

OPTIMAL CONTROL OF PARTIALLY OBSERVABLE PIECEWISE DETERMINISTIC MARKOV PROCESSES

NICOLE BÄUERLE* AND DIRK LANGE†

ABSTRACT. In this paper we consider a control problem for a Partially Observable Piecewise Deterministic Markov Process of the following type: After the jump of the process the controller receives a noisy signal about the state and the aim is to control the process continuously in time in such a way that the expected discounted cost of the system is minimized. We solve this optimization problem by reducing it to a discrete-time Markov Decision Process. This includes the derivation of a filter for the unobservable state. Imposing sufficient continuity and compactness assumptions we are able to prove the existence of optimal policies and show that the value function satisfies a fixed point equation. A generic application is given to illustrate the results.

KEY WORDS : Partially Observable Piecewise Deterministic Markov Process, Markov Decision Process, Filter, Updating-Operator
 AMS SUBJECT CLASSIFICATIONS: Primary 60J25, 90C40 Secondary 93E11

1. INTRODUCTION

Piecewise Deterministic Markov Processes (PDMP) 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.

Classical optimization problems can be formulated for PDMPs such as reward maximization or cost minimization. Minimum expected average cost problems (see, e.g., [2], [10] or [11]) as well as minimum expected total discounted cost problems (see e.g. [1], [15], [18]) have intensively been treated for PDMP control problems. Optimal policies are in general relaxed controls, i.e. a control action is a probability distribution on the action space. The idea of reducing the continuous time control problem of a PDMP to a discrete time Markov Decision Process (MDP) is due to Yushkevich, see [30]. 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 insurance [29], communication networks [9], reliability [16], neurosciences [27] and biochemics [25] to only list a very short overview that illustrates the huge variety of domains of application.

In terms of pure mathematical treatment of PDMP control problems, the status up to 1993 can be found in [14]. Since then, important steps in the further development of this theory were, amongst others: In [12] the authors 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. [18] then considers, in contrast to the earlier works, problems with only locally bounded running cost functions. They show absolute continuity for the value function and that the value

function is a (weak) solution of the Hamilton-Jacobi-Bellmann equation. In addition, they derive sufficient conditions for the existence of optimal deterministic feedback controls.

Later, with [28] and [7] 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. Remarkable about the paper [7] is, however, that they treat an optimal stopping problem for a PDMP under partial observation. Such a setting is also considered in [26] where a replacement problem under partial information is considered. Whereas in [7] new information is only received after a jump, the information in [26] is received via monitoring at equidistant inspection times. Besides these papers there are only very few works treating PDMP control problems under partial observation. In [23], a special convex hedging problem on a financial market with price processes following a geometric Poisson-distribution is considered. In the second part of this work, partial observation is modeled by assuming an unknown jump intensity. In [5], 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, the authors of [7] suggest to model partial observation by assuming only noisy measurement of the post-jump state of the PDMP which for other times than jump times, is assumed completely unobservable. Stopping, however is a very special control problem with only two control actions: stop or continue.

In this paper the first aim is to define a general model of a controlled PDMP under partial observation with the discounted cost criterion. We assume as in [7] that the controller receives a noisy measurement of the post-jump state of the PDMP. Then we show how this continuous-time control problem can be reduced to a classical discrete-time MDP with a state space consisting of probability measures. This involves the derivation of a filter for the unobservable state. We next impose some continuity and compactness assumptions along with the introduction of a regularized filter in order to guarantee the existence of optimal policies. A problem which is known to be notoriously difficult (see e.g. [17]). Finally the value function of the optimization problem is shown to be a fixed point of an operator and the minimizer of the value function defines a stationary optimal policy.

Our paper is organized as follows: In the next section we briefly introduce the notation of an uncontrolled PDMP with partial observation. In Section 3 we add controls. The optimization problem itself is explained in Section 4. In the same section we show how to reduce the problem to a Partially Observable Markov Decision Process and derive the corresponding filter. Afterwards in Section 5 we present the optimality equation for the value function and prove existence of optimal controls under our assumptions. A generic application is given in Section 6.

2. AN UNCONTROLLED PDMP WITH PARTIAL OBSERVATION

In [13] the class of PDMPs has been introduced as a *general class of non-diffusion stochastic models*. A definition of a PDMP based on its infinitesimal generator is given there and thus strongly emphasizing the fact that a PDMP is a priori a continuous-time process. Recent publications such as [7] or [18] introduce a PDMP following an axiomatic approach stating a set of properties of a PDMP. We will follow the latter approach in this paper.

We first define an uncontrolled PDMP with partial observation before we consider in Section 3 controlled PDMPs under partial observation. An informal description of a PDMP with partial observation is as follows: The process $(Y_t)_{t \geq 0}$ with values in \mathbb{R}^d first evolves in a deterministic way according to a certain drift Φ . The *drift* $\Phi : \mathbb{R}^d \times \mathbb{R}_+ \rightarrow \mathbb{R}^d$ is continuous and the mapping $t \mapsto \Phi(\cdot, t)$ is a semi-group with respect to concatenation of mappings, i.e. for all $y \in \mathbb{R}^d$ and $s, t > 0$:

$$\Phi(y, t + s) = \Phi(\Phi(y, s), t). \quad (2.1)$$

$\Phi(y, t)$ is the state of the process t time units after the last jump when the state directly after the jump was y . Often in applications the drift Φ is given by a differential equation

$$\frac{d}{dt}\Phi(y, t) = b(\Phi(y, t)), \quad \Phi(y, 0) = y \quad (2.2)$$

where $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a vector field guaranteeing for all $y \in \mathbb{R}^d$ a unique componentwise continuous solution.

At the random time T_1 the process jumps unpredictably to a new state where the deterministic evolution continues until the next jump occurs. The jump times $0 := T_0 < T_1 < \dots$ are \mathbb{R}_+ -valued random variables such that $S_n := T_n - T_{n-1}, n \in \mathbb{N}, S_0 := 0$ and $T_n < T_{n+1}$ if $T_n < \infty$ else $T_n = T_{n+1}$. The jump times are generated by a *jump rate* or *intensity* $\lambda : \mathbb{R}^d \rightarrow (0, \infty)$ which is a measurable mapping of the state. A *transition kernel* Q from \mathbb{R}^d to \mathbb{R}^d describes the probability $Q(B|y)$ that the process jumps into set B given the state before the jump is y .

We assume now that the state of the PDMP cannot be observed directly. 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, the idea is to assume that one can at least measure (or estimate) the true state of the system with some measurement noise. We assume that at jump times of the PDMP we receive new information about the state. More precisely let $(\epsilon_n)_{n \in \mathbb{N}}$ be a sequence of \mathbb{R}^d -valued independent and identically distributed random variables $\epsilon_n : \Omega \rightarrow \mathbb{R}^d$ that are independent from all other random variables. We call ϵ_n *observation noise* and denote its distribution by Q_ϵ . We assume that the agent is able to observe $X_n := Y_{T_n} + \epsilon_n$ directly after the jump at time T_n .

Given the data $(\Phi, \lambda, Q, Q_\epsilon)$, an initial state y and its observation x there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P}_{x,y})$ carrying the random variables $(T_n), (Y_{T_n})$ and (ϵ_n) such that $\mathbb{P}_{x,y}(Y_0 = y, X_0 = x) = 1$ and for all $n \in \mathbb{N}, t \geq 0, C, D \in \mathcal{B}_d$, where \mathcal{B}_d is the σ -algebra of Borel sets in \mathbb{R}^d , it holds that

$$\begin{aligned} & \mathbb{P}_{x,y}(S_n \leq t, Y_{T_n} \in C, X_n \in D \mid S_0, Y_{T_0}, X_0, \dots, S_{n-1}, Y_{T_{n-1}}, X_{n-1}) \\ &= \mathbb{P}_{x,y}(S_n \leq t, Y_{T_n} \in C, X_n \in D \mid Y_{T_{n-1}}) \\ &= \int_0^t \int_C \mathbb{P}_y(X_n \in D \mid Y_{T_n} = y') \\ & \quad \mathbb{P}_{x,y}(ds, dy' \mid Y_{T_{n-1}}) \\ &= \int_0^t \int_C Q_\epsilon(D - y') \exp(-\Lambda(Y_{T_{n-1}}, s)) \lambda(\Phi(Y_{T_{n-1}}, s)) Q(dy' \mid \Phi(Y_{T_{n-1}}, s)) ds \\ &= \int_0^t \exp(-\Lambda(Y_{T_{n-1}}, s)) \lambda(\Phi(Y_{T_{n-1}}, s)) \int_C Q_\epsilon(D - y') Q(dy' \mid \Phi(Y_{T_{n-1}}, s)) ds \quad (2.3) \end{aligned}$$

where $\Lambda(y, t) := \int_0^t \lambda(\Phi(y, s)) ds$. The (unobservable) process itself is then given by

$$Y_t := \Phi(Y_{T_n}, t - T_n), \quad \text{for } T_n \leq t < T_{n+1}, n \in \mathbb{N}_0. \quad (2.4)$$

In what follows we define the embedded process of (Y_t) by $\hat{Y}_n := Y_{T_n}$ in order to ease notation. Note that (T_n, \hat{Y}_n, X_n) is a marked point process. We call such a process *Partially Observable Piecewise Deterministic Markov Process (POPDMP)*. In the general definition of a PDMP boundary points of the state space may exist which force jumps back into the interior of the state space when reached. In order to ease the following analysis we neglect such a behavior in our model. It would have a severe impact on the filter which we need later.

3. CONTROLLED POPDMP UNDER PARTIAL OBSERVATION

Now we assume that the POPDMP can be controlled in continuous time. The set of actions is denoted by A . In order to prove existence of optimal policies later we need the following assumption.

Assumption:

(C1): The action space A is a compact metric space.

We denote by $\mathcal{P}(A)$ the set of all probability measures on (A, \mathcal{B}_A) with the weak topology. From the theory of deterministic control it is well-known that in order to prove the existence of optimal controls we have to work with relaxed controls. The space \mathcal{R} of *relaxed controls* is given by

$$\mathcal{R} := \{r : [0, \infty) \rightarrow \mathcal{P}(A) \mid r \text{ is measurable}\}.$$

On \mathcal{R} we work with the Young topology (for convergence in Young topology see the appendix). Note that under assumption (C1), the space \mathcal{R} is compact under the Young topology (see e.g. [14] Proposition 43.3 and Definition 43.4 together with the comment thereafter).

Next we define the set of observable histories up to time T_n . Let $\mathcal{H}_0 := \mathbb{R}^d$ and for $n \in \mathbb{N}$

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

and endow this space with the corresponding product σ -algebra. An element denoted by $h_n = (x_0, r_0, s_1, x_1, \dots, r_{n-1}, s_n, x_n) \in \mathcal{H}_n$ is called *observed history up to time T_n* . It consists of the received signals, the chosen controls and the inter-arrival times of jumps up to T_n . A decision rule for the period $[T_n, T_{n+1})$ is a measurable mapping

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

The upper P in the notation stands for *piecewise*. 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

$$\Pi^P := \Pi_0^P \times \Pi_1^P \times \dots$$

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

$$\pi_t := \sum_{n=0}^{\infty} 1_{\{T_n \leq t < T_{n+1}\}}(t) \cdot \pi_n^P(H_n, t - T_n), \quad (3.1)$$

where $H_n = (X_0, \pi_0^P(X_0, \cdot), S_1, X_1, \dots, \pi_{n-1}^P(H_{n-1}, \cdot), S_n, X_n)$. There is an alternative way of introducing policies which will be crucial later on and which we explain now. A *discrete time history dependent relaxed control policy* is a sequence $\pi^D := (\pi_0^D, \pi_1^D, \dots)$ of discrete time history dependent decision rules where $\pi_n^D : \mathcal{H}_n \rightarrow \mathcal{R}$ is measurable. The upper D in the notation stands for *discrete*. Note that $\pi_n^D(h_n)$ is a function in time and $\pi_n^D(h_n)(t)$ is the (randomized) action applied t time units after the n -th jump at time T_n . Here instead of a continuous-time control we have a discrete-time policy which is applied after jump time points and which now consists of functions. We write Π_n^D for the set of all discrete time history dependent decision rules at stage n and define the set of all discrete time history dependent relaxed control policies as $\Pi^D := \Pi_0^D \times \Pi_1^D \times \dots$. Note that the following statement holds which is essentially a measurability issue. For a proof see [24] Theorem 2.11.

Lemma 3.1 (Correspondence Lemma). *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*

$$\pi_n^P(h_n, t) = \pi_n^D(h_n)(t) \quad \text{a.e. on } \mathbb{R}_+ \quad \text{for all } h_n \in \mathcal{H}_n \quad (3.2)$$

and vice-versa.

Upon choosing a policy in Π^P we are able to control the data of our POPDMP in the following way. Suppose the history h_n is given up to time T_n and $\pi_n^D(h_n) = r$. Then on the time interval $[T_n, T_{n+1})$ the relaxed control r influences the drift which we denote in general by Φ^r and $\Phi^r : \mathbb{R}^d \times \mathbb{R}_+ \rightarrow \mathbb{R}^d$ is continuous and the mapping $t \mapsto \Phi^r(\cdot, t)$ is a semi-group with respect to concatenation of mappings, i.e. for all $y \in \mathbb{R}^d$ and $s, t > 0$:

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

For example let $b : \mathbb{R}^d \times A \rightarrow \mathbb{R}^d$ be a vector field such that for all $y \in \mathbb{R}^d$ and all relaxed controls $r \in \mathcal{R}$ the initial value problem

$$\frac{d}{dt} \Phi^r(y, t) = \int_A b(\Phi^r(y, t), a) r_t(da), \quad \Phi^r(y, 0) = y \quad (3.3)$$

has a unique componentwise continuous solution $\Phi^r(y, \cdot) : [0, \infty) \rightarrow \mathbb{R}^d$. Then Φ^r could be such a drift function. The relaxed control also influences the measurable *jump rate* $\lambda^A : \mathbb{R}^d \times A \rightarrow (0, \infty)$ and the action which is applied at the time point of a jump influences the transition kernel Q^A from $\mathbb{R}^d \times A$ to \mathbb{R}^d .

Definition 3.2 (Controlled POPDMP). A *Controlled Partially Observable Piecewise Deterministic Markov Process* with *local characteristics* $(\Phi^r, \lambda^A, Q^A, Q_\epsilon)$ is a stochastic process $(Y_t)_{t \geq 0}$ that satisfies the following properties: Fix $\pi \in \Pi^P$ (we write π here instead of π^P to ease notation) and an initial state y with observation x . There exists a probability space $(\Omega, \mathcal{F}, \mathbb{P}_{x,y}^\pi)$ which carries random variables $(T_n), (\hat{Y}_n), (\epsilon_n)$ such that $\mathbb{P}_{x,y}^\pi(Y_0 = y, X_0 = x) = 1$ and for all $t \geq 0, n \in \mathbb{N}_0$ and $C, D \in \mathcal{B}_d$ it holds that:

$$\begin{aligned} & \mathbb{P}_{x,y}^\pi(S_n \leq t, \hat{Y}_n \in C, X_n \in D | S_0, \hat{Y}_0, X_0, \pi_0, \dots, S_{n-1}, \hat{Y}_{n-1}, X_{n-1}, \pi_{n-1}) \\ &= \mathbb{P}_{x,y}^\pi(S_n \leq t, \hat{Y}_n \in C, X_n \in D | \hat{Y}_{n-1}, \pi_{n-1}(H_{n-1})) \\ &= \int_0^t \exp(-\Lambda^{\pi_{n-1}}(\hat{Y}_{n-1}, s)) \int_A \lambda^A(\Phi^{\pi_{n-1}}(\hat{Y}_{n-1}, s), a) \\ & \quad \int_C Q_\epsilon(D - y') Q^A(dy' | \Phi^{\pi_{n-1}}(\hat{Y}_{n-1}, s), a) \pi_{n-1}(H_{n-1}, s)(da) ds \end{aligned} \quad (3.4)$$

where $\Lambda^r(y, t) := \int_0^t \int_A \lambda^A(\Phi^r(y, s), a) r_s(da) ds$ and we use the short-hand notation $\Phi^{\pi_{n-1}}$ instead of $\Phi^{\pi_{n-1}(H_{n-1}, \cdot)}$. Note that we apply the Correspondence Lemma 3.1 here. The process (Y_t) is then defined by

$$Y_t := \Phi^{\pi_n}(Y_{T_n}, t - T_n) \text{ for } T_n \leq t < T_{n+1}, n \in \mathbb{N}_0. \quad (3.6)$$

For our existence result we need the following continuity assumptions:

Assumption:

(C2): $\lambda^A : \mathbb{R}^d \times A \rightarrow (0, \infty)$ is continuous and bounded from above by $\bar{\lambda}$ and from below by $\underline{\lambda} > 0$.

(C3): Q^A is weakly continuous, i.e. $(x, a) \mapsto \int v(z) Q^A(dz | x, a)$ is continuous and bounded for all $v : \mathbb{R}^d \rightarrow \mathbb{R}$ continuous and bounded.

Note that (C2) implies that $T_n \uparrow \infty \mathbb{P}_{xy} - a.s.$ for all $x, y \in \mathbb{R}^d$.

4. THE OPTIMIZATION PROBLEM

In this section we will introduce our optimization problem and transform it into a Markov Decision Process (MDP) which can be solved with standard techniques. We will do this in two steps: First we rewrite our continuous-time control problem for the Partially Observable Piecewise Deterministic Markov Process as a discrete-time control problem for a Partially Observable Markov Decision Process. Then we reduce this Partially Observable Markov Decision Process to a Markov Decision Process with complete observation. This problem will then be solved in the next section.

Let $\beta \in \mathbb{R}_+$ be a discount rate and $c : \mathbb{R}^d \times A \rightarrow \mathbb{R}_+$ be a measurable cost rate. The initial distribution of Y_0 given the observation $X_0 = x$ is given by the transition kernel $Q_0(\cdot | x)$. We define the *cost of policy* $\pi \in \Pi^P$ under an initial observation $x \in \mathbb{R}^d$ by (we write π instead of

π^P in order to ease notation)

$$J(x, \pi) := \int \mathbb{E}_{x,y}^\pi \left[\int_0^\infty e^{-\beta t} \int_A c(Y_t, a) \pi_t(da) dt \right] Q_0(dy|x). \quad (4.1)$$

The *value function* of the control model gives the minimal cost under an initial observation $x \in \mathbb{R}^d$ and is defined as

$$J(x) := \inf_{\pi \in \Pi^P} J(x, \pi) \quad \text{for all } x \in \mathbb{R}^d. \quad (4.2)$$

The *optimization problem* is then to find, for $x \in \mathbb{R}^d$, a policy $\pi^* \in \Pi^P$ such that we get

$$J(x) = J(x, \pi^*). \quad (4.3)$$

This problem can be rewritten as a Partially Observable Markov Decision Process in discrete time (POMDP) where we focus on the jump time points only.

Definition 4.1. Consider the following discrete-time Partially Observable Markov Decision Model:

- (i) The state space of this process is given by $\mathbb{R}_+ \times \mathbb{R}^{2d}$ and a typical state is denoted by (s, y, x) . The interpretation of the state is that y is the (unobservable) state directly after the jump which occurred s time units after the previous jump and x is the observation.
- (ii) The action space is given by \mathcal{R} and a typical action is denoted by r .
- (iii) The *substochastic* transition law is for all $x, y \in \mathbb{R}^d, t \geq 0, n \in \mathbb{N}_0$ and $C, D \in \mathcal{B}_d$ given by

$$\begin{aligned} \tilde{Q}([0, t] \times C \times D | s, y, x, r) &= \tilde{Q}([0, t] \times C \times D | y, r) \\ &= \int_0^t \exp(-\Gamma^r(y, u)) \int_A \lambda^A(\Phi^r(y, u), a) \int_C Q(D - y') Q^A(dy' | \Phi^r(y, u), a) r_u(da) du \end{aligned} \quad (4.4)$$

where $\Gamma^r(y, t) := \beta t + \int_0^t \int_A \lambda^A(\Phi^r(y, u), a) r_u(da) du$. Note that in case $\lambda^A \equiv \lambda$ we have $\tilde{Q}([0, \infty) \times \mathbb{R}^{2d} | y, r) = \frac{\lambda}{\beta + \lambda} < 1$.

- (iv) The one-stage cost depends only on $y \in \mathbb{R}^d, r \in \mathcal{R}$ and is given by

$$\begin{aligned} g(y, r) &:= \mathbb{E}_y^\pi \left[\int_0^{T_1} e^{-\beta t} \int_A c(\Phi^r(y, t), a) r_t(da) dt \right] \\ &= \int_0^\infty \exp(-\Gamma^r(y, t)) \int_A c(\Phi^r(y, t), a) r_t(da) dt. \end{aligned} \quad (4.5)$$

The last equation follows from the fact that the density of T_1 under $\mathbb{P}_{x,y}^\pi$ is given by

$$f_{T_1}(y, t) = e^{-\Lambda^r(y,t)} \int_A \lambda^A(\Phi^r(y, t), a) r_t(da)$$

and with the help of Fubini's Theorem. In order to ease notation we still denote the corresponding POMDP by (S_n, \hat{Y}_n, X_n) . According to the Theorem of Ionescu Tulcea $Q_0(\cdot|x)$ together with the transition kernel \tilde{Q} defines a probability measure $\tilde{\mathbb{P}}_x$. The difference between $\mathbb{P}_{x,y}$ and $\tilde{\mathbb{P}}_x$ is that $\tilde{\mathbb{P}}_x$ keeps track of discounting and is thus in general substochastic. For the POMDP, policies are defined as history dependent relaxed control policies $\pi^D := (\pi_0^D, \pi_1^D, \dots)$ with $\pi_n^D : \mathcal{H}_n \rightarrow \mathcal{R}$ measurable.

For a policy $\pi \in \Pi^D$ (we write π instead of π^D to ease notation) and an initial observation $x \in \mathbb{R}^d$ we define the cost of policy π as

$$\tilde{J}(x, \pi) := \tilde{\mathbb{E}}_x^\pi \left[\sum_{k=0}^\infty g(\hat{Y}_k, \pi_k(H_k)) \right] \quad (4.6)$$

where $\tilde{\mathbb{E}}_x$ is the expectation with respect to the probability measure $\tilde{\mathbb{P}}_x$.

The *value function* of the discrete time control model gives the minimal cost under an initial observation $x \in \mathbb{R}^d$ and is defined as

$$\tilde{J}(x) := \inf_{\pi \in \Pi^D} \tilde{J}(x, \pi) \quad \forall x \in \mathbb{R}^d. \quad (4.7)$$

The *discrete time optimization problem* is then to find, for $x \in \mathbb{R}^d$, a policy $\pi^* \in \Pi^D$ such that we get $\tilde{J}(x) = \tilde{J}(x, \pi^*)$. The next lemma shows that this problem is equivalent to controlling the POPDMP in (4.1).

Lemma 4.2. *Let $x \in \mathbb{R}^d$ be an initial observation, $\pi^P \in \Pi^P$ a history dependent relaxed piecewise open loop control policy for the POPDMP and $\pi^D \in \Pi^D$ its corresponding discrete-time policy according to the Correspondence Lemma. Then, it holds*

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

Proof. We obtain with the Correspondence Lemma:

$$\begin{aligned} J(x, \pi^P) &= \\ &= \int \mathbb{E}_{x,y}^{\pi^P} \left[\int_0^\infty e^{-\beta t} \int_A c(Y_t, a) \pi_t(da) dt \right] Q_0(dy|x), \\ &= \int \mathbb{E}_{x,y}^{\pi^P} \left[\sum_{k=0}^\infty \int_{T_k}^{T_{k+1}} e^{-\beta t} \int_A c(Y_t, a) \pi_k^P(H_k, t - T_k)(da) dt \right] Q_0(dy|x) \\ &= \int \mathbb{E}_{x,y}^{\pi^D} \left[\sum_{k=0}^\infty \int_{T_k}^{T_{k+1}} e^{-\beta t} \int_A c(Y_t, a) \pi_k^D(H_k)(t - T_k)(da) dt \right] Q_0(dy|x) \\ &= \int \mathbb{E}_{x,y}^{\pi^D} \left[\sum_{k=0}^\infty e^{-\beta T_k} \mathbb{E}_{\hat{Y}_k}^{\pi^D} \left[\int_{T_k}^{T_{k+1}} e^{-\beta(t-T_k)} \int_A c(Y_t, a) \pi_k^D(H_k)(t - T_k)(da) dt \middle| H_k, \hat{Y}_k, T_k \right] \right] Q_0(dy|x) \\ &= \int \mathbb{E}_{x,y}^{\pi^D} \left[\sum_{k=0}^\infty e^{-\beta T_k} g(\hat{Y}_k, \pi_k^D(H_k)) \right] Q_0(dy|x) \\ &= \tilde{\mathbb{E}}_x^{\pi^D} \left[\sum_{k=0}^\infty g(\hat{Y}_k, \pi_k^D(H_k)) \right] \end{aligned}$$

which is exactly the right hand side. Note that in the last sum there is no additional discount factor. The term $e^{-\beta T_k}$ which appears in the last but one equation is now part of the probability measure $\tilde{\mathbb{P}}_x^{\pi^D}$ (see Definition 4.1) which is substochastic. \square

In the remaining section we explain how this POMDP can be transformed into a completely observable MDP which will then be solved in the next section. We make some further simplifying assumptions. The first one implies that we later get a finite dimensional filter for our problem.

Assumption:

- (B1): There exists a finite subset $E^0 \subset \mathbb{R}^d$ with $E^0 = \{y^1, \dots, y^d\}$ such that for all $y \in \mathbb{R}^d$ and $a \in A$: $Q^A(E^0|y, a) = 1$ and Q_0 is also concentrated on E^0 .
- (B2): Q_ϵ has a bounded density f_ϵ with respect to some σ -finite measure ν .

Under Assumption (B1)-(B2) our substochastic transition law in Definition 4.1 has a density with respect to the product of Lebesgue measure, ν and the counting measure given by

$$\begin{aligned} &\tilde{q}(s, y', x|y, r) \\ &= \exp(-\Gamma^r(y, s)) f(x - y') \int_A \lambda^A(\Phi^r(y, s), a) Q^A(y'|\Phi^r(y, s), a) r_s(da). \end{aligned}$$

In order to reduce problem (4.7) to an MDP with complete observation we have to replace the unobservable state by its conditional distribution given the history so far. The computation of this conditional distribution can be done recursively. This is a Bayesian updating procedure. The conditional distribution is also called *filter*. In what follows we will introduce the updating-operator $\Psi : \mathcal{P}(E^0) \times \mathcal{R} \times \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathcal{P}(E^0)$ which maps the conditional distribution ρ of the previous step, the relaxed control r which is chosen and the received new information (this is the time point of the jump s and the observation x) onto the new conditional distribution. The updating operator essentially follows from Bayes' formula. We will later show in Lemma 4.4 that the recursive computation which is done here really yields the conditional distribution of the unobservable state. The updating-operator is defined as

$$\Psi(\rho, r, s, x)(y') := \frac{\sum_{y \in E^0} \tilde{q}(s, y', x|y, r)\rho(y)}{\sum_{\hat{y} \in E^0} \sum_{y \in E^0} \tilde{q}(s, \hat{y}, x|y, r)\rho(y)}. \quad (4.8)$$

When we denote for the history $h_n = (x_0, r_0, s_1, x_1, \dots, r_{n-1}, s_n, x_n)$ up to time T_n the following distributions

$$\begin{aligned} \mu_0(x_0) &:= Q_0(\cdot|x_0), \\ \mu_n(\cdot|h_n) = \mu_n(\cdot|h_{n-1}, r_{n-1}, s_n, x_n) &:= \Psi(\mu_{n-1}(\cdot|h_{n-1}), r_{n-1}, s_n, x_n), \end{aligned}$$

then we obtain the necessary quantity to reduce the problem to an MDP with complete observation. The previous equation is also called *filter equation*.

Definition 4.3. Consider the following discrete-time filtered Markov Decision Model with complete observation:

- (i) The state space of this process is given by $\mathcal{P}(E^0)$. A typical state is denoted by ρ . The interpretation of ρ is that it is the current conditional probability of the unobservable state.
- (ii) The action space is given by \mathcal{R} . A typical action is denoted by r .
- (iii) The transition kernel \hat{Q} from $\mathcal{P}(E^0) \times \mathcal{R}$ to $\mathcal{P}(E^0)$ is for all $r \in \mathcal{R}$ and $\rho \in \mathcal{P}(E^0)$ given by

$$\hat{Q}(B|\rho, r) = \int_{\mathbb{R}_+} \int_{\mathbb{R}^d} \sum_{y \in E^0} 1_B(\Psi(\rho, r, s, x)) \tilde{q}^{SX}(s, x|y, r) \nu(dx) ds \rho(y) \quad (4.9)$$

where $\tilde{q}^{SX}(s, x|y, r) := \sum_{y' \in E^0} \tilde{q}(s, y', x|y, r)$.

- (iv) The one-stage cost is given by

$$\hat{g}(\rho, r) := \sum_{y \in E^0} g(y, r)\rho(y). \quad (4.10)$$

The corresponding filtered MDP is denoted by (μ_n) . Policies $\pi = (f_0, f_1, \dots)$ are here defined as Markovian decision rules $f : \mathcal{P}(E^0) \rightarrow \mathcal{R}$. We denote by Π the set of all decision rules. Every $\pi = (f_0, f_1, \dots) \in \Pi^\infty$ can be seen as a special policy $\pi^D \in \Pi^D$ by setting

$$\pi_n^D(h_n) := f_n(\mu_n(\cdot|h_n)). \quad (4.11)$$

Note that in MDP theory it is well-known that we can restrict the optimization to Markovian policies (see [21], Theorem 18.4).

An initial distribution ρ on $\mathcal{P}(E^0)$ together with the transition kernels \hat{Q} define a probability measure $\hat{\mathbb{P}}_\rho$. We denote the *cost of policy* $\pi \in \Pi^\infty$ under an initial distribution $\rho \in \mathcal{P}(E^0)$ by

$$V(\rho, \pi) := \hat{\mathbb{E}}_\rho^\pi \left[\sum_{n=0}^{\infty} \hat{g}(\mu_n, f_n(\mu_n)) \right]. \quad (4.12)$$

The *value function* of the control model gives the minimal cost under an initial distribution $\rho \in \mathcal{P}(E^0)$ and is defined as

$$V(\rho) := \inf_{\pi \in \Pi^\infty} V(\rho, \pi) \quad \text{for all } \rho \in \mathcal{P}(E^0). \quad (4.13)$$

The *optimization problem* is then to find, for $\rho \in \mathcal{P}(E^0)$, a policy $\pi^* \in \Pi^\infty$ such that we get

$$V(\rho) = V(\rho, \pi^*). \quad (4.14)$$

Lemma 4.4. *Let $x \in \mathbb{R}^d$ be an initial observation, $\pi \in \Pi^\infty$ and π^D given by (4.11). Then, it holds*

$$V(Q_0(\cdot|x), \pi) = J(x, \pi^D).$$

Proof. Similar proofs can be found in [3], Theorem 5.3.2. or [4] Theorem 3.2. We first show that for any measurable $v : \mathcal{H}_n \times \mathbb{R}^d \rightarrow \mathbb{R}$ (provided the expectations exist)

$$\begin{aligned} & \tilde{\mathbb{E}}_x^{\pi^D} \left[v(X_0, R_0, S_1, X_1, \dots, R_{n-1}, S_n, X_n, \hat{Y}_n) \right] \\ &= \hat{\mathbb{E}}_{Q_0(\cdot|x)}^\pi \left[v'(X_0, R_0, S_1, X_1, \dots, R_{n-1}, S_n, X_n, \mu_n) \right] \end{aligned} \quad (4.15)$$

where $R_n := f_n(\mu_n(\cdot|H_n))$ and $v'(h_n, \rho) := \sum_{y \in E^0} v(h_n, y)\rho(y)$. This can be shown by induction on n . For $n = 0$ we have

$$\begin{aligned} \tilde{\mathbb{E}}_x^{\pi^D} \left[v(X_0, \hat{Y}_0) \right] &= \sum_{y \in E^0} v(x, y)Q_0(y|x) \\ \hat{\mathbb{E}}_{Q_0(\cdot|x)}^\pi \left[v'(X_0, \mu_0) \right] &= \sum_{y \in E^0} v(x, y)Q_0(y|x) \end{aligned}$$

so obviously both sides are equal. Now suppose the statement is true for $n - 1$ and fix $H_{n-1} = h_{n-1}$. The left-hand side of (4.15) can be written as

$$\begin{aligned} & \tilde{\mathbb{E}}_x^{\pi^D} \left[v(h_{n-1}, R_{n-1}, S_n, X_n, \hat{Y}_n) \right] \\ &= \sum_{y_{n-1}} \mu_{n-1}(y_{n-1}|h_{n-1}) \int_{\mathbb{R}^d} \int_{\mathbb{R}_+} \sum_{y_n} \tilde{q}(s_n, y_n, x_n|y_{n-1}, \pi_{n-1}^D(h_{n-1})) \\ & \quad v(h_{n-1}, \pi_{n-1}^D(h_{n-1}), s_n, x_n, y_n) ds_n \nu(dx_n). \end{aligned}$$

The right-hand side can be written as (where we use $\mu_n = \Psi(\mu_{n-1}, f_{n-1}(\mu_{n-1}), s_n, x_n)$ in the second equation)

$$\begin{aligned} & \hat{\mathbb{E}}_{Q_0}^\pi \left[v'(h_{n-1}, R_{n-1}, S_n, X_n, \mu_n) \right] \\ &= \sum_{y_{n-1}} \mu_{n-1}(y_{n-1}|h_{n-1}) \int_{\mathbb{R}_+} \int_{\mathbb{R}^d} \tilde{q}^{SX}(s_n, x_n|y_{n-1}, f_{n-1}(\mu_{n-1})) \\ & \quad v'(h_{n-1}, f_{n-1}(\mu_{n-1}), s_n, x_n, \mu_n(\cdot|h_{n-1}, f_{n-1}, s_n, x_n)) \nu(dx_n) ds_n \\ &= \sum_{y_{n-1}} \mu_{n-1}(y_{n-1}|h_{n-1}) \int_{\mathbb{R}_+} \int_{\mathbb{R}^d} \tilde{q}^{SX}(s_n, x_n|y_{n-1}, f_{n-1}(\mu_{n-1})) \\ & \quad \sum_{y_n} v(h_{n-1}, f_{n-1}(\mu_{n-1}), s_n, x_n, y_n) \\ & \quad \frac{\sum_y \tilde{q}(s_n, y_n, x_n|y, f_{n-1}(\mu_{n-1}))\mu_{n-1}(y|h_{n-1})}{\sum_{y'} \sum_y \tilde{q}(s_n, y', x_n|y, f_{n-1}(\mu_{n-1}))\mu_{n-1}(y|h_{n-1})} \nu(dx_n) ds_n. \end{aligned}$$

Note that we have

$$\begin{aligned} & \sum_{y'} \sum_y \tilde{q}(s_n, y', x_n|y, f_{n-1}(\mu_{n-1}))\mu_{n-1}(y|h_{n-1}) \\ &= \sum_y \tilde{q}^{SX}(s_n, x_n|y, f_{n-1}(\mu_{n-1}))\mu_{n-1}(y|h_{n-1}). \end{aligned}$$

Applying Fubini's Theorem to interchange the integrals we see that

$$\begin{aligned} & \hat{\mathbb{E}}_{Q_0}^\pi \left[v'(h_{n-1}, R_{n-1}, S_n, X_n, \mu_n) \right] \\ &= \int_{\mathbb{R}^d} \int_{\mathbb{R}_+} \sum_{y_n} v(h_{n-1}, f_{n-1}(\mu_{n-1}), s_n, x_n, y_n) \\ & \quad \sum_y \tilde{q}(s_n, y_n, x_n | y, f_{n-1}(\mu_{n-1})) \mu_{n-1}(y | h_{n-1}) ds_n \nu(dx_n). \end{aligned}$$

and thus both sides are equal. When we choose

$$v(H_n, \hat{Y}_n) = g(\hat{Y}_n, \pi_n^D(H_n))$$

we obtain that

$$\hat{\mathbb{E}}_x^{\pi^D} \left[g(\hat{Y}_n, \pi_n^D(H_n)) \right] = \hat{\mathbb{E}}_{Q_0}^\pi \left[\hat{g}(\mu_n, f_n(\mu_n(\cdot | H_n))) \right]$$

where we use definition (4.10) of \hat{g} on the right-hand side which implies the statement. \square

Remark 4.5. When we choose $v = 1_{B \times C}$ in the previous proof, we obtain

$$\begin{aligned} & \tilde{\mathbb{P}}_x^{\pi^D} \left((X_0, R_0, S_1, X_1, \dots, R_{n-1}, S_n, X_n) \in B, \hat{Y}_n \in C \right) \\ &= \hat{\mathbb{E}}_{Q_0}^\pi \left[1_B(X_0, R_0, S_1, X_1, \dots, R_{n-1}, S_n, X_n) \cdot \mu_n(C | X_0, R_0, S_1, X_1, \dots, R_{n-1}, S_n, X_n) \right] \end{aligned}$$

which implies that μ_n is a conditional $\tilde{\mathbb{P}}_x^{\pi^D}$ -distribution of \hat{Y}_n given the previous history $(X_0, R_0, S_1, X_1, \dots, R_{n-1}, S_n, X_n)$.

Remark 4.6. If λ^A and Q^A are not controlled i.e. do not depend on A we obtain the following special substochastic transition kernel:

$$\begin{aligned} & \tilde{q}(s, y', x | y, r) \\ &= \exp \left(-\beta t - \int_0^t \lambda(\Phi^r(y, s)) ds \right) f(x - y') \lambda(\Phi^r(y, t)) Q(y' | \Phi^r(y, t)) \end{aligned}$$

i.e. the updating-operator Ψ depends on r only through $\Phi^r(y, \cdot)$. This observation will be crucial later on (see Remark 5.7).

5. OPTIMALITY EQUATION AND EXISTENCE OF OPTIMAL POLICIES

In this section we will formulate our main theorem which states existence of an optimal policy for the original problem (4.2) and provides an optimality equation for the value function. The critical point here is to find the right continuity and compactness conditions in order to show the existence of optimal policies. In particular we have to replace the filter by a regularized version in the general case. Thus, we first make the following assumption.

Assumption:

(C4): The mapping $r \mapsto \Phi^r(y, t)$ is continuous for all $y \in E^0$ and $t \geq 0$.

(C5): The cost function $c : \mathbb{R}^d \times A \rightarrow \mathbb{R}_+$ is lower semi-continuous with respect to the product topology.

The proof of the following five lemmas can be found in the appendix.

Lemma 5.1. *Under Assumptions (C1), (C2), (C4), the mapping $r \mapsto \Gamma^r(y, t)$ is continuous for all $y \in E^0$ and $t \geq 0$.*

Lemma 5.2. *Under Assumptions (C1), (C2), (C4), (C5) the one-step cost function $(\rho, r) \mapsto \hat{g}(\rho, r)$ of the derived filtered model is lower semi-continuous.*

Lemma 5.3. *Under Assumptions (C1)-(C4), (B1), (B2) we have that $r \mapsto \int_{\mathbb{R}_+} \tilde{q}(s, y', x | y, r) ds$ is continuous for all $y, y' \in E^0$ and $x \in \mathbb{R}^d$.*

In order to prove the continuity of the transition kernel of the filtered MDP we use the following regularization of the filter. Let $h_\sigma : \mathbb{R} \rightarrow \mathbb{R}, \sigma > 0$ be a regularization kernel, i.e.

- (i) $h_\sigma(t) \geq 0$ for all $t \in \mathbb{R}$,
- (ii) $\int_{\mathbb{R}} h_\sigma(t) dt = 1$,
- (iii) $\lim_{\sigma \downarrow 0} \int_{-a}^a h_\sigma(t) dt = 1$ for all $a > 0$.

The function h_σ approximates the Dirac measure in point zero. For the general existence result we use a regularized filter of the form $\hat{\Psi} : \mathcal{P}(E^0) \times \mathcal{R} \times \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathcal{P}(E^0)$

$$\hat{\Psi}(\rho, r, s, x)(y') := \frac{\int_{\mathbb{R}} \sum_{y \in E^0} \tilde{q}(u, y', x|y, r) \rho(y) h_\sigma(s-u) du}{\sum_{\hat{y} \in E^0} \int_{\mathbb{R}} \sum_{y \in E^0} \tilde{q}(u, \hat{y}, x|y, r) \rho(y) h_\sigma(s-u) du}. \quad (5.1)$$

Note that we have $\lim_{\sigma \downarrow 0} \hat{\Psi} = \Psi$ (see e.g. [8], Theorem 1.1.7).

Lemma 5.4. *Under Assumptions (C1)-(C4), (B1), (B2) we have that $(\rho, r) \mapsto \hat{\Psi}(\rho, r, x, u)$ is continuous for all $x \in \mathbb{R}^d$.*

Finally we obtain:

Lemma 5.5. *Under all Assumptions (C1)-(C4), (B1), (B2) the stochastic transition kernel \hat{Q} in Definition 4.3 where we replace Ψ by the regularized filter $\hat{\Psi}$ is weakly continuous.*

In what follows we will always assume that the regularized filter version is used in the definition of \hat{Q} . The next step is to define the following function space

$$\mathcal{C}_{lsc}^+ := \{v : \mathcal{P}(E^0) \rightarrow [0, \infty] : v \text{ is lower semi-continuous} \}$$

and the following operators for $v \in \mathcal{C}_{lsc}^+, \rho \in \mathcal{P}(E^0), r \in \mathcal{R}$ and $f \in \Pi$:

$$\begin{aligned} (Lv)(\rho, r) &:= \hat{g}(\rho, r) + \int_{\mathcal{P}(E^0)} v(\rho') \hat{Q}(d\rho' | \rho, r), \\ (T_f v)(\rho) &:= \hat{g}(\rho, f(\rho)) + \int_{\mathcal{P}(E^0)} v(\rho') \hat{Q}(d\rho' | \rho, f(\rho)), \\ (Tv)(\rho) &:= \inf_{r \in \mathcal{R}} (Lv)(\rho, r). \end{aligned}$$

Our previous results lead now to the following observation:

Theorem 5.6. *Under all Assumptions (C1)-(C5), (B1), (B2) we have that*

- a) $T : \mathcal{C}_{lsc}^+ \rightarrow \mathcal{C}_{lsc}^+$.
- b) For all $v \in \mathcal{C}_{lsc}^+$ there exists an $f^* \in \Pi$ such that

$$(Tv)(\rho) = \inf_{r \in \mathcal{R}} (Lv)(\rho, r) = (Lv)(\rho, f^*(\rho)).$$

- c) For all $v, w \in \mathcal{C}_{lsc}^+$ with $v \leq w$ we obtain $Tv \leq Tw$.

Proof. The proof of this theorem is rather standard. If we first choose v to be continuous and bounded we obtain by Lemma 5.5 that

$$(\rho, r) \mapsto \int_{\mathcal{P}(E^0)} v(\rho') \hat{Q}(d\rho' | \rho, r)$$

is continuous and bounded. Thus using the same line of arguments as in the proof of Lemma 5.2 we obtain that the same mapping is lower semi-continuous when we plug in a lower semi-continuous function v . Since the sum of lower semi-continuous functions is again lower semi-continuous we get with Lemma 5.2 that

$$(\rho, r) \mapsto \hat{g}(\rho, r) + \int_{\mathcal{P}(E^0)} v(\rho') \hat{Q}(d\rho' | \rho, r) = (Lv)(\rho, r)$$

is lower semi-continuous. We can now use a classical measurable selection theorem for part b) (see e.g. Proposition 7.33 in [6]) and part a) follows as in Proposition 2.4.3 in [3], see also section 3.3 of [20] or Propositions 7.31 and 7.33 in [6]. Part c) is obvious. \square

Remark 5.7. Note that the existence of a minimizer $f^* \in \Pi$ in Theorem 5.6 b) cannot be shown in general if we take the original filter Ψ . In this case examples can be constructed where the filter is not continuous (for a discussion and the example see [24] Section 3.2.2). The crucial point here is that the single action which is applied at the jump time point enters the filter. This effect is incompatible with the Young topology. It does not occur when λ^A and Q^A are uncontrolled. Indeed Lemma 5.5 holds true for the original filter Ψ in case λ^A, Q^A are not controlled. In this case we have

$$\Psi(\rho, r, s, x)(y') := \frac{\sum_{y \in E^0} \tilde{p}(s, y', x|y, r)\rho(y)}{\sum_{\hat{y} \in E^0} \sum_{y \in E^0} \tilde{p}(s, \hat{y}, x|y, r)\rho(y)}.$$

with

$$\tilde{p}(t, y', x|y, r) = \exp\left(-\int_0^t \lambda(\Phi^r(y, s))ds\right) f(x - y') \lambda(\Phi^r(y, t)) Q(y'|\Phi^r(y, t)).$$

So Ψ depends on r only through $\Phi^r(y, \cdot)$ which is continuous by Assumption (C4).

In order to derive the optimality equation we need to consider the n -stage version of the optimization problems. Thus we define for a policy $\pi \in \Pi^\infty$ and the function $\underline{0} \in \mathcal{C}_{lsc}^+$ which is identical to zero, the following value functions:

$$\begin{aligned} V_n(\rho, \pi) &:= T_{f_0} \dots T_{f_{n-1}} \underline{0} \\ V_n(\rho) &:= \inf_{\pi \in \Pi^\infty} V_n(\rho, \pi) = T^n \underline{0}. \end{aligned}$$

Note that $V_n(\rho, \pi)$ is exactly the expected cost of policy π until jump time T_n . By general MDP techniques (see e.g. [3], chap. 2) we obtain the last equation $V_n(\rho) = T^n \underline{0}$ which also implies that $V_n = TV_{n-1}$. Since the cost function is non-negative we obtain by monotone convergence that the following limits exist:

$$\begin{aligned} V(\rho, \pi) &:= \lim_{n \rightarrow \infty} V_n(\rho, \pi) \\ V_\infty(\rho) &:= \lim_{n \rightarrow \infty} V_n(\rho). \end{aligned}$$

By definition we get that $V(\rho) = \inf_{\pi \in \Pi^\infty} V(\rho, \pi)$. From Theorem 5.6 it follows that $V_n \in \mathcal{C}_{lsc}^+$ because $\underline{0} \in \mathcal{C}_{lsc}^+$ and hence also $V_\infty \in \mathcal{C}_{lsc}^+$. Moreover we immediately obtain by monotonicity that $V(\rho, \pi) \geq V_n(\rho, \pi)$ for all $\pi \in \Pi^\infty$ which then implies $V(\rho) \geq V_n(\rho)$ and with $n \rightarrow \infty$ that $V(\rho) \geq V_\infty(\rho)$. The main result of this section is the following:

Theorem 5.8. *Under all Assumptions (C1)-(C4), (B1),(B2) we have that*

- a) $TV_\infty = V_\infty$.
- b) $V_\infty = V$.
- c) *There exists an $f^* \in \Pi$ with $TV = T_{f^*}V$ and the stationary policy (f^*, f^*, \dots) is optimal for problem (4.13). The optimal policy for the original problem (4.2) is thus $(\pi_0^P, \pi_1^P, \dots)$ with*

$$\begin{aligned} \pi_0^P(x, t) &= f^*(Q_0(\cdot|x))(t), \quad x \in \mathbb{R}^d \\ \pi_n^P(h_n, t) &= f^*(\mu_n(\cdot|h_n))(t), \quad h_n \in H_n. \end{aligned}$$

Proof. a) Due to monotonicity of V_n and the T -operator we obtain $V_n \leq TV_\infty$ for all $n \in \mathbb{N}$. For $n \rightarrow \infty$ we obtain $V_\infty \leq TV_\infty$. It remains to prove $V_\infty \geq TV_\infty$. Since $T^n \underline{0} \in \mathcal{C}_{lsc}^+$ we know by Theorem 5.6 that there exist decision rules f_k^* such that $T^n \underline{0} = T_{f_0^*} \dots T_{f_{n-1}^*} \underline{0}$. Now fix $\rho \in \mathcal{P}(E^0)$ and define $r^n := f_n^*(\rho)$. Then $(r^n) \subset \mathcal{R}$ and since \mathcal{R} is compact there exists a converging subsequence $\lim_{k \rightarrow \infty} r^{n_k} = r \in \mathcal{R}$. This implies by monotonicity for an arbitrary index n_k and all $n \leq n_k$:

$$V_\infty(\rho) \geq (T^{n_k+1} \underline{0})(\rho) = (LT^{n_k} \underline{0})(\rho, r^{n_k}) \geq (LT^n \underline{0})(\rho, r^{n_k}).$$

Since $T^n \underline{0} \in \mathcal{C}_{lsc}^+$ the mapping $r \mapsto (LT^n \underline{0})(\rho, r)$ is lower semi-continuous. Thus we obtain by definition of this property that

$$V_\infty(\rho) \geq \lim_{k \rightarrow \infty} (LT^n \underline{0})(\rho, r^{nk}) \geq (LT^n \underline{0})(\rho, r).$$

And with $n \rightarrow \infty$ we obtain by monotone convergence

$$V_\infty(\rho) \geq (LV_\infty)(\rho, r) \geq (TV_\infty)(\rho)$$

which finally implies the statement.

- b) Since $V \geq V_\infty$ it is sufficient to prove $V \leq V_\infty$. By part a) we know that $V_\infty \geq TV_\infty$. Since $V_\infty \in \mathcal{C}_{lsc}^+$ we know by Theorem 5.6 that there exists a decision rule f^* such that $TV_\infty = T_{f^*} V_\infty$. Iterating this equation yields:

$$V_\infty \geq T_{f^*} V_\infty = T_{f^*}^n V_\infty \geq T_{f^*}^n \underline{0}.$$

with $n \rightarrow \infty$ we obtain

$$V_\infty(\rho) \geq V(\rho, f^{*\infty}) \geq V(\rho)$$

which implies the statement.

- c) This follows from the proof of part b) and the use of the Correspondence Lemma 3.1. \square

Besides the existence of optimal policies Theorem 5.8 presents a numerical way of computing the value function V . Since $V = V_\infty$ according to part b) we can use *value iteration* to approximate V , i.e. we can start with $V_0 = \underline{0}$ and compute $V_n = TV_{n-1}$ for large n .

Since we use the regularized filter with a regularization kernel h_σ , the optimal policy depends on σ . For nice problems we expect convergence of the optimal policies for $\sigma \downarrow 0$ to an optimal policy for the original problem. A general theorem which guarantees this is as follows. Fix state $\rho \in \mathcal{P}(E^0)$ and suppose that $\sigma_n \downarrow 0$ for $n \rightarrow \infty$. We denote the value function which corresponds to σ_n by V^n . Let

$$\mathcal{R}_n := \{r \in \mathcal{R} \mid (LV^n)(\rho, r) = (TV^n)(\rho)\}$$

for $n \in \mathbb{N} \cup \{\infty\}$ be the set of maximum points of the value function V^n in state ρ . The value function V^∞ corresponds to the problem with original filter Ψ . Further let

$$Ls\mathcal{R}_n := \{r \in \mathcal{R} \mid r \text{ is an accumulation point of } (r^n) \text{ with } r^n \in \mathcal{R}_n\}.$$

The next theorem follows from Theorem A.1.5 in [3].

Theorem 5.9. *Suppose there exists a sequence $\delta_m \downarrow 0$ for $m \rightarrow \infty$ such that $(LV^n)(\rho, r) \geq (LV^m)(\rho, r) + \delta_m$ for all $n \geq m$, i.e. the sequence $(LV^n)(\rho, r)$ is weakly increasing. Then $\emptyset \neq Ls\mathcal{R}_n \subset \mathcal{R}_\infty$.*

The interpretation of Theorem 5.9 is as follows: Fix $\rho \in \mathcal{P}(E^0)$ and take $\sigma_n \downarrow 0$. Suppose the sequence $(LV^n)(\rho, r)$ is weakly increasing. The sequence of optimal relaxed controls (r^n) in state ρ has at least one accumulation point and every accumulation point is an optimal relaxed control in the original model with non-regularized filter.

6. APPLICATION

In this section we illustrate our approach by a simple example. The task is to steer a particle which moves on the real line into a target zone. At random time points the particle jumps into one of a finite number of states. However, the position of the particle cannot be observed. The only information we have is that after the random jump time points a noisy signal is received.

We consider this problem as a POPDMP where we specify the following data:

- (i) The state space is assume to be \mathbb{R} and the action space is assumed to be $A := [-1; 1]$. Here $a \in A$ refers to the speed with which the particle is moved into one of the two directions. Obviously A is compact, hence (C1) is satisfied.
- (ii) The set of possible post jump states is assumed to be $E^0 := \{-2, 0, 2\}$ and we set $y^1 := -2, y^2 := 0$ and $y^3 := 2$. Thus (B1) is valid.

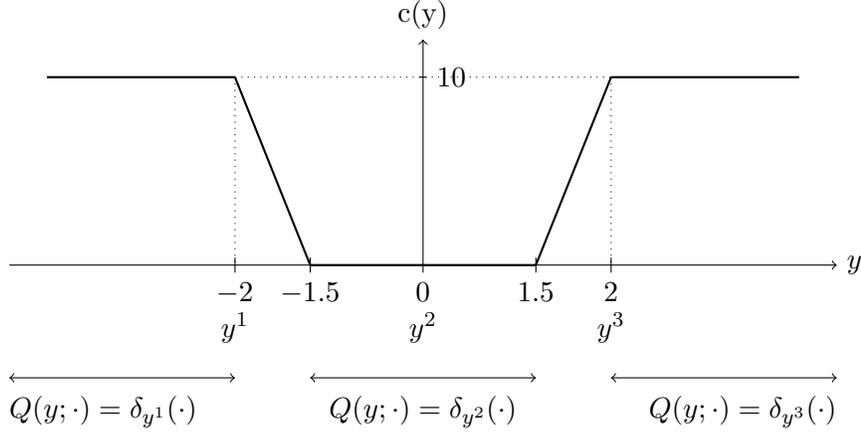


FIGURE 1. Cost function and transition kernel in concrete application example.

(iii) The controlled drift is given by

$$\frac{d}{dt}\Phi^r(y, t) = \int_A ar_t(da), \quad \Phi(y, 0) = y. \quad (6.1)$$

this implies (C4).

(iv) We set $\lambda^A \equiv 1$ and $\beta := 1$, i.e. the transition rate is uncontrolled and the discount rate is equal to one. Hence (C2) is satisfied.

(v) The jump transition kernel Q^A is also uncontrolled and specified as follows (see also Figure 1):

$$Q^A(\cdot|y) := \begin{cases} \delta_{y^1}(\cdot), & y \leq -2 \\ (-3 - 2y) \cdot \delta_{y^1}(\cdot) + (4 + 2y) \cdot \delta_{y^2}(\cdot), & -2 < y < -\frac{3}{2} \\ \delta_{y^2}(\cdot), & -\frac{3}{2} \leq y \leq \frac{3}{2} \\ (4 - 2y) \cdot \delta_{y^2}(\cdot) + (2y - 3) \cdot \delta_{y^3}(\cdot), & \frac{3}{2} < y < 2 \\ \delta_{y^3}(\cdot), & 2 \leq y, \end{cases}$$

where δ_x is the Dirac measure on point $x \in E^0$. Note that Q^A is weakly continuous, hence (C3) holds.

(vi) The cost function is independent of a and given by (see also Figure 1):

$$c(y) := \begin{cases} 10, & y \leq -2 \\ -30 - 20y, & -2 < y < -\frac{3}{2} \\ 0, & -\frac{3}{2} \leq y \leq \frac{3}{2} \\ 20y - 30, & \frac{3}{2} < y < 2 \\ 10, & y \geq 2. \end{cases}$$

Note that c is continuous which implies (C5).

(vii) For the density of the signal we take the discrete density $f_\epsilon(-1) = f_\epsilon(0) = f_\epsilon(1) = \frac{1}{3}$. Hence (B2) is satisfied.

First note that since only $\int_A ar_t(da)$ enters the equations we can restrict to deterministic controls. We still denote them by r and consider $r_t \in A$ instead of $r_t \in \mathcal{P}(A)$ which is a slight

abuse of notation. The updating operator Ψ in this case reads

$$\Psi(\rho, r, s, x)(y^j) = \frac{f(x - y^j) \sum_y Q(y^j|y + \int_0^s r_u du) \rho(y)}{\sum_{\hat{y}} f(x - \hat{y}) \sum_y Q(\hat{y}|y + \int_0^s r_u du) \rho(y)}.$$

Since $\lambda^A = \lambda$ and $Q^A = Q$ are uncontrolled we do not have to consider the regularized filter. The one-stage reward is given by

$$\hat{g}(\rho, r) = \sum_y \rho(y) \int_0^\infty e^{-2t} c(y + \int_0^t r_s ds) dt$$

and finally the transition kernel is for a measurable function $v : \mathcal{P}(E^0) \rightarrow \mathbb{R}$ given by

$$\int v(\rho') \hat{Q}(d\rho'|\rho, r) = \frac{1}{3} \int_0^\infty e^{-2t} \sum_{d=-1}^1 \sum_{y'} v(\Psi(\rho, r, t, y' + d)) \sum_y Q(y'|y + \int_0^t r_s ds) \rho(y) dt.$$

The optimization problem is

$$J(x) := \inf_{\pi \in \Pi^P} \int \mathbb{E}_{x,y}^\pi \left[\int_0^\infty e^{-t} c(Y_t) dt \right] Q_0(dy|x) \quad (6.2)$$

and the corresponding filtered MDP is defined by the T -operator which in this example reads

$$(Tv)(\rho) = \inf_{r \in \mathcal{R}} \left\{ \int_0^\infty e^{-2t} \left[\sum_y \rho(y) c\left(y + \int_0^t r_s ds\right) + \frac{1}{3} \sum_{d=-1}^1 \sum_{y'} v(\Psi(\rho, r, t, y' + d)) \sum_y Q(y'|y + \int_0^t r_s ds) \rho(y) \right] dt \right\}. \quad (6.3)$$

Since all assumptions of Theorem 5.8 are satisfied we obtain in this example.

Lemma 6.1. *In this POPDMP there exists an $f^* \in \Pi$ with $TV = T_{f^*}V$ and the stationary policy (f^*, f^*, \dots) is optimal for the filtered MDP. The optimal policy for the original problem (6.2) is thus $(\pi_0^P, \pi_1^P, \dots)$ with*

$$\begin{aligned} \pi_0^P(x, t) &= f^*(Q_0(\cdot|x))(t), \quad x \in \mathbb{R}^d \\ \pi_n^P(h_n, t) &= f^*(\mu_n(\cdot|h_n))(t), \quad h_n \in H_n. \end{aligned}$$

In this example we have also computed the value function and the optimal policy numerically by value iteration. The value function V as a function of $\rho_1 \in (0, 1)$ and $\rho_3 \in (0, 1 - \rho_1)$ can be seen in Figure 2. The optimal policy turned out to always use one of the values $\{-1, 0, 1\}$. More precisely we obtain

$$\pi_n(h_n, t) := \begin{cases} 1_{\{t \leq \frac{1}{2}\}} & \text{if } \mu_n^1 \geq \mu_n^3, \\ -1_{\{t \leq \frac{1}{2}\}} & \text{if } \mu_n^1 \leq \mu_n^3. \end{cases} \quad (6.4)$$

Recall that $\mu_n(h_n)$ is the recursively calculated conditional distribution on $\mathcal{P}(E^0)$.

7. APPENDIX

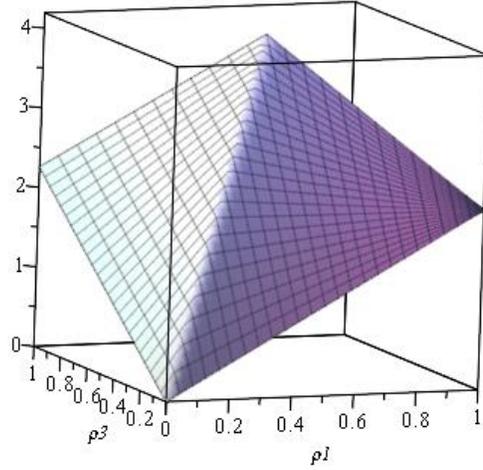
7.1. Young Topology. The Young topology is metrizable and convergence can be characterized as follows (for a proof see e.g. [24] Lemma A.21):

Definition 7.1. Let $(r^n)_{n \in \mathbb{N}}$ be a sequence in \mathcal{R} and $r \in \mathcal{R}$. Then

$$\lim_{n \rightarrow \infty} r^n = r \quad \iff \quad \lim_{n \rightarrow \infty} \int_0^\infty \int_A \psi(t, a) r_t^n(da) dt = \int_0^\infty \int_A \psi(t, a) r_t(da) dt \quad (7.1)$$

for all $\psi : \mathbb{R}_+ \times A \rightarrow \mathbb{R}$ which are measurable in the first component, continuous and bounded in the second component and satisfy

$$\int_0^\infty \sup_{a \in A} |\psi(t, a)| dt < \infty.$$

FIGURE 2. Value function $V(\rho_1, \cdot, \rho_3)$

7.2. An Auxiliary Result and some Proofs. The following auxiliary result is very helpful for our convergence statements. For a proof see e.g. [24], Lemma B.12:

Lemma 7.2. *Let X be a separable and metrizable space, Y a compact metric space and $f : X \times Y \rightarrow \mathbb{R}$ continuous. Then $\lim_{n \rightarrow \infty} x_n = x$ implies*

$$\lim_{n \rightarrow \infty} \sup_{y \in Y} |f(x_n, y) - f(x, y)| = 0.$$

Proof of Lemma 5.1:

Note that by definition $\Gamma^r(y, t) = \beta t + \Lambda^r(y, t)$. Thus, it is enough to show that the mapping $r \mapsto \Lambda^r(y, t)$ is continuous. Let $y \in E^0$ and $t \geq 0$. Further, let (r^n) be a sequence in \mathcal{R} with $\lim_{n \rightarrow \infty} r^n = r \in \mathcal{R}$. By definition of Λ^r , we then get:

$$\begin{aligned} & |\Lambda^{r^n}(y, t) - \Lambda^r(y, t)| \\ &= \left| \int_0^t \int_A \lambda^A(\Phi^{r^n}(y, s), a) r_s^n(da) ds - \int_0^t \int_A \lambda^A(\Phi^r(y, s), a) r_s(da) ds \right| \\ &\leq \left| \int_0^t \int_A \{ \lambda^A(\Phi^{r^n}(y, s), a) - \lambda^A(\Phi^r(y, s), a) \} r_s^n(da) ds \right| \\ &\quad + \left| \int_0^t \int_A \lambda^A(\Phi^r(y, s), a) r_s^n(da) ds - \int_0^t \int_A \lambda^A(\Phi^r(y, s), a) r_s(da) ds \right|. \end{aligned} \quad (7.2)$$

Looking now at the first summand of (7.2) we find that for $n \rightarrow \infty$

$$\begin{aligned} & \left| \int_0^t \int_A \{ \lambda^A(\Phi^{r^n}(y, s), a) - \lambda^A(\Phi^r(y, s), a) \} r_s^n(da) ds \right| \\ &\leq \int_0^t \sup_{a \in A} |\lambda^A(\Phi^{r^n}(y, s), a) - \lambda^A(\Phi^r(y, s), a)| ds \rightarrow 0. \end{aligned}$$

This convergence is true since by the continuity of Φ^r and λ^A and by the compactness of A we have with the help of Lemma 7.2

$$\lim_{n \rightarrow \infty} \sup_{a \in A} |\lambda^A(\Phi^{r^n}(y, s), a) - \lambda^A(\Phi^r(y, s), a)| = 0.$$

By the boundedness of λ^A , dominated convergence leads to the convergence of the integral towards zero.

Now, looking at the second summand in (7.2) we obtain by the characterization of the Young topology (see Definition 7.1) and by assumption (C2) that

$$\lim_{n \rightarrow \infty} \int_0^t \int_A \lambda^A(\Phi^r(y, s), a) r_s^n(da) ds = \int_0^t \int_A \lambda^A(\Phi^r(y, s), a) r_s(da) ds. \quad (7.3)$$

This implies the statement.

Proof of Lemma 5.2:

We first show that when c is continuous and bounded, then $(\rho, r) \mapsto \hat{g}(\rho, r)$ is continuous and bounded. Let c be continuous and bounded and suppose $\lim_{n \rightarrow \infty} (\rho^n, r^n) = (\rho, r)$ with respect to the product topology. Let us denote $\eta^r(y, t) := e^{-\Gamma^r(y, t)}$. Based on the representation of g we then get

$$\begin{aligned} |\hat{g}(\rho^n, r^n) - \hat{g}(\rho, r)| &\leq \sum_{y \in E^0} \left| \rho^n(y) \int_0^\infty \eta^{r^n}(y, t) \int_A c(\Phi^{r^n}(y, t), a) r_t^n(da) dt \right. \\ &\quad \left. - \rho(y) \int_0^\infty \eta^r(y, t) \int_A c(\Phi^r(y, t), a) r_t(da) dt \right|. \end{aligned}$$

From our assumption it follows that we obtain pointwise convergence $\lim_{n \rightarrow \infty} \rho^n(y) = \rho(y)$ and it thus remains to show that for all $y \in E^0$

$$\lim_{n \rightarrow \infty} \int_0^\infty \eta^{r^n}(y, t) \int_A c(\Phi^{r^n}(y, t), a) r_t^n(da) dt = \int_0^\infty \eta^r(y, t) \int_A c(\Phi^r(y, t), a) r_t(da) dt.$$

Hence consider

$$\begin{aligned} &\left| \int_0^\infty \eta^{r^n}(y, t) \int_A c(\Phi^{r^n}(y, t), a) r_t^n(da) dt - \int_0^\infty \eta^r(y, t) \int_A c(\Phi^r(y, t), a) r_t(da) dt \right| \\ &\leq \left| \int_0^\infty (\eta^{r^n}(y, t) - \eta^r(y, t)) \int_A c(\Phi^{r^n}(y, t), a) r_t^n(da) dt \right| \\ &\quad + \left| \int_0^\infty \eta^r(y, t) \left\{ \int_A c(\Phi^{r^n}(y, t), a) r_t^n(da) - \int_A c(\Phi^r(y, t), a) r_t(da) \right\} dt \right|. \end{aligned} \quad (7.4)$$

Now, as c is bounded by our initial assumption, the first summand satisfies

$$\begin{aligned} &\left| \int_0^\infty (\eta^{r^n}(y, t) - \eta^r(y, t)) \int_A c(\Phi^{r^n}(y, t), a) r_t^n(da) dt \right| \\ &\leq \sup_{x, a} |c(x, a)| \int_0^\infty |\eta^{r^n}(y, t) - \eta^r(y, t)| dt \rightarrow 0. \end{aligned}$$

The convergence follows from dominated convergence where $|\eta^{r^n}(y, t) - \eta^r(y, t)|$ is dominated by $2e^{-\beta t}$ and $\lim_{n \rightarrow \infty} \Gamma^{r^n}(y, t) = \Gamma^r(y, t)$ because of Lemma 5.1.

The second summand of (7.4) can be dominated by $Term_1 + Term_2$ with

$$Term_1 := \int_0^\infty \eta^r(y, t) \int_A |c(\Phi^{r^n}(y, t), a) - c(\Phi^r(y, t), a)| r_t^n(da) dt$$

and

$$Term_2 := \left| \int_0^\infty \eta^r(y, t) \int_A c(\Phi^r(y, t), a) r_t^n(da) dt - \int_0^\infty \eta^r(y, t) \int_A c(\Phi^r(y, t), a) r_t(da) dt \right|.$$

We will show that both, $Term_1$ and $Term_2$ converge to zero. First, as c is continuous and bounded and A is compact we obtain with the help of Lemma 7.2

$$\lim_{n \rightarrow \infty} \sup_{a \in A} |c(\Phi^{r^n}(y, t), a) - c(\Phi^r(y, t), a)| = 0.$$

Thus $Term_1$ converges to zero by dominated convergence applied for dominating function $t \mapsto 2 \sup_{x,a} |c(x,a)| \eta^r(y,t)$. For $Term_2$ we get convergence to zero from the characterization of the Young topology convergence in Definition 7.1 as

$$(t, a) \mapsto \eta^r(y, t) c(\Phi^r(y, t), a)$$

is measurable in t and continuous and bounded in a and because of

$$\int_0^\infty \eta^r(y, t) \sup_{a \in A} |c(\Phi^r(y, t), a)| dt \leq \sup_{(x,a) \in \mathbb{R}^d \times A} |c(x, a)| \int_0^\infty e^{-\beta t} dt < \infty.$$

We also get that \hat{g} is bounded when c is bounded.

Now, let c be lower semi-continuous (and non-negative, what we always assume). Then, there is a sequence (c_m) of continuous and bounded functions with $c_m \uparrow c$ for $m \rightarrow \infty$ (see [6], Lemma 7.14). Thus we can apply our previous findings to c_m and by monotonicity of the convergence obtain that \hat{g} is lower semi-continuous.

Proof of Lemma 5.3: Suppose $(r^n) \subset \mathcal{R}$ and $r^n \rightarrow r \in \mathcal{R}$ for $n \rightarrow \infty$. Let us denote $\eta^r(y, s) := e^{-\Gamma^r(y,s)}$ and consider $\int \tilde{q}(s, y', x|y, r) ds$. Obviously the factor $f(x - y')$ does not depend on r and can be ignored. We obtain

$$\begin{aligned} & \left| \int_0^\infty \eta^{r^n}(y, s) \int_A \lambda^A(\Phi^{r^n}(y, s), a) Q^A(y'|\Phi^{r^n}(y, s), a) r_s^n(da) ds - \right. \\ & \quad \left. - \int_0^\infty \eta^r(y, s) \int_A \lambda^A(\Phi^r(y, s), a) Q^A(y'|\Phi^r(y, s), a) r_s(da) ds \right| \\ & \leq \left| \int_0^\infty \int_A \left\{ \eta^{r^n}(y, s) \lambda^A(\Phi^{r^n}(y, s), a) Q^A(y'|\Phi^{r^n}(y, s), a) - \right. \right. \\ & \quad \left. \left. - \eta^r(y, s) \lambda^A(\Phi^r(y, s), a) Q^A(y'|\Phi^r(y, s), a) \right\} r_s^n(da) ds \right| \\ & + \left| \int_0^\infty \int_A \eta^r(y, s) \lambda^A(\Phi^r(y, s), a) Q^A(y'|\Phi^r(y, s), a) r_s^n(da) ds - \right. \\ & \quad \left. - \int_0^\infty \int_A \eta^r(y, s) \lambda^A(\Phi^r(y, s), a) Q^A(y'|\Phi^r(y, s), a) r_s(da) ds \right|. \end{aligned}$$

The first of these two terms converges to zero with Lemma 7.2. The second term converges to zero by the definition of the Young topology and the fact that

$$\int_0^\infty \eta^r(y, s) \sup_{a \in A} |\lambda^A(\Phi^r(y, s), a) Q^A(y'|\Phi^r(y, s), a)| ds < \infty.$$

Proof of Lemma 5.4: In the same way as in the proof of Lemma 5.3 it can be shown that

$$r \mapsto \int_{\mathbb{R}_+} h_\sigma(s - u) \tilde{q}(u, y', x|y, r) du$$

is continuous. The statement follows since $(\rho, r) \mapsto \hat{\Psi}(\rho, r, x, u)$ is a continuous composition of these functions.

Proof of Lemma 5.5:

Proof. We have to show that

$$(\rho, r) \mapsto \int_{\mathbb{R}_+} \int_{\mathbb{R}^d} \sum_y v(\hat{\Psi}(\rho, r, s, x)) \tilde{q}^{SX}(s, x|y, r) \nu(dx) ds \rho(y)$$

is continuous for v bounded continuous. Obviously it is enough to show for fixed $y \in E^0$ that

$$(\rho, r) \mapsto \int_{\mathbb{R}_+} \int_{\mathbb{R}^d} v(\hat{\Psi}(\rho, r, s, x)) \tilde{q}^{SX}(s, x|y, r) \nu(dx) ds$$

is continuous. Let $\lim_{n \rightarrow \infty} (\rho^n, r^n) = (\rho, r)$ w.r.t. the product topology. We obtain:

$$\begin{aligned} & \left| \int_{\mathbb{R}_+} \int_{\mathbb{R}^d} v(\hat{\Psi}(\rho^n, r^n, s, x)) \tilde{q}^{SX}(s, x|y, r^n) \nu(dx) ds - \int_{\mathbb{R}_+} \int_{\mathbb{R}^d} v(\hat{\Psi}(\rho, r, s, x)) \tilde{q}^{SX}(s, x|y, r) \nu(dx) ds \right| \\ & \leq \left| \int_{\mathbb{R}_+} \int_{\mathbb{R}^d} v(\hat{\Psi}(\rho^n, r^n, s, x)) \left(\tilde{q}^{SX}(s, x|y, r^n) - \tilde{q}^{SX}(s, x|y, r) \right) \nu(dx) ds \right| \\ & \quad + \left| \int_{\mathbb{R}_+} \int_{\mathbb{R}^d} \left(v(\hat{\Psi}(\rho^n, r^n, s, x)) - v(\hat{\Psi}(\rho, r, s, x)) \right) \tilde{q}^{SX}(s, x|y, r) \nu(dx) ds \right|. \end{aligned}$$

Since v is bounded by a constant, the first term can be bounded by

$$\sup_{\rho} |v(\rho)| \times \left| \int_{\mathbb{R}_+} \int_{\mathbb{R}^d} \tilde{q}^{SX}(s, x|y, r^n) - \tilde{q}^{SX}(s, x|y, r) \nu(dx) ds \right|$$

which converges to zero for $n \rightarrow \infty$ because of Lemma 5.3 and dominated convergence. The second term converges to zero by dominated convergence and continuity of v and $\hat{\Psi}$. \square

REFERENCES

- [1] Almudevar, A., A dynamic programming algorithm for the optimal control of piecewise deterministic Markov processes. *SIAM J. Control Optim.*, **40**, 525–539, 2001.
- [2] Bäuerle, N., Discounted stochastic fluid programs. *Math. Oper. Res.*, **26**, 401–420, 2001.
- [3] Bäuerle, N. and Rieder, U. Markov Decision Processes with Applications to Finance. Springer-Verlag, Berlin Heidelberg, 2011.
- [4] Bäuerle, N. and Rieder, U. Partially observable risk-sensitive Markov Decision Processes. To appear *Math. Oper. Res.*, 2017.
- [5] Bayraktar, E. and Ludkovski, M., Inventory management with partially observed nonstationary demand. *Ann. Oper. Res.*, **176**, 7–39, 2010.
- [6] Bertsekas, D.P. and Shreve, E. Stochastic Optimal Control: The Discrete Time Case. Academic Press, New York, 1978.
- [7] Brandejsky, A. and de Saporta, B. and Dufour, F., Optimal stopping for partially observed piecewise-deterministic Markov processes. *Stochastic Process. Appl.*, **123**, 3201–3238, 2013.
- [8] Brémaud, P. Fourier Analysis and Stochastic Processes. Springer-Verlag, Cham Heidelberg, 2014.
- [9] Chafaï, D. and Malrieu, F. and Paroux, K., On the long time behavior of the TCP window size process. *Stochastic Process. Appl.*, **120**, 1518–1534, 2010.
- [10] Costa, O.L.V. and Dufour, F. Average continuous control of piecewise deterministic Markov processes. *SIAM J. Control Optim.*, **48**, 4262–4291, 2010.
- [11] Costa, O.L.V. and Dufour, F. Continuous Average Control of Piecewise Deterministic Markov Processes. Springer Briefs in Mathematics, 2010.
- [12] Costa, O.L.V. and Raymundo, C.A.B., Impulse and continuous control of piecewise deterministic Markov processes. *Stoch. Stoch. Rep.*, **70**, 75–107, 2000.
- [13] Davis, M.H.A., Piecewise-deterministic Markov Processes: A General Class of Non-diffusion Stochastic Models. *Journal of the Royal Statistical Society B*, **46**, 353–388, 1984.
- [14] Davis, M.H.A., Markov Models and Optimization. Chapman and Hall, London, 1993.
- [15] Dempster, M.A.H. and Ye, J.J., Necessary and sufficient optimality conditions for control of piecewise deterministic processes. *Stoch. Stoch. Rep.*, **40**, 125–145, 1992.
- [16] Dufour, F. and Dutuit, Y., Dynamic reliability: A new model, *Proceedings of ESREL 2002 Lambda-Mu 13 Conference*, 350–353, 2002.
- [17] Feinberg, E.A. and Kasyanov, P.O. and Zgurovsky, M.Z., Partially Observable Total-Cost Markov Decision Processes with Weakly Continuous Transition Probabilities. *Math. Oper. Res.*, **41**, 656–681, 2016.
- [18] Forwick, L. and Schäl, M. and Schmitz, M., Piecewise deterministic Markov control processes with feedback controls and unbounded costs. *Acta Appl. Math.*, **82**, 239–267, 2004.
- [19] Hernández-Lerma, O. Adaptive Markov Control Processes, Springer-Verlag, New York, 1989.
- [20] Hernández-Lerma, O. and Lasserre, J.B., Discrete-time Markov Control Processes, Springer-Verlag, New York, 1996.
- [21] Hinderer, K. Foundations of non-stationary dynamic programming with discrete time parameter, Springer-Verlag, Berlin, 1970.
- [22] Jacobsen, M. Point Process Theory and Applications - Marked Point and Piecewise Deterministic Processes, Birkhäuser, Boston, 2006.
- [23] Kirch, M. and Runggaldier, W.J., Efficient hedging when asset prices follow a geometric Poisson process with unknown intensities. *SIAM J. Control Optim.*, **43**, 1174–1195, 2005.

- [24] Lange, D. Cost Optimal Control of Piecewise Deterministic Markov Processes under Partial Observation, PhD KIT 2017. DOI: 10.5445/IR/1000069448
- [25] Lygeros, J. and Koutroumpas, K. and Dimopolous, S. and Legouras, P. and Heichinger, C. and Nurse, P. and Lygerou, Z., Stochastic hybrid modeling of DNA replication across a complete genome. *Proceedings of the National Academy of Sciences*, **105**, 12295-12300, 2008.
- [26] Kakis, V. and Jiang, X., Optimal replacement under partial observation. *Math. Oper. Res.*, **28**, 382-394, 2003.
- [27] Pakdaman, K. and Thieullen, M. and Wainrib, G., Fluid limit theorems for stochastic hybrid systems with application to neuron models. *Adv. in Appl. Probab.*, **42**, 761-794, 2010.
- [28] de Saporta, B. and Dufour, F. and Gonzalez, K., Numerical method for optimal stopping of piecewise deterministic Markov processes. *Ann. Appl. Probab.*, **20**, 1607-1637, 2010.
- [29] Schäl, M., On piecewise deterministic Markov control processes: control of jumps and of risk processes in insurance. *Insurance Math. Econom.*, **22**, 75-91, 1998.
- [30] Yushkevich, A. A. On reducing a jump controllable Markov model to a model with discrete time. *Theory Probab. Appl.*, **25**, 58-69, 1980.

(N. Bäuerle) DEPARTMENT OF MATHEMATICS, KARLSRUHE INSTITUTE OF TECHNOLOGY, D-76128 KARLSRUHE, GERMANY

E-mail address: nicole.baeuerle@kit.edu

(D. Lange) DEPARTMENT OF MATHEMATICS, KARLSRUHE INSTITUTE OF TECHNOLOGY, D-76128 KARLSRUHE, GERMANY

E-mail address: dirk.lange@kit.edu