star\star}$ is a function of ρ. Hence, the point towards which an optimal policy is steering is different for different observations of ρ. This is a difference to the completely observable case where an optimal policy will always steer towards the global minimum $y^\star$ of the running cost function c.

5.5.2 Models with state-dependent jump transition kernel

We turn now to the case of state-dependent jump transition kernels. We start with the discussion of a very concrete application example where we can achieve a characterization of an optimal policy of bang-bang type. A deeper analysis of properties of the filter is required to solve this problem. In a second step, we then provide an outlook on how to characterize and find optimal policies for a general PO-PDMP problem with state-dependent jump transition kernel.

5.5.2.1 A three states example

In this Paragraph, we illustrate an approach how to determine an optimal policy for a partially observable PDMP control problem. To do so, we assume the following model for the rest of this Paragraph:

Assumption 5.25. *We consider a model as of Definition 5.1 where we specify the following:*

(i) The action space is assumed to be $A := [-1; 1]$, i.e. we set $a_{dec} := a_{inc} := 1$ in Definition 5.1.

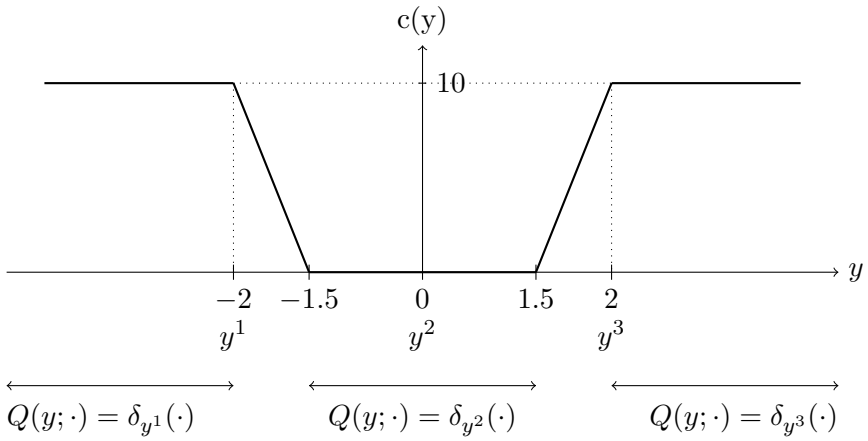


Figure 5.1: Cost function and transition kernel in concrete application example

(ii) *The set of possible post jump states is assumed to be* $E_Y^0 := \{-2, 0, 2\}$ *and we set* $y^1 := -2, y^2 := 0$ *and* $y^3 := 2$. *This is a slight modification of the model of Definition 5.1 where an even number of possible post-jump states was assumed. However, this does not influence the following results.*

(iii) *We set* $\lambda := \beta := 1$.

(iv) *The jump transition kernel* Q *is specified as follows (see also Figure 5.1):*

$$Q(y;\cdot) := \begin{cases} \delta_{y^1}(\cdot), & y \leq -2 \\ \frac{\frac{3}{2}+y}{\frac{-1}{2}} \cdot \delta_{y^1}(\cdot) + \frac{2+y}{\frac{1}{2}} \cdot \delta_{y^2}(\cdot), & -2 < y < -\frac{3}{2} \\ \delta_{y^2}(\cdot), & -\frac{3}{2} \leq y \leq \frac{3}{2} \\ \frac{2-y}{\frac{1}{2}} \cdot \delta_{y^2}(\cdot) + \frac{y-\frac{3}{2}}{\frac{1}{2}} \cdot \delta_{y^3}(\cdot), & \frac{3}{2} < y < 2 \\ \delta_{y^3}(\cdot), & 2 \leq y. \end{cases}$$

(v) *The cost function is no longer assumed to be strictly convex. We assume the following cost function throughout this example (see also Figure 5.1):*

$$c(y) := \begin{cases} 10, & y \leq -2 \\ 10 - 20(y+2), & -2 < y < -\frac{3}{2} \\ 0, & -\frac{3}{2} \leq y \leq \frac{3}{2} \\ 20(y - \frac{3}{2}), & \frac{3}{2} < y < 2 \\ 10, & y \geq 2. \end{cases}$$

The main result of this Paragraph is the following characterization of an optimal „bang-bang“-type policy for the optimization problem defined in Definition 5.5:

Theorem 5.26. *For a PO-PDMP control model as defined in Definition 5.1 and under the precisions of the model made in Assumption 5.25, an optimal policy for the*

optimization problem of Definition 5.5 is given by the following „bang-bang"-type policy $\pi = (\pi_0, \pi_1, \dots) \in \Pi^P$*, where we set, for* $n \in \mathbb{N}$ *and* $h_n \in \mathcal{H}_n$*:*

$$\pi_n(h_n, t) := \mathbb{1}_{\{\mu_n^1(h_n) \geq \mu_n^3(h_n)\}} \cdot \mathbb{1}_{\{t \leq \frac{1}{2}\}} - \mathbb{1}_{\{\mu_n^1(h_n) < \mu_n^3(h_n)\}} \cdot \mathbb{1}_{\{t \leq \frac{1}{2}\}}. \tag{5.12}$$

We use here the earlier introduced notation $\mu_n^i(h_n)$ *for* $\mu_n(h_n)(\{y^i\})$*, i.e. the conditional probability of seeing a post-jump state* y^i *at the n-th jump of the process given the observed history* h_n*. Remember that* $\mu_n(h_n)$ *is the recursively, via iteration of* χ*, calculated conditional distribution on* $\mathbf{P}(E_Y^0)$*.*

In order to prove this theorem, we will follow the approach outlined as follows:

1) We give an optimal policy for the one-step problem of the derived filtered process, i.e. we present such a policy and prove that it is optimal.

2) We prove that the one-step-minimal-cost-function $\mathcal{T}\underline{0} : \mathbf{P}(E_Y^0) \to \mathbb{R}^+$ has a set of important properties.

3) We show by induction that the value function J' of the above mentioned optimization problem has the same properties as found for $\mathcal{T}\underline{0}$ in step 2).

Based on the fixed point equation $\mathcal{T}J' = J'$, we can then prove Theorem 5.26 leveraging the properties of J' found in step 3).

Step 1: Finding an optimal policy for the one-step optimization problem:
The minimal cost for the one-step problem was derived, for $\rho = (\rho^1, \rho^2, \rho^3) \in \mathbf{P}(E_Y^0)$, as

$$\left(\mathcal{T}\underline{0}\right)(\rho) = \inf_{r \in \mathcal{R}} \left\{ \int_0^\infty e^{-2t} \sum_{i=1}^3 \rho^i c\left(\Phi^r(y^i, t)\right) \, dt \right\}, \tag{5.13}$$

see also Cost Iteration in Proposition 2.50, Corollary 2.53 and Lemma 3.6. Note further, that here, we have $\lambda = \beta = 1$.

As in the previous Sections of this Chapter, we will now first analyze the sum of cost functions appearing in the above integral:

Definition 5.27. *For* $\rho = (\rho^1, \rho^2, \rho^3) \in \mathbf{P}(E_Y^0)$*, with* $\rho^i = \mathbb{P}(\{y^i\})$ *for* $i = 1, 2, 3$*, we define*

$$C_\rho(y) := \sum_{i=1}^3 \rho^i c(y^i + y), \qquad\qquad y \in E_Y.$$

In order to simplify notations, we denote by $l[(x_1, y_1); (x_2, y_2)](x) := y_1 \cdot \frac{x_2 - x}{x_2 - x_1} + y_2 \cdot \frac{x - x_1}{x_2 - x_1}$ the linear function describing the shortest path between the points (x_1, y_1) and (x_2, y_2) of the real plane.

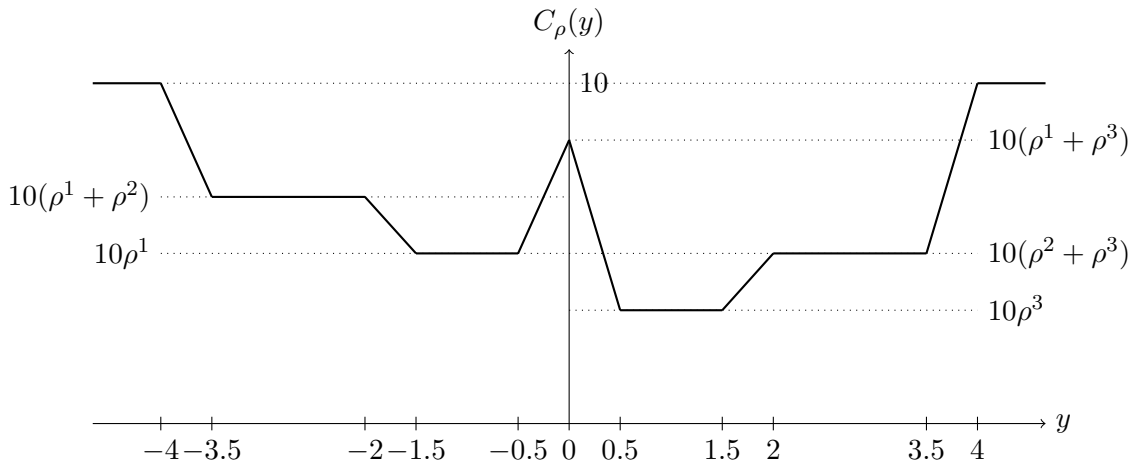


Figure 5.2: Graph of function C_ρ for $\rho = \left(\frac{1}{2}, \frac{5}{30}, \frac{1}{3}\right)$, i.e. for case $\rho^1 > \rho^3$.

Lemma 5.28. *Let $\rho \in \mathbf{P}(E_Y^0)$, then it holds (see also Figure 5.2):*

$$C_\rho(y) = \begin{cases} 10, & y \le -4 \\ l\left[(-4,10);(-\frac{7}{2},10(\rho^1+\rho^2))\right](y), & -4 < y < -\frac{7}{2} \\ 10(\rho^1+\rho^2), & -\frac{7}{2} \le y \le -2 \\ l\left[(-2,10(\rho^1+\rho^2));(-\frac{3}{2},10\rho_1)\right](y), & -2 < y < -\frac{3}{2} \\ 10\rho^1, & -\frac{3}{2} \le y \le -\frac{1}{2} \\ l\left[(-\frac{1}{2},10\rho^1);(0,10(\rho^1+\rho^3))\right](y), & -\frac{1}{2} < y < 0 \\ l\left[(0,10(\rho^1+\rho^3));(\frac{1}{2};10\rho^3)\right](y), & 0 \le y < \frac{1}{2} \\ 10\rho^3, & \frac{1}{2} \le y \le \frac{3}{2} \\ l\left[(\frac{3}{2},10\rho^3);(2,10(\rho^2+\rho^3))\right](y), & \frac{3}{2} < y < 2 \\ 10(\rho^2+\rho^3), & 2 \le y \le \frac{7}{2} \\ l\left[(\frac{7}{2},10(\rho^2+\rho^3));(4,10)\right](y), & \frac{7}{2} < y < 4 \\ 10, & y \ge 4. \end{cases}$$

Proof. Follows directly from the definition of the cost function c in this example, see Assumption 5.25, part (v). □

Remark 5.29. *In Figure 5.2, the graph of $C_\rho(y)$ is illustrated for the case $\rho^1 > \rho^3$.*

Having understood the function C_ρ, we can now give an optimal policy for the one-step optimization problem of the derived filtered process in our example:

Proposition 5.30. *The one-step minimal cost for $\rho \in \mathbf{P}(E_Y^0)$ is given by*

$$(\mathcal{T}\underline{0})(\rho) = \int_0^\infty e^{-2t} \sum_{i=1}^3 \rho^i c\left(y^i + \Phi^{u^\star}(0,t)\right)\, dt,$$

where an optimal deterministic policy $u^\star \in \mathcal{U}$ (see also Remark 5.12 to recap the Definition of a deterministic policy) is given by

$$u^\star(t) := \mathbb{1}_{\{\rho^1 \geq \rho^3\}} \cdot \mathbb{1}_{\{t \leq \frac{1}{2}\}} - \mathbb{1}_{\{\rho^1 < \rho^3\}} \cdot \mathbb{1}_{\{t \leq \frac{1}{2}\}}.$$

This is not a unique optimal policy as one may recognize when looking at the graph of C_ρ.

Proof. Assume, w.l.o.g., $\rho^1 \geq \rho^3$, otherwise analogous reasoning by symmetry of the cost function and as $|y^1| = |y^3|$.

For $\rho^1 \geq \rho^3$, the function C_ρ has a global minimum in $y^\star := \frac{1}{2}$ (follows from previous Lemma, see also Figure 5.2). Furthermore, C_ρ has a local minimum in $y' := -\frac{1}{2}$. However, as $\rho^1 \geq \rho^3$ it follows $C_\rho(y') \geq C_\rho(y^\star)$ and hence, as $|y^\star| = |y'|$ and as $a_{dec} = a_{inc}$, there is no benefit in conidering the „left branch“ of the graph of C_ρ for $y \leq 0$.

Now considering the „right branch“ of the graph of C_ρ for $y \geq 0$, we find: C_ρ is monotone decreasing on $[0, \frac{1}{2})$ and monotone increasing on $(\frac{1}{2}, \infty)$. By (an adapted version of) Lemma 5.17, an optimal drift, realizing the minimal one-step cost, is given by $\Phi^{u^\star}(y^i, t) = y^i + k(0, \frac{1}{2}, t)$, where we use the definition also presented in Lemma 5.17:

$$k(0, \frac{1}{2}, t) := \mathbb{1}_{\{0 \leq \frac{1}{2}\}} \min(0 + t; \frac{1}{2}) + \mathbb{1}_{\{0 > \frac{1}{2}\}} \max(x_0 - t; \frac{1}{2}) = \min(t; \frac{1}{2}).$$

Differentiation w.r.t. the time parameter leads to the result for the case $\rho^1 \geq \rho^3$. Analogous reasoning for the case $\rho^1 < \rho^3$, where we then get $k(0, -\frac{1}{2}, t) := \mathbb{1}_{\{0 \leq -\frac{1}{2}\}} \min(0 + t; -\frac{1}{2}) + \mathbb{1}_{\{0 > -\frac{1}{2}\}} \max(x_0 - t; -\frac{1}{2})$. □

This completes step 1) of our approach and we now turn to the analysis of properties of $\mathcal{T}\underline{0}$ as a function defined on $\mathbf{P}(E_Y^0)$.

Step 2: Analyzing the properties of the one-step minimal cost function $\mathcal{T}\underline{0}$: Knowing an optimal policy for the one-step optimization problem, a few Corollaries can be shown:

Corollary 5.31. *For $\rho \in \mathbf{P}(E_Y^0)$ it holds:*

$$(\mathcal{T}\underline{0})(\rho) = \int_0^{\frac{1}{2}} e^{-2t} \left\{ \mathbb{1}_{\{\rho^1 \geq \rho^3\}} \cdot \left[\rho^1 \; c(t-2) + 10\rho^3\right] + \mathbb{1}_{\{\rho^1 < \rho^3\}} \cdot \left[\rho^3 \; c(2-t) + 10\rho^1\right] \right\} dt$$
$$+ \int_{\frac{1}{2}}^\infty e^{-2t} \left\{ \mathbb{1}_{\{\rho^1 \geq \rho^3\}} \cdot 10 \cdot \rho^3 + \mathbb{1}_{\{\rho^1 < \rho^3\}} \cdot 10 \cdot \rho^1 \right\} dt.$$

Proof. Follows from previous Proposition: Applying the optimal policy stated in previous proposition leads to a drift of $\Phi^{u^\star}(y^i, t) = y^i + k(0, \frac{1}{2}, t)$ in the case of $\rho^1 \geq \rho^3$ and to a drift of $\Phi^{u^\star}(y^i, t) = y^i + k(0, -\frac{1}{2}, t)$ if $\rho^1 < \rho^3$. Now, as $c(y) = 0$ for $y \in [-1.5, 1.5]$ we get $c(y^2 + k(0, y^\star, t)) = 0$ for both cases, $y^\star = \frac{1}{2}$ and $y^\star = -\frac{1}{2}$. The result then follows putting in $y^1 = -2$ and $y^3 = 2$ according to their definitions. □

Corollary 5.32. *Writing δ_y for the point mass in point $y \in E_Y$ and setting $\rho_0 := \frac{1}{2}\delta_{y^1} + \frac{1}{2}\delta_{y^2} \in \mathbf{P}(E_Y^0)$, we get the following results:*

(i) $(\mathcal{T}\underline{0})(\delta_{y^2}) = 0$

(ii) $(\mathcal{T}\underline{0})(\delta_{y^1}) = (\mathcal{T}\underline{0})(\delta_{y^3}) > 0$

(iii) $(\mathcal{T}\underline{0})(\rho_0) > (\mathcal{T}\underline{0})(\delta_{y^1})$ *„Information inequality“.*

Proof. (i) is clear from previous lemma as ρ^2 does not appear in the term for $(\mathcal{T}\underline{0})(\rho)$ there. Part (ii) is clear by symmetrie of c (for the equality) and by strict positiveness of the second integral. Part (iii) clear as second integral is zero for $\rho = \delta_{y^1}$ but not for $\rho = \rho_0$ and in first integral we have $\frac{1}{2}c(t-2) + \frac{1}{2}\cdot 10 \geq c(t-2)$ for all $t \in [0, \frac{1}{2}]$. □

As ρ^1, ρ^2 and ρ^3 only appear in first order in the expression for $(\mathcal{T}\underline{0})(\rho)$ in Corollary 5.31, knowing (i),(ii) and (iii) of the previous Corollary is enough to draft the graph of $\mathcal{T}\underline{0}$. All has to be linear and we have the significant points and know about the relation of the function values of $\mathcal{T}\underline{0}$ in these points. But we can also express $\mathcal{T}\underline{0}$ as a function of only two variables as it holds $\rho^1 + \rho^2 + \rho^3 = 1$ for all $\rho \in \mathbf{P}(E_Y^0)$.

Corollary 5.33. *For $\rho = (\rho^1, \rho^2, \rho^2) \in \mathbf{P}(E_Y^0)$, it holds:*

$$(\mathcal{T}\underline{0})(\rho) = F(\rho^1, \rho^2),$$

where we define the function F by

$$\begin{aligned} F(\rho^1,\rho^2) = \int_0^{\frac{1}{2}} e^{-2t} \Big\{ & \mathbb{1}_{\{\rho^1 \geq (1-\rho^1-\rho^2)\}} \cdot \Big[\rho^1\ c(t-2) + 10(1-\rho^1-\rho^2)\Big] \\ & + \mathbb{1}_{\{\rho^1 < (1-\rho^1-\rho^2)\}} \cdot \Big[(1-\rho^1-\rho^2)\ c(2-t) + 10\rho^1\Big] \Big\}\ dt \\ + \int_{\frac{1}{2}}^{\infty} e^{-2t} \Big\{ & \mathbb{1}_{\{\rho^1 \geq (1-\rho^1-\rho^2)\}} \cdot 10 \cdot (1-\rho^1-\rho^2) + \mathbb{1}_{\{\rho^1 < (1-\rho^1-\rho^2)\}} \cdot 10 \cdot \rho^1 \Big\}\ dt. \end{aligned}$$

Proof. Follows from Corollary 5.31 by setting $\rho^3 = 1 - \rho^1 - \rho^2$. □

Remark 5.34. *A plot of the graph of F on the domain $\rho^1 \in [0;1], \rho^2 \in [0; 1-\rho^1]$ is given in Figure 5.3. As $\rho^1 + \rho^2 + \rho^3 = 1$ this domain covers all $\rho \in \mathbf{P}(E_Y^0)$. The probability mass ρ^3 is given by $\rho^3 = 1 - \rho^1 - \rho^2$.*

Lemma 5.35. *For $\rho = (\rho^1, \rho^2, \rho^3) \in \mathbf{P}(E_Y^0)$ and $\tilde{\rho} = (\tilde{\rho}^1, \tilde{\rho}^2, \tilde{\rho}^3) \in \mathbf{P}(E_Y^0)$ it holds:*

$$\Big(\tilde{\rho}^2 \geq \rho^2 \wedge \tilde{\rho}^1 = 0\Big) \vee \Big(\tilde{\rho}^2 \geq \rho^2 \wedge \tilde{\rho}^3 = 0\Big) \Longrightarrow (\mathcal{T}\underline{0})(\tilde{\rho}) \leq (\mathcal{T}\underline{0})(\rho).$$

Proof. Follows from a technical proof analyzing the gradient field of the function F. This proof is known to the author but omitted here as it is very technical and lengthy. However, from our findings in Corollary 5.32 and by the remark thereafter, one can construct the graph of F. A visualization of the graph is provided in Figure 5.3. From this graph, the statement follows. □

The two important properties of $\mathcal{T}\underline{0}$ to remember from this step of our approach are thus:

a) $(\mathcal{T}\underline{0})(\delta_{y^2}) = 0$

b) $(\tilde{\rho}^2 \geq \rho^2 \wedge \tilde{\rho}^1 = 0) \vee (\tilde{\rho}^2 \geq \rho^2 \wedge \tilde{\rho}^3 = 0) \Longrightarrow (\mathcal{T}\underline{0})(\tilde{\rho}) \leq (\mathcal{T}\underline{0})(\rho)$.

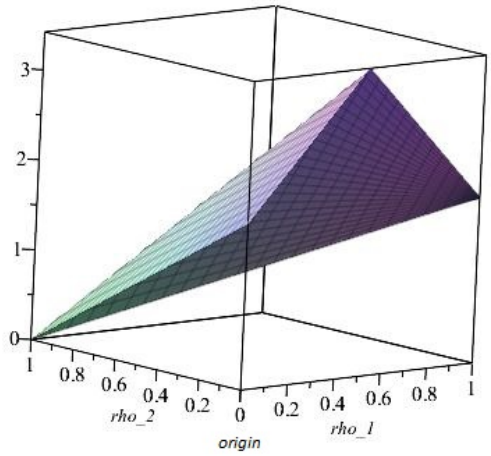


Figure 5.3: Graph of function F from Corollary 5.33, plotted for $\rho^1 \in [0;1], \rho^2 \in [0; 1-\rho^1]$

Step 3: Analyzing the properties of the value function J':
We turn now to the analysis of properties of the value function J' and the main result of this step of our approach is:

Proposition 5.36. *The value function J' satisfies:*

a) $J'(\delta_{y^2}) = 0$

and for $\rho, \tilde{\rho} \in \mathbf{P}(E_Y^0)$ we have

b) $(\tilde{\rho}^2 \geq \rho^2 \wedge \tilde{\rho}^1 = 0) \vee (\tilde{\rho}^2 \geq \rho^2 \wedge \tilde{\rho}^3 = 0) \Longrightarrow J'(\tilde{\rho}) \leq J'(\rho)$.

For the proof of this Proposition, we need to better understand the filter equation for χ. The important result here is the following:

Lemma 5.37. *For $\rho^1 \geq \rho^3$ and under the optimal policy $u^\star$ of Proposition 5.30, we get for $x \in \mathbb{R}$ and $s \geq \frac{1}{2}$:*

$$\begin{aligned}
\chi^1(\rho, x, s, u^\star) &= 0 \\
\chi^2(\rho, x, s, u^\star) &= \frac{(\rho^1+\rho^2) f_\epsilon(x-y^2)}{(\rho^1+\rho^2) f_\epsilon(x-y^2) + \rho^3 f_\epsilon(x-y^3)} \\
\chi^3(\rho, x, s, u^\star) &= \frac{\rho^3 f_\epsilon(x-y^3)}{(\rho^1+\rho^2) f_\epsilon(x-y^2) + \rho^3 f_\epsilon(x-y^3)}.
\end{aligned}$$

An analogous result (by symmetry) follows for the case $\rho^1 < \rho^3$. Furthermore, again for $\rho^1 \geq \rho^3$, and under policy u^0 defined by $u_t^0 = 0 \in A$, i.e. the executed control action is zero at every point in time, we get for $s \geq 0$ and $x \in \mathbb{R}$:

$$\begin{aligned}
\chi^1(\rho, x, s, u^0) &= \frac{\rho^1 f_\epsilon(x-y^1)}{\sum_{i=1}^3 \rho^i f_\epsilon(x-y^i)} \\
\chi^2(\rho, x, s, u^0) &= \frac{\rho^2 f_\epsilon(x-y^2)}{\sum_{i=1}^3 \rho^i f_\epsilon(x-y^i)} \\
\chi^3(\rho, x, s, u^0) &= \frac{\rho^3 f_\epsilon(x-y^3)}{\sum_{i=1}^3 \rho^i f_\epsilon(x-y^i)}.
\end{aligned}$$

An analogous result (by symmetry) follows for the case $\rho^1 < \rho^3$.

Proof. Follows from the definition of χ, see Definition 2.30 and the Defintion of the jump transition kernel Q here in this example. Note that we have λ constant. By the same reasoning as in Corollary 5.21, χ^j does not depend on λ. Furthermore, remember that under $u^\star$ for $s \geq \frac{1}{2}$, we have already moved the probability mass that initially was in y^1 into the area where Q leads to a jump to y^2 almost surely. Under control u^0, probability masses are transferred from y^i to y^i as such is Q but then, according to the filter equation, the noise modeled by $f\epsilon$ has to be considered. This leads to the expressions for $\chi(\rho, x, s, u^0)$ above. □

Corollary 5.38. *For $s \geq \frac{1}{2}$ and for $u^\star$ and u^0 as in previous Lemma, it holds for all $x \in \mathbb{R}$:*

$$\chi^2(\rho, x, s, u^\star) \geq \chi^2(\rho, x, s, u^0).$$

Proof. Follows from previous Lemma as

$$\begin{aligned}
&\chi^2(\rho, x, s, u^\star) - \chi^2(\rho, x, s, u^0) \\
=&\frac{(\rho^1 + \rho^2) f_\epsilon(x - y^2)}{(\rho^1 + \rho^2) f_\epsilon(x - y^2) + \rho^3 f_\epsilon(x - y^3)} - \frac{\rho^2 f_\epsilon(x - y^2)}{\sum_{i=1}^3 \rho^i f_\epsilon(x - y^i)} \\
=&\frac{\rho^1 f_\epsilon(x - y^2) \left[(\rho^1 + \rho^2) f_\epsilon(x - y^1) + \rho^3 f_\epsilon(x - y^3)\right]}{\left[\sum_{i=1}^3 \rho^i f_\epsilon(x - y^i)\right] \left[(\rho^1 + \rho^2) f_\epsilon(x - y^2) + \rho^3 f_\epsilon(x - y^3)\right]}
\end{aligned}$$

Now the result follows as there are only sums and products of non-negative expressions. From Lemma 5.37 we further get another immediate consequence: □

Corollary 5.39. $\chi^2(\delta_{y^2}, x, s, u^\star) = \chi^2(\delta_{y^2}, x, s, u^0) = \delta_{y^2}$.

We can now prove the main result of this step of our approach:

Proof (proof of Proposition 5.36). We will follow the analogous approach as for the proof of Proposition 5.20 for the completely observable case. The three steps here are now:

(i) We show that for $w : \mathbf{P}(E_Y^0) \to \mathbb{R}^+$ with w satisfies properties a) and b) of Proposition 5.36 we have again $\mathcal{T}w$ satisfying properties a) and b).

(ii) By induction it follows that $\mathcal{T}^k \underline{0}$ satisfies properties a) and b) for all $k \geq 0$.

(iii) As $J' = \lim_{k\to\infty} \mathcal{T}^k \underline{0}$ (pointwise limit) we also have J' satisfying properties a) and b).

We only have to show (i). Hence, let w as in (i). We first determine an optimal policy for w, i.e. some $u^\star \in \mathcal{U}$ with $(Hw)(\rho, u^\star) = \mathcal{T}w$ (remember: we know that an optimal policy can be found in the set of deterministic policies). Consider $u^\star$ of Proposition 5.30 and assume w.l.o.g. $\rho^1 \geq \rho^3$. We get

$$\begin{aligned}
&(Hw)(\rho, u^\star) \\
&= \left(\mathcal{T}\underline{0}\right)(\rho) + \lambda \int_0^\infty e^{-2t} \sum_{i=1}^3 \sum_{j=1}^3 \rho^i Q(\Phi^{u^\star}(y^i, t); \{y^j\}) \int_{\mathbb{R}} w\left(\chi(\rho, x, t, u^\star)\right) f_\epsilon(x - y^i) dx \, dt.
\end{aligned}$$

Now, under $u^\star$ and for $t \geq \frac{1}{2}$, the sum in the integrand becomes:

$$\rho^1 \int_{\mathbb{R}} w\left(\chi(\rho, x, t, u^\star)\right) f_\epsilon(x - y^2) dx + \rho^2 \int_{\mathbb{R}} w\left(\chi(\rho, x, t, u^\star)\right) f_\epsilon(x - y^2) dx \\ + \rho^3 \int_{\mathbb{R}} w\left(\chi(\rho, x, t, u^\star)\right) f_\epsilon(x - y^3) dx.$$

By Lemma 5.37, we have now $\chi^1(\rho, x, t, u^\star) = 0$ and by Corollary 5.38 we get $\chi^2(\rho, x, s, u^\star) \geq \chi^2(\rho, x, s, u^0)$. As w satisfies property b), we thus get $w(\chi(\rho, x, s, u^\star)) \leq w(\chi(\rho, x, s, u^0))$ for all $x \in \mathbb{R}$. Policy $u^\star$ is thus better than policy u^0 for the integral above. Now as $u^\star$ is optimal for the one-step problem and as steering longer to the right would not improve the one-step cost problem neither the integral above and as steering to the left would worsten both parts as well, $u^\star$ is optimal for w. Hence, $(Hw)(\rho, u^\star) = (\mathcal{T}w)(\rho)$.

Now, to show property a) for $\mathcal{T}w$, remember that $(\mathcal{T}\underline{0})(\delta_{y^2}) = 0$. Further, by Corollary 5.39, we have $\chi(\delta_{y^2}, x, s, u^\star) = \delta_{y^2}$ and as w satisfies property a), we have for all $s \geq 0$ and all $x \in \mathbb{R} : w(\chi(\delta_{y^2}, x, s, u^\star)) = 0$. Hence, $(\mathcal{T}w)(\delta_{y^2}) = 0$.

To show property b) for $\mathcal{T}w$ let $\rho, \tilde{\rho} \in \mathbf{P}(E_Y^0)$ and assume w.l.o.g. $\tilde{\rho}^3 = 0$ and $\tilde{\rho}^2 \geq \rho^2$. By Lemma 5.37, we get for $t \geq \frac{1}{2}$ then $\chi(\tilde{\rho}, x, t, u^\star) = \delta_{y^2}$. This is the best distribution one can get for the integral above, as then, as we saw, $w(\chi(\tilde{\rho}, x, t, u^\star)) = 0$ as w satisfies property a). Furthermore, we have $(\mathcal{T}\underline{0})(\tilde{\rho}) \leq (\mathcal{T}\underline{0})(\rho)$ as $(\mathcal{T}\underline{0})$ satisfies property a). We conclude $\mathcal{T}w$ satisfies property a).

Steps (ii) and (iii) are clear. □

With this result about properties of J' we can now prove the main result of this Paragraph:

Proof (of Theorem 5.26). As we have J' satisfying properties a) and b) of Proposition 5.36, the result follows. We actually showed during the proof of Proposition 5.36, that for a function w satisfying properties a) and b) an optimal policy is given by $u^\star$ as of Proposition 5.30. Now the result follows by the Correspondence Theorem. □

5.5.2.2 Outlook: The general case

In the last Paragraph we discussed a very concrete example of an optimization problem for a controlled PO-PDMP. We could derive a characterization of an optimal policy. This discussion required a deeper analysis of of the filter and its response to a control policy. We were finally able to show that the stated policy is optimal because we were able to prove that the policy being optimal for the one-step problem is also optimal for the infinite horizon problem,too. This was possible because the jump transition kernel had good properties: Jumps from the state optimal for the cost function do not go to „bad“ starting positions for a relaunched optimization problem.

In the case of a completely observable problem, we were able to give a concrete sufficient condition for the existence of optimal bang-bang policies. This condition involved the jump transition kernel and somehow required exactly this property that jumps occuring from states near the optimum of the cost function do not lead to states with very high cost.

Such a relatively simple sufficient condition for the existence of optimal bang-bang type policies cannot be given in the general model under partial observation. The principal issue is that the control policy intervenes at two points in summand (II) of (5.11): In the argument of the jump transition kernel via Φ^r and in the argument of the filter χ. Remember: In the completely observable case, the control policy only intervenes in the argument of the jump transition kernel.

A deeper analysis of the concrete kernel properties arising from the concretely selected model paramters is required as we did here in the example. No simple convexity condition

can be applied as far as we understand by now. The topic of deriving an adequate sufficient condition on the transition kernel remains an active research topic.

However, the value of the fixed point equation we derived in this thesis for the value function J' can be best illustrated by the following: As we know that $\mathcal{T}J' = J'$, we can apply numerical approximation methods such as Howard's policy iteration, see [43] or [6]. This algorithm needs a good initial guess of an optimal policy. For problems with convex cost function, we highly recommend to always initialize the algorithm with the optimal policy for the one-step problem. As we saw, this one-step problem is still an easy problem where the techniques for convex cost functions apply as detailed in this Chapter.

Chapter 6

Further applications and outlook

Having developed a theory for optimal control of PO-PDMPs under a set of important model assumptions, we end this thesis with a brief discussion of the main assumptions taken in order to derive existence of optimal policies. While some assumptions taken were of technical relevance (e.g., uncontrolled jump intensity), others were a free choice (e.g., how to model partial observation). The question on practical relevance of these assumptions in concrete applications of the theory thus arises. Therefore, we aim to end this thesis by a discussion that is twofold:

First, we highlight the main assumptions taken and discuss them in view of concrete applications. Three questions will guide our discussion in Section 6.1:

(i) What restriction does this assumption have on possible applications?

(ii) For what kind of applications, this assumption would not present any restriction?

(iii) Would there be alternative ways to model a PO-PDMP control problem leading to a different set of model assumptions?

From this discussion, we then derive, in a second step, a perspective on possible refinements and extensions of the PO-PDMP control theory developed in this thesis. This final discussion is presented in Section 6.2. We highlight, were the PO-PDMP control theory presented could be refined or even extended. By a refinement, we understand the attempt to get either more precise results under the same set of assumptions or the same kind of results under less restrictive assumptions. Extensions of the theory are the attempt to cover an even broader range of possible applications. This might lead to an even more general model or different models that, however, are similar to ours in many aspects.

6.1 Model assumptions in view of concrete applications

The range of possible applications of the general PDMP control theory is broad. There are applications in finance [57], communication networks [15] or [40], neurosciences [50] and biochemics [49] to only list a very short overview that hopefully illustrates the huge variety of domains of application for optimal control problems for PDMPs.

Many of them bear the potential of becoming applications of the PO-PDMP control theory presented in this thesis. As soon as there arises a situation where, for one of these applications, the post-jump state can only be observed by a noisy measurement and the underlying process is unobservable, we enter the context of our model. For applications of

Biology, Chemistry or Neurosciences, such applications can arise whenever an MRI[1], an X-ray or other imaging tool is required to measure the system state. In order to apply our PO-PDMP theory, a set of further assumptions has to be satisfied by the model arising from the application. We summarized these assumptions in Annex C.

The assumptions bearing the highest potential of presenting a restriction to possible applications shall be discussed in the following sections while always having in mind the three guiding questions introduced in the introduction to this chapter. We try to enrich our discussion by always presenting concrete applications where these assumptions are satisfied.

6.1.1 The finite dimensional case

The assumption of only having finitely many possible post-jump states (see Assumption 2.16) might seems as the most restrictive assumption taken throughout this thesis. However, in view of many possible applications as well as in view of numerical approaches to determine a concrete optimal policy, this assumption is meaningful.

In many applications, the „noisy measurement“ of the post-jump state might actually be an „estimation“ of a certain quantity. Think of examples in queueing theory, where the workload waiting to be processed has to be estimated. First, for many examples in queueing theory, one might assume a maximum length of the queue. This maximum length might be very long but assume a cash desk in a supermarket: There is certainly a length that, if achieved, will lead to no new customer even entering the supermarket. Hence, assuming a bounded state space makes sense in many applications.

Now, having only a discrete set of possible post-jump states might make sense in a lot of examples from queueing theory as well. Think of queues in production lines, were randomly positioned spare parts have to be processed. Depending on the concrete position of a spare part that is the next to be processed, a robot has to perform the following: Correct positioning of the part, first, followed by adding the spare part to, say the car produced. Often, the set of possible positions of the spare part right before being processed is finite and we enter the context of our model.

In view of numerical approaches the power of our model is the following: We saw that to determine a concrete optimal policy one has to study the fixed point equation of the value function. We studied this equation in the case of convex running cost functions. For more general cases with less regularity properties of the running cost function, an analytic approach becomes arbitrarily complex. In many cases, however, one can use numerical approximation to determine an optimal policy. Howard's policy iteration is one of the adequate approaches here, see [6], Theorem 7.2.1 or [43]. However, one has to pass by a discretization of the filter for this numerical approach and we deliver an adequate discretization with our theory.

6.1.2 Uncontrolled jump intensity and jump transition kernel

The assumption of having an uncontrolled jump intensity and uncontrolled jump transition kernel (see Assumption 3.9) might as well seem as a serious restriction of the model in view of possible applications. However, we still allow for both to be state dependent, hence they do not have to be constant. State dependent but uncontrolled jump intensities and jump transition kernels is what one can find in many applications. Very often, these two characteristics are out of the sphere of influence of an agent. Think of the queue at the

[1] Magnetic Resonance Imaging

cash desk of a grocery store again. Here, certainly λ is state dependent as for very long queues, less and less new customers will enter the queue.

To remain with an example from queueing theory: Controlling λ and Q would mean to control how often new work load is arriving to the queue and, in case of bulk arrivals, controlling Q would mean to control the amount of new workload arriving. In many applications this is not possible.

There are cases, however, where one could assume to have these characteristics controlled. Assume a logistics provider or even an online retailer watching its workload in its logistics queue. By means of advertisement, e.g., special offers, the online retailer could increase the jump intensity as more of the very price sensitive customers would start buying its products. Hence, jumps occur more often. By only positioning advertisements for large products that need more work steps in the logistics chain, the retailer could also control Q in a way to get jumps leading to greater bulks of work load arriving to the queue.

Very often, controlling λ or Q goes along with higher cost. In the example above, positioning advertisements is expensive. This leads to a running cost function c depending on the control action $a \in A$ executed.

6.1.3 Noisy measurement of post-jump state

The way how to model the noisy observation of the underlying PDMP is a free choice when setting up the model. Depending on the concrete application in mind, different ways how to model this observation might be adequate:

One could assume, what we did, to only get some kind of „triggered“ information whenever a jump occurs. We assume to get a noisy measurement of the post-jump state of the process right at the jump time of the process. This makes perfect sense in many applications such as break downs in production lines or communication networks, natural catastrophes, medical diagnostics and many more.

For some applications in queueing theory, it might also make sense to assume to get a noisy measurement of the jump height, i.e. of the quantity of new work load arriving to the queue. Actually, this might make sense as often, there is no good reason why to estimate or measure again the full length of the queue as knowing the estimate of arriving work loads and knowing the processing speed is enough to estimate the full length of the queue. This way of observing the underlying PDMP, however, would lead to a slightly different model than the one presented in this thesis while major parts of the theory would work analogously.

Another way to model partial observation of the underlying PDMP would be to assume noisy measurements of the current process state at deterministic, perhaps equi-distant, points in time. Hence, more measurements of the unobservable state could lead to a more precise filtered process and thus, to better optimal policies. Again, this would lead to a different model and in addition, one could raise the question if, in this case, feedback controls would be more appropriate in combination with these recurring measurements of the process state.

Finally, in order to get the best estimate possible of the state of the underlying process, one could even imagine to perform measurements in continuous time. For many applications, however, this is not possible, either because of technical restrictions or, because cost of continuous time measurements is so high that possible gains in total discounted cost over life time, compared to our model, are netted out by these high measurement cost.

6.1.4 Observation of inter-jump time

Initially, we assumed perfect observation of the inter-jump time. It turned out that we had to restrict our model to the case of uncontrolled jump intensities and jump transition kernels in order to prove existence of optimal policies. In Chapter 4, we then showed existence of optimal policies for models with unobservable inter-jump time. As discussed in Chapter 4, even assuming a small, concentrated measurement noise for te inter-jump time would lead to the existence of optimal policies even for models with controlled jump intensity and controlled jump transition kernel.

In view of possible applications, the assumption of having a noisy measurement of the inter-jump time is completely meaningful. Time has to be measured in some way and assuming measurement noise is absolutely adequate. We come back to this point in the outlook discussion for possible refinements of the theory.

6.1.5 Behavior at the border

Many classical PDMP models assume initiated jumps when the process reaches the border of the state space. Hence, a jump occurs either triggered by the intensity λ and is of random nature or, occurs at the border of the state space and is of deterministic nature. In many applications, this assumption makes perfect sense. Think again of queues at cash desks of grocery stores: Whenever the queue at cash desk number 1 achieves a pre-defined maximum length, cash desk number 2 is openend and half of the customers is transferred to cash desk number 2. Hence, when reaching the border of the state space, the process jumps to half of the maximum length.

In our model, we did not include these jumps at the border, mainly because we work under the assumption of finitely many possible post-jump states together with the limit property for the jump intensity: The intensity λ^A satisfies the following limit property:

$$\forall n \in \mathbb{N}, \pi_n \in \Pi_n^P, h_n \in \mathcal{H}_n, y \in E_Y :$$

$$\lim_{t\to\infty} \int_0^t \int_A \lambda^A(\Phi^{\pi_n(h_n,\cdot)}(y,s),a)\ \pi_n(h_n,s)(da)\ ds = \infty.$$

Almost surely, there is a jump after finite time, bringing back the process to a state in E_Y^0.

6.2 Outlook: Possible refinements and extensions of PO-PDMP control theory

Looking at the PO-PDMP control model we developed in this thesis as well as at the existence results for and characterizations of optimal policies we derived, further refinements and extensions of the PO-PDMP control theory presented seem possible.

Whereas our initial optimization problem of Definition 1.37 is a classical optimization problem for total discounted cost, a second class of optimization problems for PDMPs and MDPs is very common: Minimum average cost problems. Various authors have published results on average cost problems for PDMPs, e.g., [20], [19] or [42], as well as for MDPs (e.g., [2]) or for problems close to SFPs (e.g., [16]). An average cost theory for PO-PDMPs could be inspired by the approaches of these and other works.

Staying in the context of total dicounted cost problems, the presented PO-PDMP control model could be refined under the following three aspects:

First, as we already indicated in Chapter 4 and Section 6.1.4, one could pass to a model where the inter-jump time is not or only under noisy measurement observed. We treated the case of unobservable inter-jump times in Chapter 4 but the model under noisy observation of the inter-jump time could be developed along the lines of the present model. A filter could be determined analogously to our presented filter equation for χ. Basically, an additional integral w.r.t. the noise density of the measured inter-jump time would appear in the filter.

Second, the question of existence of optimal policies in the class $\mathcal{U}$ of deterministic controls could be addressed in a general theory, not only in the case of concrete applications as we did in Chapter 5. A possible approach could be oriented on the work of Forwick [35], where additional convexity requirements as well as a separation of the drift into a controlled part and an uncontrolled part are necessary to derive existence of optimal deterministic controls.

Third, the assumption of having finitely many possible post-jump states could be skipped. This would lead to a filter where summation w.r.t. to last observed conditional distribution is replaced by an adequate integral. Additional integrability properties are then required and one has to investigate what this implies on necessary conditions on the underlying PO-PDMP. However, as determination of concrete optimal policies will very often pass by numerical approximation (see also below), a discretization or quantization of the filter will be required again. Hence, we hold our approach of a „finite dimensional case“ for adequate.

In terms of extensions of the model presented in this thesis, the following four points seem reasonable to be addressed as next steps:

First, a more general theory for the characterization of an optimal policy, also covering the case of not convex running cost functions could be developed. Here, not only the case of „bang-bang“ policies should be covered. Also, approximation of optimal policies by, e.g., Howard's policy iteration algorithm (see [6], Theorem 7.2.1 or [43]) or other numerical methods, see, e.g., [47], could be addressed.

Second, existence of optimal policies was shown for lower semi-continuous running cost functions. The theory could be extended to other classes of running cost functions. Here, one has to keep in mind what results on selection of measurable optimizers exist as these results are at the core of the existence proof for one-step optimizers. An overview of such selection theorems offers, e.g., [11].

Third, the PO-PDMP theory developed here could be extended to other models of partial observation. We discussed in Section 6.1.3 how these models could look like. Especially the case of using feedback controls coupled with repeating noisy measurements of the unobservable state of the underlying PDMP seems promising. An extension of the PO-PDMP theory to the use of feedback controls could be based on Forwick's work [35] for completely observable PDMPs.

Finally, the theory could also be extended to the case of partial observation modeled by unknown parameters such as, e.g., unknown jump intensity λ. This extension could also contain the case of so-called hidden Markov models, see also [6], Example 5.1.2.

Appendix A

The Young topology on the space of relaxed controls

In this appendix, we develop all results related to the Young topology on $\mathcal{R}$ we use in the main part of this thesis. We define the space $\mathcal{R}$ and and the Young topology on this space. We develop the most important properties of this topological space and finally show its compactness. We further proof the correspondance Theorem required for the first reformulation of the initial optimization problem.

This annex is largely inspired by the great work Forwick did in [34], but as we work with the space of observable histories up to time T_n, we had to adapt the theory slightly, especially for the proof of the correspondence theorem. This annex is not intended to deliver all necessary basics of measure theory, probability theory or stochastic dynamic programming. We tried to present the most important results of these domains in a version adapted to the concrete situations where we will apply these results in the main part of this thesis. For a more rigorous treatment and a complete overview of foundations of probability theory we refer to the excellent book of Kallenberg [45]. For more background on the investigated measurability questions and on foundations of stochastic dynamic programming we refer to the book of Bertsekas and Shreve [11].

A.1 The action space A and the relaxed action space $\mathbf{P}(A)$

We assume the action space A to be a compact metric space and denote by d_A its metric. Let $\mathcal{T}[d_A]$ denote the topology on A induced by the metric d_A. The following notations will be used throughout the rest of this work: The couple $(A, \mathcal{T}[d_A])$ will refer to the topological space and the couple $(A, \mathcal{B}_A)$ will refer to the measurable space endowed with the Borel-σ-algebra $\mathcal{B}_A$ generated by the open sets of the topology $\mathcal{T}[d_A]$.

Lemma A.1. *A is a compact Polish space with fixed metric d_A.*

Proof. The completeness and separability of A follow directly from the assumed compactness of the metric space A. (See, e.g., Corollary 7.6.2 of [11]). □

A.1.1 The space $\mathbf{C}(A)$

We denote by $\mathbf{C}(A)$ the set of continuous and bounded functions from $(A, \mathcal{T}[d_A])$ to $(\mathbb{R}, \mathcal{T}_{\mathbb{R}})$. For $f \in \mathbf{C}(A)$, we define the usual norm

$$\|f\|_\infty := \sup_{a \in A} |f(a)|_{\mathbb{R}}.$$

This norm induces a metric on $\mathbf{C}(A)$ and we will denote the topology induced by this metric by $\mathcal{T}[\|\cdot\|_\infty]$.

Lemma A.2. *The space* $(\mathbf{C}(A), \mathcal{T}[\|\cdot\|_\infty])$ *is separable.*

Proof. For a proof see, e.g., [11], Proposition 7.7. □

A.1.2 The space $\mathbf{P}(A)$

We will now turn into a short introduction of the weak topology on $\mathbf{P}(A)$. There is a lot more details about this "standard" topology for spaces of probability measures in books like, e.g., [11]. As the space $\mathbf{P}(A)$ is at the core of "relaxed control strategies", however, we will present here at least the definition and some basic properties of the weak topology on $\mathbf{P}(A)$.

Remember the definition of $\mathbf{P}(A)$ to be the space of all probability measures on the measurable space $(A, \mathcal{B}_A)$ as introduced above.

Definition A.3 (Weak topology on $\mathbf{P}(A)$). *The* weak topology *on* $\mathbf{P}(A)$ *is* $\mathcal{T}[\mathbf{C}(A)]$*, where we define for a subset* $D \subset \mathbf{C}(A)$ *the topology* $\mathcal{T}[D]$ *to be the topology with subbase*

$$\mathcal{V}(D) := \{V_\epsilon(p, f) : \epsilon > 0, p \in \mathbf{P}(A), f \in D\} \tag{A.1}$$

where we use for $\epsilon > 0$ *and* $f \in D$ *the notation*

$$V_\epsilon(p, f) := \left\{ q \in \mathbf{P}(A) : \left| \int f dq - \int f dp \right| < \epsilon \right\}. \tag{A.2}$$

In a sense, one can understand $V_\epsilon(p, f)$ as a "from f induced ϵ-neighborhood of $p \in \mathbf{P}(A)$" and the subbase $\mathcal{V}(D)$ is then the "collection of all ϵ-neighborhoods induced by an $f \in D$ around all $p \in \mathbf{P}(A)$".

An example of a probability measure $q \in V_\epsilon(p, f)$ would be a q that only differs from p outside of the support of f. Generally speaking, $V_\epsilon(p, f)$ contains all $q \in \mathbf{P}(A)$ such that the "f-weighted" probability mass of q does not differ too much (i.e. less than ϵ) from the "f-weighted" probability mass of p.

An important property of the so-defined weak topology on $\mathbf{P}(A)$ is that, for all $f \in \mathbf{C}(A)$, the mapping

$$\theta_f : \mathbf{P}(A) \to \mathbb{R},\ p \mapsto \int f dp$$

is continuous w.r.t. the standard topology on $\mathbb{R}$ and $\mathcal{T}[\mathbf{C}(A)]$ on $\mathbf{P}(A)$, as shown by the next lemma.

Lemma A.4. *Let* A *a metrizable space and* $D \subset \mathbf{C}(A)$*. Let further* $\{p_\alpha\}$ *a net in* $\mathbf{P}(A)$ *and* $p \in \mathbf{P}(A)$*. Then* $p_\alpha \to p$ *relative to the topology* $\mathcal{T}[D]$ *if and only if* $\int f dp_\alpha \to \int f dp$ *for every* $f \in D$*.*

Proof. If $p_\alpha \to p$ relative to $\mathcal{T}[D]$ then, for every open neighborhood U of p we find β such that for all $\alpha \geq \beta$ we have $p_\alpha \in U$. Thus, for all $\epsilon > 0$ and $f \in D$, there exists β such that for all $\alpha \geq \beta$, we have $p_\alpha \in V_\epsilon(p, f)$. As $p_\alpha \in V_\epsilon(p, f)$ implies $|\int f dp_\alpha - \int f dp| < \epsilon$ we conclude $\int f dp_\alpha \to \int f dp$.

If $\int f dp_\alpha \to \int f dp$ for all $f \in D$, we know that for $\epsilon > 0$ and $f \in D$ we can find a β such that for all $\alpha \geq \beta$ we have $|\int f dp_\alpha - \int f dp| < \epsilon$. Let now $U \in \mathcal{T}[D]$ with $p \in U$. As $\mathcal{V}(D)$ is a subbase of $\mathcal{T}[D]$, p is contained in some basic open set $\cap_{k=1}^n V_{\epsilon_k}(p, f_k) \subset U$ where $\epsilon_k > 0$ and $f_k \in D$ for $k = 1, \ldots, n$. Choose β such that for all $\alpha \geq \beta$ we have $|\int f_k dp_\alpha - \int f_k dp| < \epsilon_k, k = 1, \ldots, n$. Then $p_\alpha \in U$ for $\alpha \geq \beta$ and thus, $p_\alpha \to p$ realtive to $\mathcal{T}[D]$. □

The weak topology still being rather "abstract" in how it is defined and the space $\mathbf{C}(A)$ being too large to be manipulated easily, some further results re. the weak topology shall be cited here to illustrate how to make this topology more "tangible" and finally, how to work with convergent sequences rather than with convergent nets.

The first result shows that already a countable set is sufficient to generate $\mathcal{T}[\mathbf{C}(A)]$. We denote by $U_d(A)$ the subset of $\mathbf{C}(A)$ of all uniformly continuous functions w.r.t a metric d on A.

Lemma A.5. *Let A a separable and metrizable space. There is a metric d on A consistent with its topology and a countable dense subset $D \subset U_d(A)$ such that $\mathcal{T}[D]$ is the weak topology $\mathcal{T}[\mathbf{C}(A)]$ on $\mathbf{P}(A)$.*

Proof. The full proof can be found in [11], Proposition 7.19. The major incredients of the proof are the fact the functions $f \in \mathbf{C}(A)$ can be approximated from below and from above by functions $g_n, h_n \in U_d(A)$, see Lemma 7.7 in [11]. By the help of such approximating functions, one can show that $\mathcal{T}[\mathbf{C}(A)] = \mathcal{T}[U_d(A)]$, see Lemma 7.8 in [11]. For a dense subset D of $U_d(A)$ it is then easy to show, that $\mathcal{T}[D] = \mathcal{T}[U_d(A)]$, see Lemma 7.9 in [11]. The existence of such a dense set $D \subset U_d(A)$ follows then from the fact that the separable metrizable space A has a totally bounded metrization d and from the fact that $U_d(A)$ is separable if (A, d) is totally bounded as metric space, see Corollary 7.6.1 and Proposition 7.9 of [11]. □

In a sense, $\mathbf{P}(A)$ inherits separability and metrizability from A:

Lemma A.6. *If A is separable and metrizable, then $\mathbf{P}(A)$ is separable and metrizable.*

Proof. The proof is mainly based on the previous result that $\mathcal{T}[\mathbf{C}(A)] = \mathcal{T}[D]$ for a dense subset D of $U_d(A)$. See Proposition 7.20 of [11] for further details. □

This last result guarantees that, in our setting of the controlled PO-PDMP, where the action space A is a separable and metrizable space, the weak topology on $\mathbf{P}(A)$ can be characterized in terms of convergent sequences rather than nets.

Proposition A.7. *Let A be a separable metrizable space and let d a metric on A consistent with its topology. Let $\{p_n\}$ a sequence in $\mathbf{P}(A)$ and $p \in \mathbf{P}(A)$. Then, the following statements are equivalent:*

(a) $p_n \to p$;

(b) $\int f dp_n \to \int f dp$ for every $f \in \mathbf{C}(A)$;

(c) $\int g dp_n \to \int g dp$ for every $g \in U_d(A)$;

(d) $\limsup_{n\to\infty} p_n(F) \leq p(F)$ *for every closed set* $F \subset A$*;*

(e) $\liminf_{n\to\infty} p_n(G) \geq p(G)$ *for every open set* $G \subset A$*.*

Proof. The equivalence of (a), (b) and (c) follows from Lemma A.4 (a sequence is a net) and from the fact that $\mathcal{T}[\mathbf{C}(A)] = \mathcal{T}[U_d(A)]$. Equivalence of (d) and (e) follows by complementation. For the rest see Proposition 7.21 of [11]. □

The main result of this section states that $\mathbf{P}(A)$ is inheriting all good properties from the action space A we selected for our controlled PO-PDMP model:

Proposition A.8. *The space* $\mathbf{P}(A)$ *of probability measures on the compact metric action space* A *endowed with the Borel-*σ*-algebra* $\mathcal{B}_A$ *as introduced above is a compact Polish space.*

Proof. As A is compact and metric, compactness of $\mathbf{P}(A)$ follows from Proposition 7.22 of [11]. According to Lemma A.6, $\mathbf{P}(A)$ is separable and metrizable. As A is complete and separable (see Lemma A.1), this properties follow for $\mathbf{P}(A)$ as shown in Proposition 7.23 of [11]. □

A.2 The space $\mathcal{R}$ of relaxed controls

Definition A.9. *We define the space* $\mathcal{R}$ *of* relaxed controls *as*

$$\mathcal{R} := \{[r] \mid r : [0,\infty) \to \mathbf{P}(A), r \text{ measurable}\},$$

where $[r]$ *denotes the* λ^1 *equivalence class of* r*. To simplify notations we will write* r_t *instead of* $r(t)$*.*

Some remarks regarding this definition:

- We use λ^1 equivalence classes in this definition, that means $\tilde{r} \in [r] \Leftrightarrow \tilde{r} = r$ for λ_1-almost all $t \in [0,\infty)$.
- r shall be measurable w.r.t. the Borel-σ-algebras $\mathcal{B}([0,\infty))$ and $\mathcal{B}_{\mathcal{T}[\mathbf{C}(A)]}(\mathbf{P}(A))$, i.e.

  $$r^{-1}(V_\epsilon(p,f)) \in \mathcal{B}([0,\infty)) \quad \forall \epsilon > 0, f \in \mathbf{C}(A), p \in \mathbf{P}(A).$$

  (Remind that $\mathbf{P}(A)$ is separable and metrizable and thus $\mathcal{B}_{\mathcal{T}[\mathbf{C}(A)]}(\mathbf{P}(A)) = \sigma(\mathcal{V}(\mathbf{C}(A)))$.)

A.3 Definition of the Young Topology on $\mathcal{R}$

The definition of the Young topology on $\mathcal{R}$ is a rather abstract definition involving an L^1 function space and its dual space. We first introduce these two spaces before we turn to the definition of the Young topology.

Definition A.10. *Let* $\mathcal{B}(\mathbf{C}(A))$ *the Borel-*σ*-algebra on* $\mathbf{C}(A)$ *induced by the* $\|\cdot\|_\infty$ *norm on* $\mathbf{C}(A)$*. We then define*

$$\mathbb{X} := L^1([0,\infty), \mathbf{C}(A)) \tag{A.3}$$

as the space of λ^1*-equivalence classes of* $(\mathcal{B}([0,\infty)), \mathcal{B}(\mathbf{C}(A)))$*-measurable functions such that*

$$\|\psi\|_{\mathbb{X}} := \int_0^\infty \|\psi(t,\cdot)\|_\infty \, dt < \infty \quad \forall \psi \in \mathbb{X}. \tag{A.4}$$

It is easy to see, that $(\mathbb{X}, \|\cdot\|_{\mathbb{X}})$ is a normed linear space. Thus, the norm $\|\cdot\|_{\mathbb{X}}$ induces a topology $\mathcal{T}[\|\cdot\|_{\mathbb{X}}]$ on $\mathbb{X}$.

Remark A.11. *According to lemma B.2, the measurability requirement for $\psi \in \mathbb{X}$ is equivalent to the measurability of $t \mapsto \psi(t, a)$ for fix $a \in A$.*

Definition A.12. *We define the dual space $\mathbb{X}^\star$ of $\mathbb{X}$ as*

$$\mathbb{X}^\star := \{F : \mathbb{X} \to \mathbb{R} \mid F \text{ linear and continuous}\}, \tag{A.5}$$

and endow this dual space with the weak-$\star$-topology, i.e. the topology induced by the family of mappings

$$\{E_\psi : \mathbb{X}^\star \to \mathbb{R}, F \mapsto F(\psi) \mid \psi \in \mathbb{X}\}. \tag{A.6}$$

Definition A.13. *The Young topology on $\mathcal{R}$ is the topology induced by the mapping $i : \mathcal{R} \to \mathbb{X}^\star$, where we define for $\psi \in \mathbb{X}$:*

$$i(r)(\psi) := \int_0^\infty \int_A \psi(t, a) r_t(da)\, dt, \tag{A.7}$$

and $\mathbb{X}^\star$ is endowed with the above defined weak-$\star$-topology.

Remark A.14. *Two remarks on this definition of the Young topology:*

- *The mapping i is well-defined with regard to its dependence on the selection of a representative of $[r]$, as one can easily see when rather writing $\int_{[0,\infty)} \ldots \lambda^1(dt)$ instead of $\int_0^\infty \ldots dt$ in the definition of $i(r)(\psi)$ above.*
- *The Young topology is induced from the standard topology on $\mathbb{R}$ in two steps:*

$$\mathcal{R} \xrightarrow{i} \mathbb{X}^\star \xrightarrow{(E_\psi)_{\psi \in \mathbb{X}}} \mathbb{R}, \tag{A.8}$$

step 1 induces the weak-$\star$-topology $\mathcal{T}[(E_\psi)_{\psi \in \mathbb{X}}]$ on $X^\star$ and step 2 induces the Young topology $\mathcal{T}[i]$ on $\mathcal{R}$.

A.4 Properties of the Young Topology on $\mathcal{R}$

We will list and prove here some important properties of the Young topology on $\mathcal{R}$. We will need these properties to prove the compactness of $\mathcal{R}$ under the Young topology and for continuity and measurability investigations. The latter ones will be important for the existence of optimal relaxed control strategies for our PO-PDMP.

As the Young topology is deduced form the weak-$\star$-topology on $\mathbb{X}^\star$ by some mapping $i : \mathcal{R} \to \mathbb{X}^\star$, we will first have a closer look on $\mathbb{X}^\star$ and the mapping i.

Definition A.15. *Let B_1 denote the closed unit ball in $\mathbb{X}^\star$ w.r.t. the operator norm on $\mathbb{X}^\star$, i.e.*

$$B_1 := \{F \in \mathbb{X}^\star : \|F\|_{\mathbb{X}^\star} \leq 1\}, \tag{A.9}$$

where the operator norm on $\mathbb{X}^\star$ is defined for all $F \in \mathbb{X}^\star$ as

$$\|F\|_{\mathbb{X}^\star} := \sup\left\{\frac{|F(\psi)|_{\mathbb{R}}}{\|\psi\|_{\mathbb{X}}} : \psi \in \mathbb{X}, \psi \neq 0\right\}. \tag{A.10}$$

This unit ball of the dual space $\mathbb{X}^\star$ plays an important role in the proof for the weak-$\star$-compactness of $\mathcal{R}$. That's why we will present here separately two results regarding this unit ball B_1. We will need both results for the compactness proof of $\mathcal{R}$ later.

Before we give these results, however, the reader should note that $\mathbb{X}^\star$ becomes a normed vector space by the operator norm $\|\cdot\|_{\mathbb{X}^\star}$ used in Definition A.15. Under this norm, $\mathbb{X}^\star$ is clearly a metric space. However, we are not interested in this metric topology coming from the operator norm in our investigations regarding the Young topology. As stated before, we look at $\mathbb{X}^\star$ as topological space endowed with the weak-$\star$-topology. It is a well-known result that in general, $\mathbb{X}^\star$ *is not metrizable under this topology.* If $\mathbb{X}$ is a *separable* Banach space, however, one can show that B_1 is metrizable and even separable under the weak-$\star$-topology. This is what the following two results show.

We start with a classic result of functional analysis, called Alaoglu's theorem:

Proposition A.16 (Alaoglu's Theorem). *For any normed linear space X, the closed unit ball (w.r.t. the operator norm) B_1 of the dual $X^\star$ is compact under the weak-$\star$-topology.*

Proof. For a proof see, e.g. [28], chapter II, The weak-$\star$-topology, Alaoglu's Theorem. The proof relies on Tychonoff's theorem and thus on the axiom of choice. □

The next result tells us that at least the closed unit ball B_1 of $\mathbb{X}^\star$ is metrizable under the relative weak-$\star$-topology whenever $\mathbb{X}$ is separable.

Lemma A.17. *If X is a separable normed space then the closed unit ball (w.r.t. the operator norm) B_1 of the dual $X^\star$ is metrizable under the relative weak-$\star$-topology.*

Proof. As X is separable, there is a countable dense subset $D = \{x_n \mid n \in \mathbb{N}\}$ of the open unit ball in X. For $F, G \in X^\star$, define (using the duality pairing $< \cdot, \cdot >$)

$$d(F,G) := \sum_{n\in\mathbb{N}} \frac{| < F - G, x_n > |}{2^n}. \tag{A.11}$$

One can show that D is dense in the closed unit ball of X, that d is a metric on B_1, that the topology generated by d is a subset of the relative weak-$\star$-topology on B_1 and that the relative weak-$\star$-topology on B_1 is included in the topology generated by d. All these steps are well known to the author but we omit the lengthy proof here.

The reader might also refer to [28], chapter III, Exercise 2(i) as well as to page 226 of [28], where this result can also be found. □

The important result for us is now the combination of the latter two results, namely that B_1 is separable and metrizable under the weak-$\star$-topology if $\mathbb{X}$ is separable.

Lemma A.18. *If X is a separable Banach space, then the closed unit ball (w.r.t. the operator norm) B_1 of the dual $X^\star$ is compact, separable and metrizable under the relative weak-$\star$-topology.*

Proof. Compactness (under the relative weak-$\star$-topology) of B_1 follows without separability of X from Alaoglu's Theorem, theorem A.16. Metrizability (under the relative weak-$\star$-topology) of B_1 follows from lemma A.17 as X is supposed to be separable. For a metric space, compactness implies separability. □

Lemma A.19. *The mapping $i : \mathcal{R} \to \mathbb{X}^\star$ is injective with $i(\mathcal{R}) \subset B_1$. In particular, $i(\mathcal{R})$ is homeomorphic to $\mathcal{R}$ w.r.t. the relative weak-$\star$-topology (restriction of weak-$\star$-topology on $\mathbb{X}^\star$ to $i(\mathcal{R})$).*

Proof. We first show $i(\mathcal{R}) \subset B_1$. Let $r \in \mathcal{R}$, then $\|i(r)\|_{\mathbb{X}^\star} \leq 1$ as for $\psi \in \mathbb{X}$ we get:

$$|i(r)(\psi)| = \left| \int_0^\infty \int_A \psi(t,a) r_t(da)dt \right| \leq \int_0^\infty \int_A |\psi(t,a)| r_t(da)dt \leq \int_0^\infty \|\psi(t,\cdot)\|_\infty \, dt = \|\psi\|_{\mathbb{X}} . \tag{A.12}$$

To show that i is injective, let $r, r' \in \mathcal{R}$ with $i(r) = i(r')$. By definition of i we then get:

$$\begin{aligned} & i(r)(\psi) = i(r')(\psi) \quad \forall \psi \in \mathbb{X} \\ \iff & \int_0^\infty \int_A \psi(t,a) r_t(da)dt = \int_0^\infty \int_A \psi(t,a) r'_t(da)dt \quad \forall \psi \in \mathbb{X} \end{aligned}$$

If $f \in L^1([0,\infty))$ and $c \in \mathbf{C}(A)$ then $[0,\infty) \ni t \mapsto f(t) \cdot c(\cdot) \in \mathbf{C}(A)$ belongs (as mapping) to the space $\mathbb{X}$ and thus

$$\begin{aligned} & \implies \int_0^\infty f(t) \left\{ \int_A c(a) r_t(da) - \int_A c(a) r'_t(da) \right\} dt = 0 \quad \forall f \in L^1([0,\infty)), c \in \mathbf{C}(A) \\ & \implies \int_A c(a) r_t(da) = \int_A c(a) r'_t(da) \quad \lambda^1\text{−a.e. } \forall c \in \mathbf{C}(A) \end{aligned}$$

By lemma A.2, the space $\mathbf{C}(A)$ is separable and we can choose a dense subset $C' \subset \mathbf{C}(A)$. With this, we find (second step of the following uses the fact that a countable union of λ^1 null sets is still a λ^1 null set):

$$\begin{aligned} & \implies \int_A c(a) r_t(da) = \int_A c(a) r'_t(da) \quad \lambda^1\text{−a.e. } \forall c \in C' \\ & \implies \int_A c(a) r_t(da) = \int_A c(a) r'_t(da) \quad \forall c \in C' \ \lambda^1\text{−a.e.} \\ & \implies \int_A c(a) r_t(da) = \int_A c(a) r'_t(da) \quad \forall c \in \mathbf{C}(A) \ \lambda^1\text{−a.e.} \\ & \implies r_t = r'_t \quad \lambda^1\text{−a.e.} \\ & \implies [r] = [r']. \end{aligned}$$

□

Lemma A.20. *The (topological) spaces $\mathbf{P}(A)$, B_1 (endowed withe the weak-$\star$-topology) and $\mathcal{R}$ are separable and metrizable.*

Proof. According to lemma A.1, A is separable and metrizable and thus, $\mathbf{P}(A)$ is separable and metrizable (see Proposition A.6).

If we show that $\mathbb{X}$ is a separable Banach space, then B_1 is separable and metrizable under the weak-$\star$-topology according to lemma A.18.

With B_1, we then have as well $i(\mathcal{R})$ and $\mathcal{R}$ separable and metrizable according to lemma A.19.

We end by proving that $\mathbb{X} = L^1([0,\infty), \mathbf{C}(A))$ is a separable Banach space. It is clearly a normed vector space and completeness follows from [29], Theorem III.6.6.

To show separability of $\mathbb{X}$, we start with the spaces $L^1([0,T], \mathbf{C}(A))$. As $\mathbf{C}(A)$ is a separable Banach space (lemma A.2), $[0,T]$ is a compact metric space and λ^1 a positive Radon measure on $\mathcal{B}([0,T])$, we can apply [60], Theorem I.5.18, and $L^1([0,T], \mathbf{C}(A))$ is separable for all $T > 0$. Thus, let M^k a countable dense subset of $L^1([0,k], \mathbf{C}(A))$, then the set

$$M := \bigcup_{k \in \mathbb{N}} M^k$$

is still countable and dense in $L^1([0,\infty], \mathbf{C}(A))$. Note, that one can extend every mapping $[0,k] \to \mathbf{C}(A)$ to a mapping $[0,\infty) \to \mathbf{C}(A)$ by mapping $(k,\infty) \ni t \mapsto \underline{0}$, where $\underline{0} \in \mathbf{C}(A)$ is the mapping $A \ni a \mapsto 0 \in \mathbb{R}$. □

Having shown that $\mathcal{R}$ is separable, we can use sequences rather than nets to characterize the young Topology on $\mathcal{R}$.

Lemma A.21 (Characterization of Young Topology). *Let $(r^n)_{n\in\mathbb{N}}$ a sequence in $\mathcal{R}$ and $r \in \mathcal{R}$. We then have the following equivalence of convergence in $\mathcal{R}$ resp. $\mathbb{R}$:*

$$r^n \overset{n\to\infty}{\longrightarrow} r \quad \iff \quad \int_0^\infty \int_A \psi(t,a)\; r_t^n(da)\; dt \overset{n\to\infty}{\longrightarrow} \int_0^\infty \int_A \psi(t,a)\; r_t(da)\; dt \quad \forall \psi \in \mathbb{X} \tag{A.13}$$

Proof. We have the following equivalences:

$$\begin{aligned} r^n \to r \text{ (convergence in } \mathcal{R}) &\iff i(r^n) \to i(r) \text{ (convergence in } \mathbb{X}^\star) \\ &\iff i(r^n)(\psi) \to i(r)(\psi)\ \forall \psi \in \mathbb{X} \text{ (convergence in } \mathbb{R}), \end{aligned}$$

where we used that the mapping $i : \mathcal{R} \to i(\mathcal{R})$ is a homeomorphism (first equivalence) and that the mappings $E_\psi : \mathbb{X}^\star \to \mathbb{R}, F \mapsto E_\psi(F) := F(\psi)$ generating the weak-$\star$-topology on $\mathbb{X}^\star$ are continuous (second equivalence). □

Corollary A.22 (Generating functions for Young Topology). *The Young Topology on $\mathcal{R}$ is generated by the mappings $\mathcal{R} \to \mathbb{R}, r \mapsto \int_0^\infty \int_A \psi(t,a) r_t(da)\; dt$ for $\psi \in \mathbb{X}$.*

Corollary A.23 (measurability criteria). *The following equivalences for measurability hold:*

(1) $r : [0,\infty) \to \mathbf{P}(A)$ is measurable $\iff$
$[0,\infty) \ni t \mapsto \int_A c(a)\; r_t(da) \in \mathbb{R}$ is measurable $\forall c \in \mathbf{C}(A)$

(2) $\pi_n : \mathcal{H}_n \to \mathcal{R}$ is measurable $\iff$
$\mathcal{H}_n \ni h_n \mapsto \int_0^\infty \int_A \psi(t,a)\; \pi_n(h_n,t;da)\; dt \in \mathbb{R}$ is measurable $\forall \psi \in \mathbb{X}$.

Proof. With $\mathbf{P}(A)$ and $\mathcal{R}$ being separable and metrizable according to lemma A.20, we can apply lemma B.3 to show these equivalences:

(1) Choose the index set to be $\mathbf{C}(A)$, $X_c = \mathbb{R}$ for all $c \in \mathbf{C}(A)$ and $X = \mathbf{P}(A)$. Choose further g to be r and thus, Y to be $[0,\infty)$. We then can apply lemma B.3 with $[0,\infty) \overset{r}{\to} \mathbf{P}(A) \overset{f_c}{\to} \mathbb{R}, t \mapsto r_t \mapsto \int_A c(a)\; r_t(da)$ to obtain the result. Note that the weak topology on $\mathbf{P}(A)$ is indeed the topology generated by the mappings $f_c : \mathbf{P}(A) \to \mathbb{R}, p \mapsto \int_A c\; dp$ for $c \in \mathbf{C}(A)$ (see Proposition A.7 (b)).

(2) We are looking at the composition of mappings $\mathcal{H}_n \overset{\pi_n}{\to} \mathcal{R} \overset{f_\psi}{\to} \mathbb{R}$ where $\psi \in \mathbb{X}$ and $h_n \mapsto \pi_n(h_n) \mapsto \int_0^\infty \int_A \psi(t,a)\; \pi_n(h_n,t;da)\; dt$. We apply lemma B.3 for the index set $\mathbb{X}$ and note that the Young topology on $\mathcal{R}$ is generated by the mappings $\{f_\psi, \psi \in \mathbb{X}\}$ (see corollary A.22). □

We end this section by giving a characterization of the Young topology and the corresponding σ-algebra based on families of functions that generate the topology or the σ-algebra respectively. We need the following definition:

Definition A.24. *The set of* Charatheodory-functions *on* $\mathbb{R}^+ \times A$ *is the set*

$$\mathbf{Car}\left(\mathbb{R}^+ \times A\right) := \left\{ g \in \mathbf{B}\left(\mathbb{R}^+ \times A\right) \;\middle|\; g(t,\cdot) \text{ continuous } \forall a \in A \right\}.$$

Theorem A.25 (Characterization Theorem). *The following families of functions generate the Young topology and the corresponding* σ*-algebra respectively:*

(1) The Young topology is generated by either one of the following families of functions:

$$(i)\quad \mathcal{R} \ni r \mapsto \int_0^\infty \int_A \psi(t,a)\; r_t(da)\; dt \in \mathbb{R}, \qquad \psi \in \mathbb{X} = L^1(\mathbb{R}^+, \mathbf{C}(A))$$
$$(ii)\quad \mathcal{R} \ni r \mapsto \int_0^\infty e^{-t} \int_A g(t,a)\; r_t(da)\; dt \in \mathbb{R}, \qquad g \in \mathbf{Car}(\mathbb{R}^+ \times A).$$

(2) The corresponding σ*-algebra is generated by either one of the following families of functions:*

$$(i)\quad \mathcal{R} \ni r \mapsto \int_0^\infty \int_A \psi(t,a)\; r_t(da)\; dt \in \mathbb{R}, \qquad \psi \in \mathbb{X} = L^1(\mathbb{R}^+, \mathbf{C}(A))$$
$$(ii)\quad \mathcal{R} \ni r \mapsto \int_0^\infty e^{-t} \int_A g(t,a)\; r_t(da)\; dt \in \mathbb{R}, \qquad g \in \mathbf{Car}(\mathbb{R}^+ \times A)$$
$$(iii)\quad \mathcal{R} \ni r \mapsto \int_0^\infty \int_A \psi(t,a)\; r_t(da)\; dt \in \mathbb{R}, \qquad \psi : \mathbb{R}^+ \times A \to \mathbb{R} \text{ measurable},\ \int \sup_{a\in A} |\psi(t,a)|\; dt < \infty$$
$$(iv)\quad \mathcal{R} \ni r \mapsto \int_0^\infty e^{-t} \int_A g(t,a)\; r_t(da)\; dt \in \mathbb{R}, \qquad g \in \mathbf{B}(\mathbb{R}^+ \times A).$$

Proof. To proof **part (1)**, note that the functions in $(1)(i)$ generate the Young topology according to corollary A.22. Now, $(i) \Rightarrow (ii)$ is obvious as

$$\left\{ (t,a) \mapsto e^{-t} g(t,a) \mid g \in \mathbf{Car}\left(\mathbb{R}^+ \times A\right) \right\} \in \mathbb{X}.$$

We show $(ii) \Rightarrow (i)$:

First, for all $\psi \in \mathbb{X}$ such that there is some $K > 0$ with

$$\|\psi_t\|_\infty \le K\; e^{-t}, \qquad \forall\; t \ge 0,$$

the function $g(t,a) := e^t\, \psi(t,a)$ satisfies $|g(t,a)| \le e^t\, \|\psi_t\|_\infty \le K$ and thus, $g \in \mathbf{B}(\mathbb{R}^+ \times A)$. By definition of $\mathbb{X}$, we have $\psi(t,\cdot) \in \mathbf{C}(A)$ and thus, $g(t,\cdot) = e^t \psi(t,\cdot)$ is continuous for all $a \in A$. This means, we can write

$$\psi(t,a) = e^{-t}\; g(t,a) \qquad \forall\; t \ge 0, a \in A,$$

with $g \in \mathbf{Car}(\mathbb{R}^+ \times A)$ and according to (ii), functions of this type are continuous w.r.t. the Young topology.

For arbitrary $\psi \in \mathbb{X}$, we define the following approximating sequence ψ^n by

$$\psi^n(t,a) := \mathbb{1}_{t \le n}\left((-n) \vee \psi(t,a) \wedge n\right)\ \forall t \ge 0, a \in A.$$

Then, for $n \in \mathbb{N}$ and $K_n := ne^n$, the above defined function ψ^n belongs to the previously considered subset of $\mathbb{X}$ as it satisfies

$$\|\psi^n_t\|_\infty \le K_n e^{-t} \qquad \forall t \ge 0.$$

Furthermore, for $n \geq \|\psi_t\|_\infty$ and $a \in A$, we get

$$|\psi(t,a) - \psi^n(t,a)| = |\psi(t,a) - \mathbb{1}_{t\leq n}\psi(t,a)| = \mathbb{1}_{t>n}\,|\psi(t,a)| \leq \mathbb{1}_{t>n}\,\|\psi_t\|_\infty\,,$$

and as a was arbitrary, this inequality also holds when transitioning to the supremum, thus

$$\|\psi_t - \psi_t^n\|_\infty \leq \mathbb{1}_{t>n}\,\|\psi_t\|_\infty\,.$$

For $n \to \infty$, the right hand side is converging to zero pointwise in $t \geq 0$. With dominated convergence (take $t \mapsto \|\psi_t\|_\infty$), the integral $\int_0^\infty \|\psi_t - \psi_t^n\|_\infty \; dt$ is converging to zero and thus,

$$\psi^n \overset{n\to\infty}{\longrightarrow} \psi \text{ in } \mathbb{X}.$$

This, however, implies the uniform convergence of the mappings

$$r \mapsto \int_0^\infty \psi^n(t,a)\; r_t(da)\; dt,$$

and thus, the mappings in (i) are continuous.

To proof **part (2)**, note that the functions in $(2)(i)$ generate $\mathcal{B}(\mathcal{R})$: According to part (1), they generate the Young topology on $\mathcal{R}$, thus, define a subbase for this topology. As $\mathcal{R}$ is separable and metrizable (see lemma A.20), every subbase of the Young topology generates $\mathcal{B}(\mathcal{R})$. We now show $(i) \Rightarrow (ii) \Rightarrow (iv) \Rightarrow (iii) \Rightarrow (i)$.

Implication $(i) \Rightarrow (ii)$ is obvious as $\{(t,a) \mapsto e^{-t}g(t,a) \mid g \in \mathbf{Car}(\mathbb{R}^+ \times A)\} \subset \mathbb{X}$, implication $(iii) \Rightarrow (i)$ is obvious as $\mathbb{X} \subset \{\psi \mid \psi \text{ measurable, } \int \sup_{a\in A} |\psi(t,a)| \; dt < \infty\}$. The proof for $(iv) \Rightarrow (iii)$ follows the same lines as the proof of $(1)(ii) \Rightarrow (1)(i)$.

Implication $(ii) \Rightarrow (iv)$ follows from the equivalence $(b) \Leftrightarrow (c)$ of theorem B.4. To see this, define a stochastic kernel $p : \mathcal{R} \to \mathbb{P}(\mathbb{R}^+ \times A)$ by setting $p := \mathrm{Exp}(1) \otimes \; r$, i.e., for $g \in \mathbf{B}(\mathbb{R}^+ \times A)$, we get:

$$\int g \; dp(r;\cdot) = \int_0^\infty e^{-t} \int_A g(t,a)\; r_t(da)\; dt.$$

Assuming that (ii) holds, the mappings

$$r \mapsto \int g \; dp(r;\cdot)$$

are measurable for all $g \in \mathbf{Car}(\mathbb{R}^+ \times A)$ and thus, as well for all $g \in \mathbf{C}(\mathbb{R}^+ \times A)$. Now, (iv) follows from theorem B.4, (c) as $\mathcal{R}$ is separable and metrizable according to lemma A.20 □

A.5 Compactness of $\mathcal{R}$ under Young Topology

Base for all our investigations regarding existence of optimal relaxed control strategies for our PO-PDMP will be the following result, stating that the space $\mathcal{R}$ of relaxed control strategies is compact w.r.t. the Young topology.

This compactness result, together with some (semi-) continuity conditions we'll establish later on in this work, will actually guarantee the existence of such optimal relaxed control strategies.

Proposition A.26. *The space $\mathcal{R}$ of relaxed controls is compact w.r.t. the Young topology.*

Proof. For a proof of this result see [23], Proposition 43.3 together with Definition 43.4 and Comment thereafter. □

A.6 Correspondence Theorem

We give here the proof of the correspondence theorem 2.11 that we repeat here:

Theorem A.27 (Correspondence Theorem). *Let $n \in \mathbb{N}_0$. For every $\pi_n^P \in \Pi_n^P$ there exists $\pi_n^D \in \Pi_n^D$ such that*

$$\pi_n^P(h_n, \cdot) = \pi_n^D(h_n)(\cdot) \quad \lambda^1\text{−a.e. on } \mathbb{R}^+ \ \ \forall h_n \in \mathcal{H}_n \tag{A.14}$$

and vice-versa.

For the proof of this theorem we need the following:

Lemma A.28. *The space of observable histories $\mathcal{H}_n$ is a Borel-space for all $n \in \mathbb{N}_0$.*

Proof. The spaces $\mathbb{R}^+$ and E_X are complete, separable and metrizable spaces, thus Borel spaces. The space $\mathcal{R}$ is separable and metrizable (see lemma A.20) and compact (see Proposition A.26), thus $\mathcal{R}$ is complete and therefore a Borel space. Finite products of Borel spaces are again Borel spaces (see [11], proposition 7.13) w.r.t. the product topology and the Borel σ-algebra of the product space coincides with the product σ-algebras and so,

$$\mathcal{H}_0 = \mathbb{R}^+ \times E_X \qquad \text{and} \qquad \mathcal{H}_n = \Big(\mathbb{R}^+ \times E_X \times \mathcal{R}\Big)^n \times \mathbb{R}^+ \times E_X, n \geq 1$$

are Borel spaces. □

We now give the proof of the correspondence theorem

Proof. (Correspondence Theorem). We first show that every π_n^P **has a corresponding** π_n^D: Let $n \in \mathbb{N}_0$ and $\pi_n^P : \mathcal{H}_n \times [0, \infty) \to \mathbf{P}(A)$ measurable.

By lemma A.28, $\mathcal{H}_n$ and thus, $\mathcal{H}_n \times [0, \infty)$ is a Borel space. The compact metric space A is a Borel space as well and thus, we can apply [11], proposition 7.29 and get

$$\lambda_g : \mathcal{H}_n \times [0, \infty) \to \overline{\mathbb{R}}, \qquad \lambda_g(h_n, t) := \int_A \tilde{g}(h_n, t, a)\ \pi_n^P(h_n, t)(da),$$

is measurable for $\tilde{g}(h_n, t, a) := g(t, a)$, where $g \in \mathbf{B}(\mathbb{R}^+ \times A)$.

With λ_g measurable, we also have the measurability of

$$(h_n, t) \mapsto \int_0^\infty e^{-t} \int_A g(t, a)\ \pi_n^P(h_n, t)(da)\ dt,$$

and this implies measurability of

$$\mathcal{H}_n \ni h_n \mapsto \tilde{\lambda_g}(h_n) := \int_0^\infty e^{-t} \int_A g(t, a)\ \pi_n^P(h_n, t)(da)\ dt.$$

With the notation $R_n : \mathcal{H}_n \to \mathcal{R}, h_n \mapsto \pi_n^P(h_n, \cdot)$ and for $g \in \mathbf{B}(\mathbb{R}^+ \times A)$ defining $I_g : \mathcal{R} \to \mathbb{R}, r \mapsto \int_0^\infty e^{-t} \int_A g(t, a)\ r_t(da)\ dt$, we get

$$\tilde{\lambda_g} = I_g \circ R_n : \mathcal{H}_n \xrightarrow{R_n} \mathcal{R} \xrightarrow{I_g} \mathbb{R} \qquad \forall g \in \mathbf{B}(\mathbb{R}^+ \times A).$$

As $\tilde{\lambda_g}$ is measurable for all $g \in \mathbf{B}(\mathbb{R}^+ \times A)$ and the family of mappings $(I_g)_{g \in \mathbf{B}(\mathbb{R}^+ \times A)}$ is generating the Borel-σ-algebra on $\mathcal{R}$ (see theorem A.25, (2)(iv)), the measurability of R_n follows and we can simply select, as corresponding time-discrete history dependent decision rule,

$$\pi_n^D(h_n) := R_n(h_n) = \pi_n^P(h_n, \cdot).$$

To show that every π_n^D **has a corresponding** π_n^P, let $n \in \mathbb{N}_0$ and $\pi_n^D : \mathcal{H}_n \to \mathcal{R}$ measurable. We define a stochastic kernel

$$p : \mathcal{H}_n \to \mathbb{P}(\mathbb{R}^+ \times A),\ p(h_n; \cdot) := \mathrm{Exp}(1) \otimes \pi_n^D(h_n),$$

with other words,

$$\int g(t,a)\ dp(h_n; d(t,a)) = \int_0^\infty e^{-t} \int_A g(t,a)\ \pi_n^D(h_n)(t)(da)\ dt \qquad \forall g \in \mathbf{B}(\mathbb{R}^+ \times A).$$

The measurability of p follows here from theorem B.4, $(c) \Rightarrow (d)$ as follows: For all $g \in \mathbf{B}(\mathbb{R}^+ \times A)$, the mapping

$$h_n \mapsto \int g\ dp(h_n; \cdot) \tag{A.15}$$

□

can be written as

$$h_n \mapsto (I_g \circ \pi_n^D)(h_n),$$

where I_g as defined in the first part of the proof. Now, I_g (generating the σ-algebra on $\mathcal{R}$) and π_n^D are measureable and so is the concatenation of both. The measurability of (A.15) for all $g \in \mathbf{B}(\mathbb{R}^+ \times A)$ implies the measurability of

$$p : \mathcal{H}_n \to \mathbb{P}(\mathbb{R}^+ \times A)$$

according to theorem B.4.

Now applying [11], Corollary 7.27.1 to p and to the Borel spaces $\mathcal{H}_n, \mathbb{R}^+$ and A, there exist Borel-measurable mappings

$$\pi_n^P : \mathcal{H}_n \times [0,\infty) \to \mathbf{P}(A) \qquad \text{and} \qquad \sigma : \mathcal{H}_n \to \mathbb{P}(\mathbb{R}^+),$$

such that we have

$$p(h_n; B_1 \times B_2) = \int_{B_1} \pi_n^P(h_n, t)(B_2)\ \sigma(h_n)(dt) \qquad \forall\ B_1 \in \mathcal{B}(\mathbb{R}^+), B_2 \in \mathcal{B}(A).$$

Here, $\sigma(h_n)(\cdot)$ is the marginal distribution of p, i.e.,

$$\sigma(h_n)(B_1) = p(h_n; B_1 \times A) = \int_{B1} e^{-t}\ dt \qquad \forall B_1 \in \mathcal{B}(\mathbb{R}^+).$$

This disaggregation of p combined with the initial definition of p leads to

$$\begin{aligned}
\int_{B_1} e^{-t}\ \pi_n^D(h_n)(t)(B_2)\ dt &= \int_{B_1} e^{-t}\ \pi_n^P(h_n,t)(B_2)\ dt \qquad \forall\ B_1 \in \mathcal{B}(\mathbb{R}^+), B_2 \in \mathcal{B}(A) \\
\Rightarrow e^{-t}\ \pi_n^D(h_n)(t)(B_2) &= e^{-t}\ \pi_n^P(h_n,t)(B_2) \qquad \lambda^1 - \text{almos all } t \geq 0, \forall\ B_2 \in \mathcal{B}(A) \\
\Rightarrow \pi_n^D(h_n)(t)(B_2) &= \pi_n^P(h_n,t)(B_2) \qquad \lambda^1 - \text{almos all } t \geq 0, \forall\ B_2 \in \mathcal{B}(A)
\end{aligned}$$

As A is separable and metrizable, there exists a countable base $\mathcal{S}$ for the topology $\mathcal{T}[d_A]$ on A (see [11], Proposition 7.1). Thus, every open set in $\mathcal{T}[d_A]$ can be written as a countable union of elements of $\mathcal{S}$ which means $\mathcal{B}(A) = \mathcal{B}(\mathcal{T}[d_A]) = \sigma(\mathcal{S})$ and therefore

$$\Rightarrow\ \pi_n^D(h_n)(t)(B_2) = \pi_n^P(h_n,t)(B_2) \qquad \forall\ B_2 \in \mathcal{S}, \lambda^1 - \text{almos all } t \geq 0.$$

We can assume, w.l.o.g., $\mathcal{S}$ to be closed under finite intersections (see remark A.29), and therefore the theorem for uniqueness of probability measures can be applied to deduct

$$\begin{aligned}
\Rightarrow\ \pi_n^D(h_n)(t)(B_2) &= \pi_n^P(h_n,t)(B_2) \qquad \forall\ B_2 \in \mathcal{B}(A), \lambda^1 - \text{almos all } t \geq 0, \\
\Rightarrow\ \pi_n^D(h_n)(\cdot) &= \pi_n^P(h_n,\cdot) \qquad \lambda^1 - \text{almos all } t \geq 0.
\end{aligned}$$

Remark A.29. *In case, $\mathcal{S}$ is not closed under finite intersections, we can take $\mathcal{S}$ as subbase, i.e. take the following set as a base for the topology:*

$$\mathcal{S}' := \left\{ \bigcap_{j \in J} S_j \middle| J \subset \mathbb{N}, |J| < \infty \right\}.$$

Clearly, $\mathcal{S}'$ is stable under finite intersections and $\mathcal{S}'$ is still countable, as its cardinality is the cardinality of $\left\{ J \subset \mathbb{N} \middle| |J| < \infty \right\}$ which is the same as the cardinality of

$$\bigcup_{k \in \mathbb{N}} \mathbb{N}^k,$$

and this is a countable union of countable sets, thus countable.

Appendix B

Basics and useful results from various mathematical disciplines

In this Annex, we summarize and (partly) give proofs of all results from measurability theory, functional analysis as well as from some other domains of mathematics if these results go beyond basic knowledge in these domains.

B.1 Useful measurability results

Lemma B.1. *Let X a compact metric space endowed with the corresponding Borel-σ-algebra $\mathcal{B}(X)$. For every $n \in \mathbb{N}$, we can then find a finite partition*

$$X = \sum_{i=1}^{k(x)} X_i^n, \tag{B.1}$$

of Borel subsets $X_i^n \in \mathcal{B}(X), X_i^n \neq \emptyset$ with $\operatorname{diam}(X_i^n) \leq \frac{1}{n}$ for $1 \leq i \leq k(n)$.

Proof. Let $n \in \mathbb{N}$ arbitrary but fix. For $x \in X$ we define the open ball around x with radius $\frac{1}{2n}$ by $B(x, \frac{1}{2n})$. Thus, the set $\left\{B(x, \frac{1}{2n}) \mid x \in X\right\}$ is an open coverage of X. As X is compact metric, we can extract a finite open coverage $\left\{B_i^n \mid 1 \leq i \leq \tilde{k}(n)\right\}$. To get the finite partition of X, we define $\tilde{X}_i^n := B_i^n \setminus \cup_{j=1}^{i-1} B_j^n$. We re-enumerate after eliminating all such empty sets $\tilde{X}_i^n$ and get the requested result. □

Lemma B.2. *Let (A, d) a compact metric space and $\psi : [0, \infty) \times A \to \mathbb{R}$ a mapping such that $\psi_t := \psi(t, \cdot) \in \mathbf{C}(A)$ $\forall t \geq 0$. Then, the following properties are equivalent:*

(1) ψ is product-measurable,

(2) the mapping $[0, \infty) \to \mathbb{R}, t \mapsto \psi(t, a)$ is measurable $\forall a \in A$,

(3) the mapping $[0, \infty) \to \mathbf{C}(A), t \mapsto \psi_t$ is measurable.

Proof. **(3) ⇒ (2):** For $a \in A$ define $\delta_a : \mathbf{C}(A) \to \mathbb{R}$ by $\delta_a(f) := f(a)$ $\forall f \in \mathbf{C}(A)$. We first show that for all $a \in A$ this mapping δ_a is continuous and thus measurable. With $f_n \to f$ in $\mathbf{C}(A)$ we have convergence of $\|f_n - f\|_\infty \to 0$ and thus, $f_n(a) \to f(a)$ $\forall a \in A$. With this, continuity of δ_a follows from $\delta_a(f_n) = f_n(a) \to f(a) = \delta_a(f)$.

This property of δ_a together with property (3) lead to the measurability of the composition of mappings

$$t \mapsto \psi_t \overset{\delta_a}{\mapsto} \psi_t(a) \equiv \psi(t,a) \tag{B.2}$$

for all $a \in A$.

(2) ⇒ (1): We show the product measurability of ψ by defining product measurable functions ψ^n converging (pointwise) to ψ. According to lemma B.1, we can find, for every $n \in \mathbb{N}$, a finite partition $A = \sum_{i=1}^{k(n)} A_i^n$. For every $1 \leq i \leq k(n)$, we chose a fix element $a_i^n \in A_i^n$ and define

$$\psi^n(t,a) := \sum_{i=1}^{k(n)} \mathbb{1}_{A_i^n}(a)\psi(t,a_i^n) \ \forall t \geq 0, a \in A. \tag{B.3}$$

By property (2) and the measurability of indicator functions of Borel subsets, this ψ^n is clearly product measurable.

Let now $a \in A$ fix and (i_n) a sequence of indices such that $a \in A_{i_n}^n$ for all $n \in \mathbb{N}$. We then have

(i) $\psi^n(t,a) = \psi(t,a_{i_n}^n) \ \forall n \in \mathbb{N}$,

(ii) $a_{i_n}^n \to a (n \to \infty)$ because $d(a,a_{i_n}^n) \leq \operatorname{diam}(A_{i_n}^n) \leq \frac{1}{n}$.

Bringing this together with the continuity of $\psi(t,\cdot)$ in the second component (remember $\psi_t \in \mathbf{C}(A) \ \forall t \geq 0$), we finally get the pointwise convergence of ψ^n to ψ by

$$\psi^n(t,a) = \psi(t,a_{i_n}^n) \to \psi(t,a) \quad (n \to \infty). \tag{B.4}$$

(1) ⇒ (3): According to lemma A.1 A is separable, thus a countable dense subset $A' \subset A$ exists. Let's further choose $f_0 \in \mathbf{C}(A)$ and $\epsilon > 0$. It is then enough to show that $\psi_t^{-1}(B(f_0,\epsilon))$ is measurable for

$$B(f_0,\epsilon) := \{f \in \mathbf{C}(A) \mid \|f - f_0\|_\infty \leq \epsilon\}.$$

This follows from

$$\begin{aligned}
\{t \geq 0 \mid \psi(t,\cdot) \in B(f_0,\epsilon)\} &= \left\{t \geq 0 \mid \sup_{a\in A} |\psi(t,a) - f_0(a)| \leq \epsilon\right\} \\
&= \left\{t \geq 0 \mid \sup_{a\in A'} |\psi(t,a) - f_0(a)| \leq \epsilon\right\} \\
&= \bigcap_{a\in A'} \{|\psi(\cdot,a) - f_0(a)| \leq \epsilon\}.
\end{aligned}$$

□

Lemma B.3. *Let I an arbitrary index set and $(X,\mathcal{T})$ as well as $(X_i,\mathcal{T}_i)$ topological spaces where $\mathcal{T} = \mathcal{T}[f_i, i \in I]$ is the topology generated by the mappings $f_i : X \to X_i, i \in I$. Denote by σ_X and σ_{X_i} the Borel-σ-algebras on X and X_i respectively.*

If X is separable metrizable, then, for every measurable space (Y,σ_Y) and every mapping $g : Y \to X$, it holds:

$$g \text{ is } \sigma_Y - \sigma_X - \text{ measurable } \Leftrightarrow f_i \circ g \text{ is } \sigma_Y - \sigma_{X_i} - \text{measurable } \forall i \in I \tag{B.5}$$

Proof. **"⇒":** By definition of $\mathcal{T}$ as topology generated by $\{f_i; i \in I\}$, every $f_i : X \to X_i$ is $\mathcal{T}$-$\mathcal{T}_i$-continuous and thus σ_X-σ_{X_i}-measurable. By composition with g the result follows.

"⇐": By definition of $\mathcal{T}$ being the topology generated by $\{f_i; i \in I\}$, the following set is a base of $\mathcal{T}$:

$$\left\{ \bigcap_{k=1}^{n} f_{i_k}^{-1}(O_{i_k}) \mid n \in \mathbb{N}, i_k \in I, O_{i_k} \in \mathcal{T}_{i_k} \right\}. \tag{B.6}$$

As X is separable metrizable, every base for $\mathcal{T}$ contains a countable subcollection that is also a base for $\mathcal{T}$ (see [11], Proposition 7.1). Thus, every open set $O \subset X$ can be written as

$$O = \bigcup_{j=1}^{\infty} \bigcap_{k=1}^{n_j} f_{i_{j,k}}^{-1}(O_{i_{j,k}}), \tag{B.7}$$

where $n_j \in \mathbb{N}, i_{j,k} \in I$ and $O_{i_{j,k}} \in \mathcal{T}_{i_{j,k}}$. As a consequence, we can write $g^{-1}(O)$ as

$$g^{-1}(O) = \bigcup_{j=1}^{\infty} \bigcap_{k=1}^{n_j} (f_{i_{j,k}} \circ g)^{-1}(O_{i_{j,k}}). \tag{B.8}$$

By assumption, $f_{i_{j,k}} \circ g$ are all measurable and $\mathcal{T}_{i_{j,k}} \subset \sigma_{X_{i_{j,k}}}$ and thus, $g^{-1}(O) \in \sigma_Y$ which means $g^{-1}(\mathcal{T}) \subset \sigma_Y$ and thus, $g^{-1}(\sigma_X) \subset \sigma_Y$. □

Theorem B.4. *Let (X, σ_X) a measurable space and Y a separable and metrizable space with Borel-σ-algebra $\mathcal{B}(Y)$. Then, the following measurability statements are equivalent:*

(a) $p : X \to \mathbb{P}(Y)$ is a σ_X-measurable stochastic kernel,

(b) $x \mapsto \int f \, dp(x, \cdot)$ is σ_X-measurable $\forall f \in \mathbf{C}(Y)$,

(c) $x \mapsto \int f \, dp(x; \cdot)$ is σ_X-measurable $\forall f \in \mathbf{B}(Y)$,

(d) $x \mapsto p(x; B)$ is σ_X-measurable $\forall B \in \mathcal{B}(Y)$.

Proof. We show $(d) \Rightarrow (a) \Rightarrow (b) \Rightarrow (c) \Rightarrow (d)$.

$(d) \Rightarrow (a)$ follows from [11], proposition 7.26. As Y is separable and metrizable, so is $\mathbb{P}(Y)$ (see, e.g., [11], proposition 7.20), and thus, $(a) \Leftrightarrow (b)$ follows from theorem B.3. Having this equivalence between (a) and (b), the implication $(b) \Rightarrow (c)$ is shown, if $(a) \Rightarrow (c)$ holds. This latter implication follows from [11], proposition 7.29. Finally, $(c) \Rightarrow (d)$ is obvious.

Looking closer and the cited propositions of [11], one will notice that these propositions are formulated for Borel spaces. For our implications though, we do not need the completeness of the spaces, thus, separable and metrizable spaces are sufficient. □

B.2 Other useful technical results

Lemma B.5. *Let X a separable and metrizable space, Y a compact metric space and $f : X \times Y \to \mathbb{R}$ continuous. Then, the following continuity property holds:*

$$x_n \to x \Longrightarrow \sup_{y \in Y} |f(x_n, y) - f(x, y)| \to 0 \ (n \to \infty).$$

Proof. Let $x_n \to x$ and assume there is $\epsilon > 0$ and a subsequence (x_{n_k}) of (x_n) such that

$$\sup_{y \in Y} |f(x_{n_k}, y) - f(x, y)| > \epsilon \ \forall k \in \mathbb{N}.$$

This implies, that for all $k \in \mathbb{N}$, one can find $y_k \in Y$ with

$$|f(x_{n_k}, y_k) - f(x, y_k)| > \epsilon \ \forall k \in \mathbb{N}.$$

By compactness of Y, there exists a convergent subsequence (y_{k_l}) of (y_k) with $y_{k_l} \to y_\infty \in Y$ $(l \to \infty)$. By continuity of f we then get

$$\epsilon < |f(x_{n_{kl}}, y_{k_l}) - f(x, y_{k_l})| \to |f(x, y_\infty) - f(x, y_\infty)| = 0,$$

which is a contradiction to $\epsilon > 0$. □

Appendix C

Summary of model assumptions

The following Assumptions and Definitions were made throughout the development of the general PO-PDMP theory and are sufficient for Theorem 2.51 (existence of optimal policies for the N-stage problem) and Theorem 2.60 (existence of optimal policies for the infinite time horizon problem) to hold:

(SP) ***Polish state space:*** *(see Definition 1.1)*

The state space E_Y is assumed to be a Polish space.

(SF) ***Finite set of post-jump states:*** *(see Assumption 2.16)*

We assume the set E_Y^0 of possible post-jump states to be finite, i.e. $\exists q \in \mathbb{N}$: $E_Y^0 := \{y_1, \dots, y_q\} \subset E_Y$ and $Q_Y^A(y, a; E_Y^0) = 1$ for all $y \in E_Y, a \in A$. We further assume $Y_0 \in E_Y^0$.

(OS) ***Observation space:*** *(see Definition 1.8)*

Let $(E_X, +, 0_X)$ a Polish space endowed with a commutative group structure with neutral element 0_X and $\psi : E_Y \to E_X$ a homeomorphism of topological spaces. We call $E_{SX} := \mathbb{R}^+ \times E_X$ the *observation space*.

(ON) ***Observation noise:*** *(see Definition 1.9 and Assumption 1.13)*

(i) Let $(\epsilon_n)_{n\geq 0}$ a sequence of E_X-valued i.i.d. random variables $\epsilon_n : \Omega \to E_X$, that are independent from $(S_n, Z_n)_{n\geq 0}$. We call ϵ_n *observation noise* and denote its distribution by Q_ϵ.

(ii) We assume the distribution Q_ϵ of the noise ϵ_n to have a bounded density function $f_\epsilon : E_X \to \mathbb{R}$ with respect to some σ-finite measure ν on $(E_X, \mathcal{B}(E_X))$, i.e. $Q_\epsilon(B) = \int_B f_\epsilon(x)\nu(dx)$ for all $B \in \mathcal{B}(E_X)$.

(A) ***Action space:*** *(see Definition 1.20)*

Let A a compact metric space.

(D) ***Drift:*** *(see Definition 1.29)*

Let $[r] \in \mathcal{R}$ an arbitrary relaxed control. We assume that the drift Φ^r of the controlled PDMP (Y_t^r) is a continuous mapping $\Phi^r : E_Y \times \mathbb{R}^+ \to E_Y$ and satisfies:

(i) The mapping $t \mapsto \Phi^r(\cdot, t)$ is a semi-group, i.e. for all $y \in E_Y$:

$$\Phi^r(y, s + t) = \Phi^r(\Phi^r(y, s), t).$$

(ii) The controlled drift Φ^r is λ^1-a.e. independent of the choice of a representative of $[r]$, i.e., for all $r' \in [r]$ and all $y \in \mathbb{E}_Y$:

$$\Phi^r(y,t) = \Phi^{r'}(y,t) \text{ for } \lambda^1-\text{almost all } t \in [0,\infty).$$

(DC) ***Continuous dependence of drift on relaxed control:*** *(see Assumption 3.1)*

We assume that $\mathcal{R} \ni r \mapsto \Phi^r(y,t) \in E_Y$ is continuous for all $y \in E_Y^0$ and all $t \geq 0$.

(IC) ***Continuous and bounded intensity:*** *(see Definition 1.32)*

Let $\lambda^A : E_Y \times A \to (0,\infty)$ a continuous and bounded function.

(IL) ***Limit property of intensity:*** *(see Definition 1.32)*

The intensity λ^A satisfies the following limit property:

$$\forall n \in \mathbb{N}, \pi_n \in \Pi_n^P, h_n \in \mathcal{H}_n, y \in E_Y :$$

$$\lim_{t\to\infty} \int_0^t \int_A \lambda^A(\Phi^{\pi_n(h_n,\cdot)}(y,s),a)\ \pi_n(h_n,s)(da)\ ds = \infty.$$

(QC) ***Weak continuity of transition kernel:*** *(see Definition 1.33)*

Let $Q^A : E_Y \times A \to \mathbf{P}(E_Y^0)$ a weakly continuous transition kernel.

(CC) ***Lower semi-continuity of cost function:*** *(see Assumption 3.2)*

We assume the cost function $c : E_X \times E_Y \times A \to \mathbb{R}^+$ to be lower semi-continuous w.r.t. the product topology.

(IQ) ***Uncontrolled intensity and transition kernel:*** *(see Assumption 3.9)*

We assume the jump intensity as well as the transition kernel for the unobservable state of the PO-PDMP to be uncontrolled, i.e. there exists $a_0 \in A$ such that for all $a \in A, y^j \in E_Y^0, y \in E_Y$:

$$\lambda^A(y,a) = \lambda^A(y,a_0) \qquad \text{and} \qquad Q^A(y,a;\{y^j\}) = Q^A(y,a_0;\{y^j\}).$$

Frequently used notations

Integers and real numbers

$\mathbb{N}$	Set of positive integers	11
$\mathbb{N}_0$	Set of positive integers including zero	11
$\mathbb{R}$	Set of real numbers	97
$\mathbb{R}^+$	Set of non negative real numbers	10
$\overline{\mathbb{R}}$	$\mathbb{R} \cup \{\infty\}$	10

State spaces, action spaces, observable histories

E_X	State space of observable component, see Definition 1.8	15
E_Y	State space of unobservable component, see Definition 1.1	10
E_Y^0	Space of possible post-jump states, see Assumption 2.16	42
$\mathbf{P}(E_Y^0)$	Space of probability measures on E_Y^0, see Definition 2.18	42
E_{SXY}	State space of pseudo-embedded process, see Definition 2.6	37
E'	State space of derived filtered process, see Definition 2.32	48
A	Action space, see Definition 1.20	21
$\mathbf{P}(A)$	Relaxed action space, see Definition 1.21	21
$\mathcal{R}$	Space of relaxed controls, see Definition 1.25	23
$\mathcal{U}$	Space of deterministic controls, see Remark 5.2	97
$\mathcal{H}_n$	Space of observable histories up to n-th jump time, see Definition 1.27	24

Function spaces and operators

$\mathbf{C}(A)$	Space of continuous and bounded functions from A to $\mathbb{R}$, see Section A.1.1	128
$\mathbb{X}$	Space $L^1([0,\infty), \mathbf{C}(A))$, see Definition A.10	130
$\hat{\mathbf{B}}^+(E')$	Space of non negative measurable functions on E', see Definition 2.41	59
$\hat{\mathbf{C}}^+_{low}(E')$	Space of lower semi-continuous functions on E', see Definition 2.41	59
H	Cost operator, partially observable model, see Definition 2.42	59
$\mathcal{T}_f$	Cost operator of decision rule f, partially observable model, see Definition 2.42	59
$\mathcal{T}$	Minimum cost operator, partially observable model, see Definition 2.42	59
K	Cost operator, completely observable model	100
$\mathcal{L}_f$	Cost operator of decision rule f, completely observable model	100
$\mathcal{L}$	Minimum cost operator, completely observable model	100

Policies and controls

Π^P, Π^P_n	Set of history dependent relaxed piecewise open loop policies for PO-PDMP (for period $[T_n; T_{n+1})$), see Definition 1.27	24
π^P_n	Decision rule for PO-PDMP for period period $[T_n; T_{n+1})$, see Definition 1.27	24
Π^D, Π^D_n	Set of history dependent relaxed piecewise open loop policies for pseudo-embedded process (for at stage n), see Definition 2.8	37
π^D_n	Decision rule for pseudo-embedded process at stage n, see Definition 2.8	37
Π^0	Set of (stationary) decision rules for deterministic controls, see Definition 5.11	103
π^0	Decision rule (stationary) for deterministic controls, see Definition 5.11	103
$r \in \mathcal{R}$	Relaxed control, see Definition 1.25	23
$u \in \mathcal{U}$	Deterministic control, see Remark 5.2	97

Characteristics and parameters of PO-PDMP

Φ	Drift of underlying PDMP, see Definition 1.1	10
Φ^π, Φ^r	Controlled drift, see Assumption 1.29	24
λ, λ^A	Jump intensity (uncontrolled and controlled), see Definitions 1.1 and 1.32	26
Λ^r	Abbreviation, cumulative intensity	73
Q, Q^A	Jump transition kernel (uncontrolled and controlled), see Definitions 1.1 and 1.33	26
Q_0	Initial conditional distribution of unobservable state, see Assumption 1.19	18
Q_ϵ, f_ϵ	Distribution and density of measurement noise, see Assumption 1.13	17
ν	A σ-finite measure on E_X, see Assumption 1.13	17
q	Cardinality of E^0_Y, see Assumption 2.16	42
$f^\pi_{T_1}(t \mid y)$	Density of first jump-time given $Y_0 = y$, see Lemma 1.35	27
$\mathbb{P}_y$	Probability measure on $(\Omega, \mathcal{F})$ given $Y_0 = y$, see Definition 1.1	10
$\eta^r(y,t)$	$= e^{-\beta t - \Lambda^r(y,t)}$	73

Transition laws and process states

Q_{SXY}	Transition law of embedded process of PO-PDMP, see Lemma 2.4	35
$\tilde{Q}_{SXY}$	Transition law of pseudo-embedded process, see Definition 2.6	37
Q'_{SXM}	Transition law of derived filtered process, see Definition 2.32	48
$\hat{Q}'_{SXM}$	Transition law of derived filtered process under filter $\hat{\chi}$, see (4.9)	91

S_n, X_n, Y_n	Components of embedded process of PO-PDMP	34
$\tilde{S}_n, \tilde{X}_n, \tilde{Y}_n$	Components of pseudo-embedded process, see Definition 2.6	37
M_n	Filter, see Definition 2.21	42
h_n	Observable history up to stage n	24
y^i	Possible post-jump state, element of E_Y^0, see Assumption 2.16	42
χ	Filter equation, see Definition 2.30	46
$\hat{\chi}$	Filter equation for unobservable inter-jump time, see Assumption 4.5	89
$\mu_n(h_n)$	Factorization of filter, see Remark 2.23	43

Value functions and parameters of optimization problems

J	Value function of initial optimization problem	28
$\tilde{J}$	Value function for pseudo-embedded process	40
J'	Value function for derived filtered process	50
J^N	Value function up to stage N	62
J_n^N	Value function up to stage N if current state is n	63
$J'^N_{\pi n}$	As above for derived filtered process	63
v	Value function for completely observable PDMP control problem	101
c	Running cost function	28
g	One period cost function for PO-PDMP	33
G	Discounted one period cost function for PO-PDMP	33
g'	One step cost function for derived filtered process	48
G'	Discounted one step cost function for derived filtered process	48
β	Discount factor	28

Index

Bibliography

[1] A. Almudevar. A dynamic programming algorithm for the optimal control of piecewise deterministic Markov processes. *SIAM J. Control Optim.*, 40(2):525–539, 2001.

[2] A. Arapostathis, V.S. Borkar, E. Fernández-Gaucherand, M.K. Ghosh, and S.I. Marcus. Discrete-time controlled Markov processes with average cost criterion: a survey. *SIAM Journal on Control an Optimization*, 31(2):282–344, 1993.

[3] A. Bain and D. Crisan. *Fundamentals of stochastic filtering.* Springer, New York, 2009.

[4] N. Bäuerle. Convex stochastic fluid programs with average cost. *Journal of mathematical analysis and applications*, 259:137–156, 2001.

[5] N. Bäuerle. Discounted stochastic fluid programs. *Mathematics of Operations Research*, 26:401–420, 2001.

[6] N. Bäuerle and U. Rieder. *Markov Decision Processes with Applications to Finance.* Springer-Verlag, 2010.

[7] E. Bayraktar and M. Ludkovski. Inventory management with partially observed nonstationary demand. *Ann. Oper. Res.*, 176:7–39, 2010.

[8] R. Bellman. *Dynamic programming.* Princeton University Press, Princeton, NJ, 1957.

[9] R. Bellman. *Dynamic programming.* Dover Publications, Mineola, NY, 2003.

[10] A. Bensoussan and J.L. Lions. Nouvelles méthodes en contrôle impulsionnel. *Applied Math. and Optimization*, 1:289–312, 1975.

[11] D.P. Bertsekas and E. Shreve. *Stochastic Optimal Control: The Discrete Time Case.* Academic Press, New York, 1978.

[12] P. Billingsley. *Convergence of probability measures.* Wiley series in probability and mathematical statistics. Wiley, New York [u.a.], 1968.

[13] A. Brandejsky, B. de Saporta, and F. Dufour. Optimal stopping for partially observed piecewise-deterministic Markov processes. *Stochastic Process. Appl.*, 123(8):3201–3238, 2013.

[14] P. Brémaud. *Point Processes and Queues: Martingale Dynamics.* Springer Series in Statistics. Springer, New York, 1981.

[15] D. Chafaï, F. Malrieu, and K. Paroux. On the long time behavior of the tcp window size process. *Stochastic Process. Appl.*, 120(8):1518–1534, 2010.

[16] K. Chernysh. Average cost optimization for a power supply management model. *Modern Trends in Controlled Stochastic Processes*, 2:307–323, 2015.

[17] K. Chernysh. *Stochastic average-cost control, with energy-related applications.* PhD Thesis, Heriot-Watt University, 2016.

[18] E. Çinlar and J. Jacod. *Representation of semimartingale Markov processes in terms of Wiener processes and Poisson random measures*, volume 1 of *Progr. Prob. Statist.* Birkhäuser Boston, Mass., 1981.

[19] O.L.V. Costa and F. Dufour. Average continuous control of piecewise deterministic Markov processes. *SIAM J. Control Optim.*, 48(7):4262–4291, 2010.

[20] O.L.V. Costa and F. Dufour. *Continuous Average Control of Piecewise Deterministic Markov Processes*. Springer Briefs in Mathematics, 2010.

[21] O.L.V. Costa and C.A.B. Raymundo. Impulse and continuous control of piecewise deterministic Markov processes. *Stochastics Stochastics Rep.*, 70(1-2):75–107, 2000.

[22] M.H.A. Davis. Piecewise-deterministic markov processes: A general class of non-diffusion stochastic models. *Journal of the Royal Statistical Society B*, 46:353–388, 1984.

[23] M.H.A. Davis. *Markov Models and Optimization*. Chapman and Hall, London, 1993.

[24] M.H.A. Davis, M.A.H. Dempster, S.P. Sethi, and D. Vermes. Optimal capacity expansion under uncertainty. *Adv. in Appl. Probab.*, 19(1):156–176, 1987.

[25] B. de Saporta, F. Dufour, and K. Gonzalez. Numerical method for optimal stopping of piecewise deterministic Markov processes. *Ann. Appl. Probab.*, 20(5):1607–1637, 2010.

[26] M.A.H. Dempster. Optimal control of piecewise deterministic processes. *Applied Stochastic Analysis*, pages 303–325, 1991.

[27] M.A.H. Dempster and J.J. Ye. Necessary and sufficient optimality conditions for control of piecewise deterministic markov processes. *Stochastics Stochastics Rep.*, 40:125–145, 1992.

[28] J. Diestel. *Sequences and Series in Banach Spaces*. Springer-Verlag, Graduate Texts in Mathematics, New York, 1984.

[29] N. Dunfort and J.T. Schwartz. *Linear Operators - Part 1*. Interscience Publishers Inc., New York, 1976.

[30] E.B. Dynkin and A.A. Juškevič. *Controlled Markov processes*. Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen ; 235. Springer, Berlin [u.a.], 1979.

[31] R.J. Elliott, L. Aggoun, and J.B. Moore. *Hidden Markov models*. Springer-Verlag, New York, 1995.

[32] E. Feinberg. *Handbook of Markov decision processes : methods and applications*. International series in operations research and management science ; 40. Kluwer Academic, Boston, 2002.

[33] E.A. Feinberg, P.O. Kasyanov, and M.Z. Zgurovsky. Partially observable total-cost markov decision processes with weakly continuous transition probabilities. *Mathematics of Operations Research*, 41(2):656–681, 2016.

[34] L. Forwick. *Optimale Kontrolle stückweise deterministischer Prozesse*. Dissertationsschrift, Rheinische Friedrich-Wilhelms-Universität Bonn, 1997.

[35] L. Forwick, M. Schäl, and Schmitz M. Piecewise deterministic markov control processes with feedback controls and unbounded costs. *Acta Appl. Math.*, 82(3):239–267, 2004.

[36] B. Fristedt, N. Jain, and N. Krylov. *Filtering and prediction: a primer*. American Mathematical Society, Providence, RI, 2007.

[37] U.S. Gugerli. Optimal stopping of a piecewise-deterministic Markov process. *Stochastics*, 19(4):221–236, 1986.

[38] O. Hernández-Lerma. *Adaptive Markov Control Processes*. Springer-Verlag, New York, 1989.

[39] O. Hernández-Lerma and R. Romera. Limiting discounted-cost control of partially observable stochastic systems. *SIAM Journal on Control and Optimization*, 40(2):348–369, 2001.

[40] J.P. Hespanha. A model for stochastic hybrid systems with applications to communication networks. *Nonlinear Analysis*, 62(8):1353–1383, 2005.

[41] K. Hinderer. *Foundations of non-stationary dynamic programming with discrete time parameter*. Lecture notes in operations research and mathematical systems ; 33. Springer, Berlin, 1970.

[42] A. Hordijk and F.A.V.D.D. Schouten. Average optimal policies in Markov decision drift processes with applications to queueing and a replacement model. *Advances in Applied Probability*, pages 274–303, 1983.

[43] R.A. Howard. *Dynamic programming and Markov processes*. M.I.T. Pr., Cambridge, Mass., 1960.

[44] M. Jacobsen. *Point Process Theory and Applications - Marked Point and Piecewise Deterministic Processes*. Probability and Its Applications. Birkhäuser, Boston, 2006.

[45] O. Kallenberg. *Foundations of modern probability - second edition*. Springer, New York, 2002.

[46] M. Kirch and W.J. Runggaldier. Efficient hedging when asset prices follow a geometric poisson process with unknown intensities. *SIAM J. Control Optim.*, 43(4):1174–1195, 2005.

[47] H.J. Kushner and P. Dupuis. *Numerical methods for stochastic control problems in continuous time*. Applications of mathematics ; 24. Springer, New York, 2001.

[48] G. Last and A. Brandt. *Marked Point Processes on the Real Line: The dynamic Approach*. Springer, New York, 1995.

[49] J. Lygeros, K. Koutroumpas, S. Dimopolous, P. Legouras, C. Heichinger, P. Nurse, and Z. Lygerou. Stochastic hybrid modeling of dna replication across a complete genome. *Proceedings of the National Academy of Sciences*, 105(34):12295–12300, 2008.

[50] K. Pakdaman, M. Thieullen, and G. Wainrib. Fluid limit theorems for stochastic hybrid systems with application to neuron models. *Adv. in Appl. Probab.*, 42(3):761–794, 2010.

[51] K.R. Parthasarathy. *Probability measures on metric spaces*. Probability and mathematical statistics ; 3. Acad. Press, New York [u.a.], 1967.

[52] H. Pham, W. Runggaldier, and A. Sellami. Approximation by quantization of the filter process and applications to optimal stopping problems under partial observation. *Monte Carlo Methods Appl.*, 11(1):57–81, 2005.

[53] M.L. Puterman. *Markov decision processes : discrete stochastic dynamic programming*. Wiley series in probability and mathematical statistics : Applied probability and statistics. A Wiley-interscience publication. Wiley, New York [u.a.], 1994.

[54] D. Rhenius. Incomplete information in markovian decision models. *The Annals of Statistics*, 2(6):1327–1334, 1974.

[55] H.L. Royden. *Real analysis*. MacMillan, New York [u.a.], 2. ed. edition, 1968.

[56] Y. Sawaragi and T. Yoshikawa. Discrete-time markovian decision processes with incomplete state observation. *The Annals of Mathematical Statistics*, 41(1):78–86, 1970.

[57] M. Schäl. On piecewise deterministic Markov control processes: control of jumps and of risk processes in insurance. *Insurance Math. Econom.*, 22(1):75–91, 1998.

[58] A.N. Shiryaev. Some new results in the theory of controlled random processes. *Trans. Fourth Prague Conf. on Information Theory, Statistical Decision Functions, Random Processes*, pages 131–203, 1965.

[59] E.J. Sondik. The optimal control of partially observable markov processes over the infinite horizon: Discounted costs. *Operations Research*, 26(2):282–304, 1978.

[60] J. Warga. *Optimal Control of Differential and Functional Equations*. Academic Press, 1972.

[61] A. A. Yushkevich. On reducing a jump controllable markov model to a model with discrete time. *Theory Probab. Appl.*, 25:58–69, 1980.

[62] A. A. Yushkevich. Verification theorems for markov decision processes with controllable deterministic drift, gradual and impulse controls. *Teor. Veroyatnost. i Primenen*, 34:528–551, 1989.

[63] A.A. Yushkevich. Reduction of a controlled markov model with incomplete data to a problem with complete information in the case of borel state and control space. *Theory of Probability & Its Applications*, 21(1):153–158, 1976.

[64] A.A. Yushkevich. Bellman inequalities in markov decision deterministic drift processes. *Stochastics*, 23:25–77, 1987.