Optimal Control of Piecewise Deterministic Markov Processes with Finite Time Horizon

Nicole Bäuerle*[1] and Ulrich Rieder[2]

[1] Karlsruhe Institute of Technology, 76128 Karlsruhe, Germany
nicole.baeuerle@kit.edu

[2] University of Ulm, 89069 Ulm, Germany
ulrich.rieder@uni-ulm.de

Abstract. In this paper we study controlled Piecewise Deterministic Markov Processes with finite time horizon and unbounded rewards. Using an embedding procedure we reduce these problems to discrete-time Markov Decision Processes. Under some continuity and compactness conditions we establish the existence of an optimal policy and show that the value function is the unique solution of the Bellman equation. It is remarkable that this statement is true for unbounded rewards and without any contraction assumptions. Further conditions imply the existence of optimal nonrelaxed controls. We highlight our findings by two examples from financial mathematics.

Keywords: Piecewise Deterministic Markov Processes, Markov Decision Process, Bellman Equation, Portfolio Problems.
AMS 2010 subject classification: Primary 90C40, Secondary 60J25, 91G10

1 Introduction

In this paper we deal with optimization problems where the state process is a Piecewise Deterministic Markov Process (PDMP). These processes evolve through random jumps at random time points while the behavior between jumps is governed by an ordinary differential equation. They form a general and important class of non-diffusions. It is assumed that both the jump behavior as well as the drift behavior between jumps can be controlled. Hence this leads to a control problem in continuous-time which can be tackled for example via the Hamilton-Jacobi-Bellman equation. However, since the evolution between jumps is deterministic these problems can also be reduced to a discrete-time Markov Decision Process (MDP) where however the action space is now a function space. Since this will turn out to be a Borel space we can treat these problems with general MDP methods. More precisely we will restrict the presentation to problems with finite time horizon and we will establish two main theorems (Theorem 2 and 3) which state that under some continuity and compactness conditions the

* Corresponding author

value function of the Piecewise Deterministic Markov Decision Process is the unique fixed point of a dynamic programming operator. The reward function is allowed to be unbounded, and we work with bounding functions and a weighted supremum norm.

A first systematic study of controlled PDMP is done in [13, 14]. The idea of reducing the control problems of this type to an MDP is due to [21]. For a recent paper on this topic see [1]. In [22, 23] optimality conditions are given in a weak form based on a continuous-time approach. Davis introduced the name *Piecewise Deterministic Markov Process* (see e.g. [7]) and summarized the state of the art in his book [8]. [19] and [10] extended the existing results to unbounded reward problems. They impose certain assumptions on the drift which imply (using a time transformation) the existence of optimal nonrelaxed controls. Relaxed controls are known from deterministic control and allow to define a topology on the action space (Young topology) which simplifies the task to have a compact action space and continuous functions at the same time. It is well-know that concavity conditions imply the existence of optimal nonrelaxed controls (see e.g. [9], [2]). An important subclass of PDMPs (with uncontrolled drift) is the control of continuous-time Markov Chains (for a recent book on this topic see [11]).

There has been a renewed interest into PDMPs recently, in particular as far as applications in finance, insurance and queueing are concerned. For applications in insurance see in particular [19] and the monograph [20]. [16] used the reduction technique to solve a hedging problem in a continuous-time jump market (with uncontrolled drift). A utility maximization problem in a PDMP financial market is treated in [3]. Such financial markets are extensively studied in [15]. Applications in queueing can be found e.g. in [17], [2] and [18].

The paper is organized as follows: In the next section we introduce our PDMP optimization problem with finite time horizon. This problem is then reduced in Section 3 to a discrete-time MDP. We introduce relaxed controls and show the existence of an optimal policy within this class and that the value function is the unique solution of the Bellman equation. Moreover, we show that in the case of uncontrolled drift or in the case of concavity assumptions, the optimal policy can be found in the smaller class of nonrelaxed controls. In the last section we provide two examples from finance: One example is the maximization of expected terminal wealth in a PDMP financial market and the second example is the liquidation of a large amount of shares in so-called dark pools.

2 Piecewise Deterministic Markov Decision Processes

In this first section we introduce the problem of controlling a Piecewise Deterministic Markov process (PDMP) with finite time horizon T where we restrict to a simple formulation in order to highlight the solution procedure in the following sections. We assume that the state space E of the process is a Borel subset of $\mathbb{R}^d$ and actions can be taken from the *control action space* $\mathcal{U}$ which is assumed to be a Borel subset of a Polish space. The stochastic evolution of the PDMP is based on a marked point process (T_n, Z_n), where (T_n) is the increasing sequence of jump time points of a Poisson process with fixed rate $\lambda > 0$. At these time points the PDMP may also jump. The marks (Z_n) describe the post jump states. We set $T_0 := 0$. Between the jump times T_n and T_{n+1} the process is described by a *deterministic flow* which is constructed from a given drift function $\mu(x, u)$ (see below). A stochastic kernel Q from $E \times \mathcal{U}$ to E describes the distribution of the jump goals, i.e. $Q(B|x, u)$ is the probability that the process jumps in the set B given the state $x \in E$ immediately before the jump and the control action $u \in \mathcal{U}$ at the jump time.

At time T_n the evolution of the process up to time T_{n+1} is known to the decision maker who can therefore fix the control action for $T_n + t < T_{n+1}$ by choosing some

$$\alpha \in A := \{\alpha : \mathbb{R}_+ \to \mathcal{U} \text{ measurable }\}. \tag{1}$$

It is known that A becomes a Borel space if A is endowed with the coarsest σ-algebra such that

$$\alpha \mapsto \int_0^\infty e^{-t} w\big(t, \alpha_t\big) dt$$

is measurable for all bounded and measurable functions $w : \mathbb{R}_+ \times \mathcal{U} \to \mathbb{R}$ (see e.g. [21]).

We assume that for all $\alpha \in A$ there exists a unique solution $\phi_t^\alpha(x) \in E$ of the initial value problem

$$dx_t = \mu(x_t, \alpha_t)dt, \quad x_0 = x \in E.$$

Then $\phi_t^\alpha(x)$ is the state of the piecewise deterministic process at time $T_n + t < T_{n+1}$ if $Z_n = x$. It is assumed that $\phi_t^\alpha(x)$ is measurable in (x, α) and continuous in t.

We restrict here to *Markovian policies* (or piecewise open loop policies) $\pi = (\pi_t)$ which are defined by a sequence of measurable functions $f_n : \mathbb{R}_+ \times E \to A$ such that

$$\pi_t = f_n(T_n, Z_n)(t - T_n) \text{ for } t \in (T_n, T_{n+1}].$$

Note that $f : \mathbb{R}_+ \times E \to A$ is measurable if and only if there exists a measurable function $\tilde{f} : \mathbb{R}_+ \times E \times \mathbb{R}_+ \to \mathcal{U}$ such that

$$f(t, x)(s) = \tilde{f}(t, x, s) \quad \text{for } s, t \in \mathbb{R}_+, x \in E.$$

In the sequel we will not distinguish between f and $\tilde{f}$. It can be shown that more general policies which depend on the complete history of the process do not increase the value of our maximization problem.

For a policy π we write $\pi = (\pi_t) = (f_n)$ and

$$\phi^\pi_{t-T_n}(Z_n) := \phi^{f_n(T_n, Z_n)}_{t-T_n}(Z_n) \text{ for } t \in [T_n, T_{n+1}).$$

Then the piecewise deterministic process (X_t) is given by

$$X_t = \phi^\pi_{t-T_n}(Z_n) \text{ for } t \in [T_n, T_{n+1}).$$

Note that $Z_n = X_{T_n}$.

Given a policy π and an initial state $x \in E$ there is a probability space $(\Omega, \mathcal{F}, \mathbb{P}^\pi_x)$ on which the random variables T_n and Z_n are defined such that $X_0 = Z_0 = x$ and for all Borel sets $C \subset E$ and $t \geq 0$

$$\begin{aligned}
&\mathbb{P}^\pi_x\big(T_{n+1} - T_n \leq t, Z_{n+1} \in C \mid T_0, Z_0, \dots, T_n, Z_n\big) = \\
&= \lambda \int_0^t e^{-\lambda s} Q\big(C | X_{T_n+s}, \pi_{T_n+s}\big) ds = \\
&= \lambda \int_0^t e^{-\lambda s} Q\big(C | \phi^\pi_s(Z_n), f_n(T_n, Z_n)(s)\big) ds.
\end{aligned}$$

Moreover we assume that there is a measurable reward function $r : E \times \mathcal{U} \to \mathbb{R}$ such that $r(x,u)$ gives the reward rate in state x if control action u is taken. Since we consider here control problems with a finite time period $[0,T]$ we have a measurable terminal reward $g : E \to \mathbb{R}$. We impose now the following (we set $x^+ := \max(x,0)$)

Integrability Assumption (A):

$$\sup_\pi \mathbb{E}^\pi_x \left[\int_0^T r^+(X_s, \pi_s) ds + g^+(X_T) \right] < \infty, \quad x \in E.$$

Then the expected total reward when we start at time t in state x is well-defined for all π by

$$V_\pi(t,x) := \mathbb{E}^\pi_{tx} \left[\int_t^T r(X_s, \pi_s) ds + g(X_T) \right], \quad x \in E, t \in [0,T]$$

where $\mathbb{E}_{t,x}$ denotes the conditional expectation that $X_t = x$. The value function of the Piecewise Deterministic Markov Decision Process (PDMDP) is given by

$$V(t,x) := \sup_\pi V_\pi(t,x), \quad x \in E, t \in [0,T] \tag{2}$$

where the supremum is taken over all Markovian policies.

It holds that $V_\pi(T,x) = g(x) = V(T,x)$.

Remark 1. We restrict here the presentation to the problem of maximizing the expected integrated reward rate over a finite time interval. The theory also allows to include instantaneous rewards at the jump time points, i.e. to look at the objective

$$\mathbb{E}^\pi_x \left[\int_0^T r(X_t, \pi_t) dt \right] + \mathbb{E}^\pi_x \left[\int_0^T \bar{r}(X_{t-}, \pi_t) dN_t \right]$$

where the measurable function $\bar{r} : E \times \mathcal{U} \to \mathbb{R}$ gives the reward for each jump and (N_t) is the Poisson process with rate $\lambda > 0$. PDMDPs with infinite horizon can also be treated (see e.g. [4], Chapter 8).

The optimization problem (2) is a continuous-time control problem. However, we can show that the value function $V(t,x)$ can be obtained by a discrete-time MDP. This point of view implies a number of interesting results. The first one is that under some conditions the value function can be characterized as the unique solution of the Bellman equation. Differentiability of the value function is not needed in contrast to the classical continuous-time stochastic control approach. Second, the existence of an optimal policy is rather easy to prove. Moreover, several different computational approaches arise. Value iteration or Howard's policy improvement algorithm can be used to solve the continuous-time PDMDPs (see e.g. [4]).

3 Solution via a discrete-time Markov Decision Process

We introduce here a discrete-time MDP which is equivalent to the control problem of the previous section. The idea is to look at the time points (T_n) and choose actions $\alpha \in A$ at time T_n, since the evolution of the state process between jumps is deterministic. For general MDP theory see e.g. [6], [12] and [4].

Now suppose a Piecewise Deterministic Markov Decision Model is given as described in the previous section. Let us define the following infinite-stage Markov Decision Model (E', A, Q', r'):

- $E' = [0,T] \times E$ is the *state space*. A state (t,x) describes the time point of a jump and the state of the process directly after the jump.
- A is the action space. Recall that the function space A is a Borel space.
- For all Borel sets $B \subset \mathbb{R}_+, C \subset E$ and $(t,x) \in E'$, $\alpha \in A$, the stochastic kernel Q' is given by
$$Q'\big(B \times C | t,x,\alpha\big) := \lambda \int_0^{T-t} e^{-\lambda s} 1_B(t+s) Q\big(C | \phi_s^\alpha(x), \alpha_s\big) ds. \tag{3}$$
This is obviously a substochastic transition law. In order to make it stochastic, we may add an artificial absorbing state Δ to the state space and define
$$Q'\big(\{\Delta\} | t,x,\alpha\big) := e^{-\lambda(T-t)}, \quad Q'\big(\{\Delta\} | \Delta, \alpha\big) := 1.$$
- The reward function $r' : E' \times A \to \mathbb{R}$ is defined by
$$r'(t,x,\alpha) := \int_0^{T-t} e^{-\lambda s} r\big(\phi_s^\alpha(x), \alpha_s\big) ds + e^{-\lambda(T-t)} g\big(\phi_{T-t}^\alpha(x)\big). \tag{4}$$
$$r'(x,\Delta) := 0.$$

In what follows we treat the problem as a substochastic problem and skip the state Δ. As usual we define now for this discrete-time Markov Decision Model for a policy (f_n):

$$J'_{\infty(f_n)}(t,x) = \mathbb{E}_{t,x}^{(f_n)} \left[\sum_{n=0}^{\infty} r'\big(T'_n, Z'_n, f_n(T'_n, Z'_n)\big) \right]$$
$$J'_\infty(t,x) = \sup_{(f_n)} J'_{\infty(f_n)}(t,x), \quad (t,x) \in E'.$$

where (T_n', Z_n') is the corresponding state process of the MDP up to absorption in Δ. Note that $(T_n', Z_n') = (T_n, Z_n)$ as long as $T_n \le T$.

Theorem 1. *For a Markovian policy $\pi = (f_n)$ we have*

$$V_\pi(t,x) = J'_{\infty (f_n)}(t,x), \quad (t,x) \in E'.$$

Moreover, it holds: $V = J'_\infty$.

Proof. Let $H_n := (T_0, Z_0, \ldots, T_n, Z_n)$ and $T_n \le T$. We consider only the time point $t = 0$. Arbitrary time points can be treated similarly by adjusting the notation. We obtain:

$$\begin{aligned} V_\pi(0,x) &= \mathbb{E}_x^\pi \left[\sum_{n=0}^\infty \Big(\int_{T_n \wedge T}^{T_{n+1} \wedge T} r(X_s, \pi_s) ds + 1_{[T_n \le T < T_{n+1}]} g(X_T) \Big) \right] \\ &= \sum_{n=0}^\infty \mathbb{E}_x^\pi \left[\mathbb{E}_x^\pi \left[\int_{T_n}^{T_{n+1} \wedge T} r(X_s, \pi_s) ds \Big| H_n \right] \right] \\ &\quad + \sum_{n=0}^\infty \mathbb{E}_x^\pi \left[\mathbb{E}_x^\pi \left[1_{[T_n \le T < T_{n+1}]} g(X_T) | H_n \right] \right] \\ &= \sum_{n=0}^\infty \mathbb{E}_x^{(f_n)} \left[r'\big(T_n', Z_n', f_n(T_n', Z_n')\big) \right] \end{aligned}$$

since the transition kernel of (T_n', Z_n') is given by (3) and r' by (4). □

Theorem 1 implies that $V(t,x)$ can be computed from the value function J'_∞ of the discrete-time Markov Decision Model. Note that the Integrability Assumption for the Piecewise Deterministic Markov Decision Process implies that the discrete-time Markov Decision Model is well-defined. The maximal reward operator $\mathcal{T}$ is given by

$$\begin{aligned} (\mathcal{T} v)(t,x) = \sup_{\alpha \in A} \Big\{ & e^{-\lambda(T-t)} g\big(\phi^\alpha_{T-t}(x)\big) + \\ & + \int_0^{T-t} e^{-\lambda s} \Big[r\big(\phi^\alpha_s(x), \alpha_s\big) + \lambda \int v(t+s, z) Q\big(dz | \phi^\alpha_s(x), \alpha_s\big) \Big] ds \Big\}. \end{aligned}$$

From now on we assume that $\mathcal{U}$ is compact. In order to prove existence of optimal controls we need certain continuity and compactness conditions. To achieve this, we have to enlarge the action space and we introduce

$$\mathcal{R} := \{\alpha : \mathbb{R}_+ \to \mathbb{P}(\mathcal{U}) \text{ measurable}\}, \tag{5}$$

the set of *relaxed controls* where $\mathbb{P}(\mathcal{U})$ is the set of all probability measures on $\mathcal{U}$ equipped with the σ-algebra of the Borel sets, i.e. α_t can be seen as a randomized action. The problem is to define a topology on A which allows for a compact action space and a continuous target function - two competing aims. The set

A of deterministic controls is a measurable subset of $\mathcal{R}$ in the sense that for $\alpha \in A$ the measures α_t are one-point measures on $\mathcal{U}$. A suitable topology on $\mathcal{R}$ is given by the so-called *Young topology*. Definition and important properties of this topology are summarized in the following remark. It can be shown that the set A of deterministic controls is dense in $\mathcal{R}$ with respect to the Young topology.

Remark 2 (Young Topology). The Young topology on $\mathcal{R}$ is the coarsest topology such that all mappings of the form

$$\mathcal{R} \ni \alpha \mapsto \int_0^\infty \int_{\mathcal{U}} w(t,u)\alpha_t(du)dt$$

are continuous for all functions $w : [0,\infty] \times \mathcal{U} \to \mathbb{R}$ which are continuous in the second argument and measurable in the first argument and satisfy

$$\int_0^\infty \max_{u \in \mathcal{U}} |w(t,u)| dt < \infty.$$

We denote this class by $Car(\mathbb{R}_+ \times \mathcal{U})$, the so-called Carathéodory functions. With respect to the Young topology $\mathcal{R}$ is a separable metric and compact Borel space. In order to have well-defined integrals the following characterizations of measurability are important:

(i) A function $\alpha : \mathbb{R}_+ \to \mathbb{P}(\mathcal{U})$ is measurable if and only if

$$t \mapsto \int_{\mathcal{U}} v(u)\alpha_t(du)$$

is measurable for all bounded and continuous $v : \mathcal{U} \to \mathbb{R}_+$.

(ii) A function $f : E' \to \mathcal{R}$ is measurable if and only if

$$(t,x) \mapsto \int_{\mathbb{R}_+} \int_{\mathcal{U}} w(s,u) f(t,x,s;du)ds$$

is measurable for all $w \in Car(\mathbb{R}_+ \times \mathcal{U})$.

Moreover, the following characterization of convergence in $\mathcal{R}$ is crucial for our applications. Suppose $(\alpha_n) \subset \mathcal{R}$ and $\alpha \in \mathcal{R}$. Then $\lim_{n\to\infty} \alpha_n = \alpha$ if and only if

$$\lim_{n\to\infty} \int_0^\infty \int_{\mathcal{U}} w(t,u)\alpha_t^n(du)dt = \int_0^\infty \int_{\mathcal{U}} w(t,u)\alpha_t(du)dt$$

for all $w \in Car(\mathbb{R}_+ \times \mathcal{U})$.

Now we have to extend the domain of functions already defined on A. In particular we define for $\alpha \in \mathcal{R}$

$$d\phi_t^\alpha(x) = \int \mu\big(\phi_t^\alpha(x), u\big)\alpha_t(du)dt, \quad \phi_0^\alpha(x) = x, \tag{6}$$

$$r'(t,x,\alpha) = e^{-\lambda(T-t)} g\big(\phi_{T-t}^\alpha(x)\big) + \int_0^{T-t} e^{-\lambda s} \int r\big(\phi_s^\alpha(x), u\big)\alpha_s(du)ds,$$

$$Q'(B \times C|t,x,\alpha) = \lambda \int_0^{T-t} e^{-\lambda s} 1_B(t+s) \int Q\big(C|\phi_s^\alpha(x), u\big)\alpha_s(du)ds$$

where we again assume that a unique solution of (6) exists (according to the Theorem of Carathéodory this is the case if e.g. $\mu(x,u)$ is Lipschitz-continuous in x uniformly in u). If α belongs to the subset A, then the definitions of ϕ^α, r' and Q' coincide with those we have used so far. In case α is a relaxed control there is no physical interpretation of the model. The operator $\mathcal{T}$ has the following form:

$$(\mathcal{T}v)(t,x) = \sup_{\alpha\in\mathcal{R}} \Big\{ e^{-\lambda(T-t)} g\big(\phi^\alpha_{T-t}(x)\big) + \\ + \int_0^{T-t} \Big[e^{-\lambda s} \int \Big(r\big(\phi^\alpha_s(x),u\big) + \lambda \int v(t+s,z) Q\big(dz|\phi^\alpha_s(x),u\big)\Big) \alpha_s(du)\Big] ds \Big\}.$$

In the Markov Decision Model with relaxed controls the decision maker can thus do at least as well as in the case without relaxed controls. When we denote by J^{rel}_∞ the corresponding value function we obtain

$$J^{rel}_\infty(t,x) \geq J'_\infty(t,x) = V(t,x), \quad (t,x)\in E'.$$

We will show that these value functions are equal under some conditions (cp. Theorem 3). Next we introduce the notion of a bounding function for the PDMDP:

Definition 1. *A measurable function $b: E \to \mathbb{R}_+$ is called a* bounding function *for the Piecewise Deterministic Markov Decision Model, if there exist constants $c_r, c_g, c_Q, c_\phi \in \mathbb{R}_+$ such that*

(i) $|r(x,u)| \leq c_r b(x)$ *for all* $(x,u) \in E \times \mathcal{U}$.
(ii) $|g(x)| \leq c_g b(x)$ *for all* $x \in E$.
(iii) $\int b(z) Q(dz|x,u) \leq c_Q b(x)$ *for all* $(x,u)\in E\times\mathcal{U}$.
(iv) $b\big(\phi^\alpha_t(x)\big) \leq c_\phi b(x)$ *for all* $(t,x,\alpha)\in E'\times\mathcal{R}$.

If b is a bounding function for the PDMDP, then

$$b(t,x) := b(x) e^{\gamma(T-t)} \text{ for } \gamma \geq 0$$

is a bounding function for the discrete-time MDP (with and without relaxed controls), since

$$|r'(t,x,\alpha)| \leq b(t,x) c_\phi \big(\frac{c_r}{\lambda} + c_g\big)$$
$$\int b(s,z) Q'(d(s,z)|t,x,\alpha) \leq b(t,x) c_Q c_\phi \frac{\lambda}{\lambda+\gamma}\big(1 - e^{-(\lambda+\gamma)T}\big).$$

For a measurable $v: E' \to \mathbb{R}$ we denote the *weighted supremum norm* by

$$\|v\|_b := \sup_{(t,x)\in E'} \frac{|v(t,x)|}{b(t,x)}$$

(with the convention $\frac{0}{0} := 0$) and define the set

$$\mathbb{B}_b := \{v: E' \to \mathbb{R} \mid v \text{ is measurable and } \|v\|_b < \infty\}.$$

Moreover, let us define

$$\alpha_b := \sup_{(t,x,\alpha)\in E'\times\mathcal{R}} \frac{\int b(s,z)Q'(d(s,z)|t,x,\alpha)}{b(t,x)}.$$

From the preceding considerations it follows that

$$\alpha_b \le c_Q c_\phi \frac{\lambda}{\lambda+\gamma}\big(1-e^{-(\lambda+\gamma)T}\big).$$

Hence when a bounding function exists, we have $\alpha_b < 1$ for γ large. From now on we assume that $\alpha_b < 1$. The existence of a bounding function then implies

$$\sup_\pi \mathbb{E}_{tx}^\pi \left[\sum_{k=n}^\infty \big|r'\big(T_k', Z_k', f_k(T_k', Z_k')\big)\big|\right] \le \frac{\alpha_b^n}{1-\alpha_b} b(t,x), \quad (t,x)\in E',$$

which means that the discrete-time MDP (with and without relaxed controls) is contracting. Moreover, it holds

$$\|\mathcal{T}v - \mathcal{T}w\|_b \le \alpha_b \|v-w\|_b, \quad v,w\in \mathbb{B}_b.$$

We introduce next the following

Continuity and Compactness Assumptions:

(i) $\mathcal{U}$ is compact.
(ii) $(t,x,\alpha)\mapsto \phi_t^\alpha(x)$ is continuous on $E'\times\mathcal{R}$.
(iii) $(x,u) \mapsto \int v(z)Q(dz|x,u)$ is continuous for all continuous v on E with $|v(x)|\le c_v b(x)$ for some $c_v \ge 0$.
(iv) $(x,u)\mapsto r(x,u)$ is continuous.
(v) $x\mapsto g(x)$ is continuous.

It is possible to weaken the continuity assumptions to upper semicontinuity (see e.g. the recent book [4]).

Lemma 1. *Let $b(x)$ be a continuous bounding function and $(t,x,\alpha)\mapsto\phi_t^\alpha(x)$ be continuous. Let $w: E'\times\mathcal{U}\to\mathbb{R}$ be continuous with $|w(t,x,u)|\le c_w b(t,x)$ for some $c_w\ge 0$. Then*

$$(t,x,\alpha)\mapsto \int_0^{T-t} e^{-\lambda s}\left(\int w\big(t+s,\phi_s^\alpha(x),u\big)\alpha_s(du)\right) ds$$

is continuous on $E'\times\mathcal{R}$.

Proof. First we prove that the function

$$W(t,x,\alpha) := \int_0^{T-t} e^{-\lambda s}\left(\int w\big(t+s,\phi_s^\alpha(x),u\big)\alpha_s(du)\right) ds$$

is bounded and continuous if w is bounded and continuous. Boundedness is obvious. Now suppose $(t^n, x^n, \alpha^n) \to (t, x, \alpha)$. Let $w_n(s,u) := w\big(t^n + s, \phi_s^{\alpha^n}(x^n), u\big)$ and $w(s,u) := w\big(t+s, \phi_s^\alpha(x), u\big)$. We consider with $a_n := \min\{T-t, T-t^n\}$ and $b_n := \max\{T-t, T-t^n\}$:

$$\begin{aligned}
&|W(t^n,x^n,\alpha^n) - W(t,x,\alpha)| \leq \\
\leq &\Big| \int_{a_n}^{b_n} e^{-\lambda s} \int w_n(s,u)\alpha_s^n(du)ds\Big| + \\
&+ \int_0^{T-t} e^{-\lambda s} \int |w_n(s,u) - w(s,u))|\alpha_s^n(du)ds + \\
&+ \Big| \int_0^{T-t} e^{-\lambda s} \int w(s,u)\alpha_s^n(du)ds - \int_0^{T-t} e^{-\lambda s} \int w(s,u)\alpha_s(du)ds\Big|.
\end{aligned}$$

The first term on the right-hand side converges to zero for $n \to \infty$ since the integrand is bounded. The second term can be further bounded by

$$\int_0^{T-t} e^{-\lambda s} \sup_{u \in \mathcal{U}} |w_n(s,u) - w(s,u)| ds$$

which converges to zero for $n \to \infty$ due to dominated convergence and the continuity of ϕ and w. The last term converges to zero in view of the definition of convergence w.r.t. the Young topology and the fact that w is continuous.

Now let w be continuous with $|w| \leq c_w b$. Then $w^b(t,x,u) := w(t,x,u) - c_w b(t,x) \leq 0$ and continuous. According to Lemma 7.14 in [6], there exists a sequence (w_n^b) of bounded and continuous functions with $(w_n^b) \downarrow w^b$. From the first part of the proof we know that

$$W_n(t,x,\alpha) := \int_0^{T-t} e^{-\lambda s} \left(\int w_n^b\big(t+s, \phi_s^\alpha(x), u\big)\alpha_s(du) \right) ds$$

is bounded and continuous and decreases for $n \to \infty$ against

$$W(t,x,\alpha) - c_w \int_0^{T-t} e^{-\lambda s} b\big(t+s, \phi_s^\alpha(x)\big) ds$$

which is thus upper semicontinuous. Since b is a continuous bounding function, it follows from generalized dominated convergence that

$$(t,x,\alpha) \mapsto \int_0^{T-t} e^{-\lambda s} b\big(t+s, \phi_s^\alpha(x)\big) ds$$

is continuous which implies that W is upper semicontinuous. Considering the function $w_b(t,x,u) := -c_w b(t,x) - w(t,x,u) \leq 0$ in the same way we obtain that W is lower semicontinuous, thus W is continuous. $\square$

Now we are able to formulate the main results for the control problem (with relaxed controls). Denote

$$\mathbb{C}_b(E') := \{ v \in I\!B_b \mid v \text{ is continuous } \}.$$

Theorem 2. *Suppose the Piecewise Deterministic Markov Decision Process has a continuous bounding function b and the continuity and compactness assumptions are satisfied. Then it holds:*

a) $J_\infty^{rel} \in \mathbb{C}_b(E')$ *and* J_∞^{rel} *is the unique fixed point of* $\mathcal{T}$ *in* $\mathbb{C}_b(E')$.
b) *There exists an optimal relaxed control* $\pi^* = (\pi_t^*)$ *such that*

$$\pi_t^* = f(T_n, Z_n)(t - T_n) \quad \textit{for } t \in (T_n, T_{n+1}]$$

for a measurable control function $f : E' \to \mathcal{R}$.

Proof. Recall from Remark 2 that $\mathcal{R}$ is compact. Then it follows from Lemma 1 that

$$(t, x, \alpha) \mapsto r'(t, x, \alpha) + \int v(s, z) Q'(d(s, z)|t, x, \alpha)$$

is continuous for $v \in \mathbb{C}_b(E')$. This implies that $\mathcal{T} : \mathbb{C}_b(E') \to \mathbb{C}_b(E')$. Moreover $\mathbb{C}_b(E')$ is a closed subset of the Banach space $I\!B_b$. Hence the statement follows with Banach's fixed point theorem (see also Theorem 7.3.6 in [4]). □

Note that the optimal control π_t^* takes values in $\mathbb{P}(\mathcal{U})$. In applications the existence of optimal *nonrelaxed* controls is more interesting. Here we are able to prove the following result:

Theorem 3. *Suppose the Piecewise Deterministic Markov Decision Process has a continuous bounding function b and the continuity and compactness assumptions are satisfied. If $\phi_t^\alpha(x)$ is independent of α (uncontrolled flow) or if $\mathcal{U}$ is convex, $\mu(x, u)$ linear in u and*

$$u \mapsto r(x, u) + \lambda \int J_\infty^{rel}(t, z) Q(dz|x, u)$$

is concave on $\mathcal{U}$, then there exists an optimal nonrelaxed *policy* $\pi^* = (\pi_t^*)$ *such that*

$$\pi_t^* = f(T_n, Z_n)(t - T_n) \quad \textit{for } t \in (T_n, T_{n+1}]$$

for a measurable control function $f : E' \to A$. Note that π_t^* takes values in $\mathcal{U}$ and that $V = J_\infty' = J_\infty^{rel}$. In particular, V is the unique fixed point of $\mathcal{T}$ in $\mathbb{C}_b(E')$.

Proof. For $v \in \mathbb{C}_b(E')$ define

$$w(t, x, u) := r(x, u) + \lambda \int v(t, z) Q(dz|x, u), \quad (t, x) \in E', u \in \mathcal{U}.$$

Then

$$(Lv)(t, x, \alpha) := e^{-\lambda(T-t)} g\big(\phi_{T-t}^\alpha(x)\big) + \int_0^{T-t} e^{-\lambda s} \int_{\mathcal{U}} w\big(t + s, \phi_s^\alpha(x), u\big) \alpha_s(du) ds$$

and

$$(\mathcal{T}v)(t, x) = \sup_{\alpha \in \mathcal{R}} (Lv)(t, x, \alpha).$$

a) Let $\phi_t^\alpha(x)$ be independent of α (uncontrolled flow). There exists a measurable function $\tilde{f} : E' \to \mathcal{U}$ such that

$$w\big(t,x,\tilde{f}(t,x)\big) = \sup_{u\in\mathcal{U}} w(t,x,u), \quad (t,x) \in E'.$$

Define $f(t,x)(s) := \tilde{f}\big(t+s,\phi_s(x)\big)$ for $s \geq 0$. Then $f : E' \to A$ is measurable and it is easily shown (by a pointwise maximization) that

$$\begin{aligned}\sup_{\alpha\in\mathcal{R}}(Lv)(t,x,\alpha) &= e^{-\lambda(T-t)}g\big(\phi_{T-t}(x)\big) + \int_0^{T-t} e^{-\lambda s} w\big(t+s,\phi_s(x),f(t,x)(s)\big)ds\\ &= (Lv)\big(t,x,f(t,x)\big), \quad (t,x)\in E'.\end{aligned}$$

Hence the statements follow as in the proof of Theorem 2.

b) Let $u \mapsto w(t,x,u)$ be concave on $\mathcal{U}$. There exists a measurable function $f^{rel} : E' \to \mathcal{R}$ such that

$$\sup_{\alpha\in\mathcal{R}}(Lv)(t,x,\alpha) = (Lv)\big(t,x,f^{rel}(t,x)\big), \quad (t,x)\in E'.$$

Define $f(t,x) := \int_{\mathcal{U}} u f^{rel}(t,x)(du)$ for $(t,x) \in E'$. Then $f(t,x) \in A$ since $\mathcal{U}$ is convex, and $f : E' \to A$ is measurable. Moreover, since the drift $\mu(x,u)$ is linear in u we obtain $\phi_t^\alpha = \phi_t^{\bar{\alpha}}$ with $\bar{\alpha}_t := \int u\alpha_t(du)$. From the concavity of $w(t,x,\cdot)$ we conclude

$$(Lv)(t,x,\alpha) \leq e^{-\lambda(T-t)}g\big(\phi_{T-t}^\alpha(x)\big)+\int e^{-\lambda s}w\big(t+s,\phi_s^{\bar{\alpha}}(x),\bar{\alpha}_s\big)ds = (Lv)(t,x,\bar{\alpha})$$

and hence

$$\sup_{\alpha\in\mathcal{R}}(Lv)(t,x,\alpha) = (Lv)\big(t,x,f(t,x)\big), \quad (t,x)\in E'.$$

For $v = J_\infty^{rel}$, the (nonrelaxed) control function f is a maximizer of J_∞^{rel}, hence optimal and $V(t,x) = J_\infty^{rel}(t,x)$. □

4 Applications

In this section we consider two applications in mathematical finance.

4.1 Terminal Wealth Problem

We consider a special class of continuous-time financial markets where asset dynamics follow a PDMP: Suppose there is a Poisson process $N = (N_t)$ with rate $\lambda > 0$ and a sequence of independent and identically distributed random vectors (marks) (Y_n) with values in $(-1,\infty)^d$. The Y_n are assumed to have a distribution Q_Y and to be independent of (N_t). Thus, we can define the $\mathbb{R}^d$-valued compound Poisson process

$$C_t := \sum_{n=1}^{N_t} Y_n.$$

By (C_t^k) we denote the k-th component of this process. We suppose that we have d risky assets and one riskless bond with the following dynamics for $t \in [0, T]$:

- The price process (S_t^0) of the riskless bond is given by

$$S_t^0 := e^{\rho t},$$

  where $r \geq 0$ denotes the continuous interest rate.
- The price processes (S_t^k) of the risky assets $k = 1, \ldots, d$ satisfy the stochastic differential equation:

$$dS_t^k = S_{t-}^k \big(\mu_k dt + dC_t^k\big)$$

  where $\mu_k \in \mathbb{R}$ are given constants. The initial prices S_0^k are assumed to be strictly positive.

In this financial market, the price processes show a deterministic evolution between jumps and the jumps occur at Poisson epochs and have random sizes. If we denote again by $0 := T_0 < T_1 < T_2 < \ldots$ the jump time points of the Poisson process and if $t \in [T_n, T_{n+1})$, then for $k = 1, \ldots, d$

$$S_t^k = S_{T_n}^k \exp\big(\mu_k (t - T_n)\big).$$

At the time of a jump we have

$$S_{T_n}^k - S_{T_n-}^k = S_{T_n-}^k Y_n^k.$$

Thus, Y_n^k gives the relative jump height of asset k at the n-th jump. Since $Y_n^k > -1$ almost surely, our asset prices stay positive. Note that the distribution Q_Y might well have probability mass on points $Y^k e_k,\ k = 1, \ldots, d$ where e_k is the k-th unit vector. The assets are allowed to have common jumps - a way of modeling dependence.

Now we want to invest into this financial market. To ensure that the wealth process stays positive we do not allow short-sellings. Thus, the portfolio strategy can be given in terms of fractions of invested capital. In what follows, an admissible portfolio strategy $\pi = (\pi_t)$ is given as a Markovian policy (see Section 2) with values in $\mathcal{U} := \{u \in \mathbb{R}^d \mid u \geq 0, u \cdot e \leq 1\}$ (we denote $e = (1, \ldots, 1)$ and $x \cdot y$ the scalar product), where $\pi_t = (\pi_t^1, \ldots, \pi_t^d)$ gives the fractions of wealth invested in the stocks at time t. The quantity $1 - \pi_t \cdot e$ is the fraction invested in the bond, thus the portfolio is self-financing. The dynamics of the wealth process is

$$dX_t = X_{t-}\Big(\rho dt + \pi_t \cdot (\bar{\mu} - \rho e)dt + \pi_t dC_t\Big)$$

where $\bar{\mu} := (\mu_1, \dots, \mu_d) \in \mathbb{R}^d$. The wealth process is a controlled PDMP with $\mu(x,u) := x(\rho + u \cdot (\bar{\mu} - \rho e))$, and hence $\phi_t^\alpha(x) = x e^{\int_0^t (\rho + \alpha_s \cdot (\bar{\mu} - \rho e)) ds}$. We obtain the following explicit expression for the wealth process:

$$X_t = x_0 \exp\Big(\int_0^t \big(\rho + \pi_s \cdot (\bar{\mu} - \rho e)\big) ds \Big) \prod_{j=1}^{N_t} \Big(1 + \pi_{T_j} \cdot Y_j\Big).$$

The aim of the investor is now to maximize her expected utility of the terminal wealth. Thus, we denote by $U : (0,\infty) \to \mathbb{R}_+$ a strictly increasing and strictly concave utility function and define for a policy π and $(t,x) \in E' := [0,T] \times (0,\infty)$:

$$V_\pi(t,x) := \mathbb{E}_{t,x}^\pi U(X_T)$$

the expected utility of terminal wealth. The maximal expected utility is given by

$$V(t,x) := \sup_\pi V_\pi(t,x).$$

Obviously we have $V_\pi(T,x) = U(x) = V(T,x)$. Throughout we assume that $\int \|y\| Q_Y(dy) < \infty$ where for $x \in \mathbb{R}^d$, $\|x\| := |x_1| + \ldots + |x_d|$. This implies that V is well-defined. This problem has been considered in [3].

It follows now directly from Section 3 that the corresponding discrete-time Markov Decision Process has transition kernel

$$Q'\big(B \times C \mid t,x,\alpha\big) := \lambda \int_0^{T-t} e^{-\lambda s} \int 1_B(t+s) 1_C\Big(\phi_s^\alpha(x)\big(1 + \alpha_s \cdot y\big)\Big) Q_Y(dy) ds$$

and the one-stage reward function

$$r'\big(t,x,\alpha\big) := e^{-\lambda(T-t)} U\big(\phi_{T-t}^\alpha(x)\big).$$

In this model the dynamic programming operator has the form

$$\begin{aligned}(\mathcal{T}v)(t,x) = \sup_{\alpha \in A} \Big\{ & e^{-\lambda(T-t)} U\big(\phi_{T-t}^\alpha(x)\big) \\ & + \lambda \int_0^{T-t} e^{-\lambda s} \int v\Big(t+s, \phi_s^\alpha(x)\big(1 + \alpha_s \cdot y\big)\Big) Q_Y(dy) ds \Big\}.\end{aligned}$$

Since U is concave it can be bounded by a linear function and thus the discrete-time MDP (with and without relaxed controls) has a bounding function b given by

$$b(t,x) := e^{\gamma(T-t)}(1+x), \quad (t,x) \in E'$$

for $\gamma \geq 0$, and $\alpha_b < 1$ if γ is large enough (see Section 3). Let us define

$$I\!M_{cv} := \{ v \in \mathbb{C}_b(E') \mid v(t,x) \text{ is concave and increasing in } x \text{ and decreasing in } t \}.$$

Note that the continuity and compactness conditions are satisfied. In particular, $\mathcal{U}$ is compact and convex, $(t,x,\alpha) \mapsto \phi_t^\alpha(x)$ is continuous, $b(x)$ is continuous and $\mu(x,u)$ is linear in u. Then we obtain from Theorem 2 and Theorem 3:

Theorem 4. *The following statements hold for the terminal wealth problem:*

a) *The value function $V(t,x)$ is the unique fixed point of $\mathcal{T}$ in $\mathbb{M}_{cv}$.*
b) *There exists an optimal (nonrelaxed) portfolio strategy $\pi^* = (\pi_t^*)$ such that*

$$\pi_t^* = f(T_n, Z_n)(t - T_n) \;\; \text{for} \;\; t \in (T_n, T_{n+1}]$$

for a measurable $f : E' \to A$.

For further properties of the value function see [3].

Example 1 (Power utility). Let $U(x) := x^\beta$ with $0 < \beta < 1$. In this case we obtain the explicit solution:

$$V(t,x) = x^\beta e^{\delta(T-t)}, \quad (t,x) \in E'$$
$$\pi_t^* \equiv u^*, t \in [0,T]$$

where u^* is the maximum point of

$$u \mapsto \beta u \cdot (\bar{\mu} - \rho e) + \lambda \int (1 + u \cdot y)^\beta Q_Y(dy)$$

on $\mathcal{U}$ and $\delta := \beta\rho - \lambda + \beta u^* \cdot (\bar{\mu} - \rho e) + \lambda \int (1 + u^* \cdot y)^\beta Q_Y(dy)$.

4.2 Trade Execution in Illiquid Markets

Suppose we have an agent who wants to sell a large amount of shares during a given time interval. Placing a large order in an order book will certainly lead to a price impact. Moreover in traditional markets, other participants may examine the order book and see the intention of the agent and then may try to trade against her. As a consequence it is recently possible to trade in *dark pools* where there is no order book and orders are matched electronically. This reduces the risk of adverse price manipulations but on the other hand may lead to lower liquidity since there is no market-maker. This problem has been considered in [5] but solved with different methods.

We will set-up a simple mathematical model to describe this situation. Suppose that the agent has initially $x_0 \in \mathbb{N}_0$ shares and is able to sell them in blocks only at the jump time points of a Poisson process to account for illiquidity. The execution horizon is T. All shares which have not been sold until time T will be placed at a traditional market and the order will be executed at once. The cost of selling a shares is given by $C(a)$ where $C : \mathbb{N}_0 \to \mathbb{R}_+$ is strictly increasing and strictly convex and satisfies $C(0) = 0$. Note that strictly convex means that

$$C(x) - C(x-1) < C(x+1) - C(x), \quad x \in \mathbb{N}. \tag{7}$$

The cost function C can be interpreted as a market depth function. Obviously this implies that it is better to sell small blocks, however if there are no trading epochs arriving anymore this will yield a large amount of shares which have to be liquidated at time T.

Let us now formalize this optimization problem: Suppose $N = (N_t)$ is a Poisson-process with fixed intensity $\lambda > 0$ and jump time points $0 = T_0 < T_1 < T_2 < \ldots$. A control process $\pi = (\pi_t) = (f_n)$ is defined as in Section 2 where π_t denotes the number of shares the agent would like to sell at time t, i.e. $\mathcal{U} := \mathbb{N}_0$. This order is only executed if t is also a jump time-point of the Poisson process. The state process (X_t) represents the number of shares which still have to be sold. Thus, if (π_t) is a control process we obtain

$$X_t = x_0 - \int_0^t \pi_s dN_s$$

where $\pi_t \le X_t$ has to be satisfied for all $t \in [0,T]$. A control process with this property is called admissible. The problem is now to minimize the value function

$$V_\pi(t,x) = \mathbb{E}_{tx}^\pi \left[\int_t^T C(\pi_s) dN_s + C(X_T) \right],$$

i.e. to find

$$V(t,x) = \inf_\pi V_\pi(t,x), \quad (t,x) \in [0,T] \times \mathbb{N}_0 =: E'$$

where the infimum is taken over all admissible Markovian policies.

Obviously this control problem is a controlled PDMDP. Since $\phi_t^\alpha(x) = x$ the flow is uncontrolled and we may consider this problem also as a controlled CTMC (cf. [11]). We will solve it by a discrete-time MDP along the lines in Section 3. Let us denote

$$A := \{\alpha : \mathbb{R}_+ \to \mathbb{N}_0 \text{ measurable } \}$$

and $D(x) := \{\alpha \in A \mid \alpha_t \le x \text{ for all}\, t \ge 0\}$ and $D := \{(t,x,\alpha) \in E' \times A | \alpha \in D(x)\}$. In contrast to the previous section we have a restriction on the actions here but this will be easy to handle. We obtain now for a Markovian policy $\pi = (\pi_t) = (f_n)$ that

$$\begin{aligned} V_\pi(t,x) &= \mathbb{E}_{tx}^\pi \left[\int_t^T C(\pi_s) dN_s + C(X_T) \right] \\ &= \sum_{n=1}^\infty \mathbb{E}_{tx}^\pi \left[c\Big(T_n', Z_n', f_n(T_n', Z_n')\Big) \right], \end{aligned}$$

where for $(t,x,\alpha) \in D$

$$c(t,x,\alpha) := \int_0^{T-t} \lambda e^{-\lambda s} C(\alpha_s) ds + e^{-\lambda(T-t)} C(x). \tag{8}$$

Thus we can consider the discrete-time MDP with the substochastic transition kernel Q'

$$Q'\big(B \times C \mid t,x,\alpha\big) := \int_0^{T-t} \lambda e^{-\lambda s} 1_B(t+s) 1_C(x - \alpha_s) ds$$

and one-stage reward function $r'\big(t,x,\alpha\big) := -c\big(t,x,\alpha\big)$ where c is given in (8). Theorem 1 implies that $V = J'_\infty$ and that an optimal policy of the discrete-time MDP defines an optimal control process for the PDMDP.

The function $b(t,x) := C(x), (t,x) \in E'$ is a bounding function for the discrete-time MDP and $\alpha_b < 1$ since

$$|r'\big(t,x,\alpha\big)| \le \int_0^{T-t} \lambda e^{-\lambda s} C(x) ds + e^{-\lambda(T-t)} C(x) = C(x)$$

and

$$\alpha_b = \sup_{(t,x,\alpha)\in D} \frac{\int_0^{T-t} \lambda e^{-\lambda s} C(x-\alpha_s) ds}{C(x)} \le 1 - e^{-\lambda T} < 1.$$

Properties of the value function $V(t,x)$ which can immediately be seen are:

$$V(t,x) \le C(x) \text{ and } V(t,0) = 0 \text{ and } V(T,x) = C(x).$$

Now the dynamic programming operator $\mathcal{T}$ reads for $v \in \mathbb{B}_b$:

$$\begin{aligned}(\mathcal{T}v)(t,x) &= \inf_{\alpha\in D(x)} \left\{ \int_0^{T-t} \lambda e^{-\lambda s}\Big(C(\alpha_s) + v(t+s, x-\alpha_s)\Big) ds + e^{-\lambda(T-t)} C(x) \right\} \\ &= \int_0^{T-t} \lambda e^{-\lambda s} \min_{a\in\{0,\dots,x\}} \Big(C(a) + v(t+s, x-a)\Big) ds + e^{-\lambda(T-t)} C(x).\end{aligned}$$

Since the drift of the PDMDP cannot be controlled, it is not necessary to consider relaxed controls (cp. Theorem 3). Let us denote by

$$f^*(t,x) := argmin_{a\in\{0,\dots,x\}} \Big(C(a) + v(t, x-a)\Big), \quad (t,x) \in E' \tag{9}$$

the smallest minimizer of the right-hand side. Define

$$\mathbb{M}_{cx} := \Big\{ v \in \mathbb{C}_b(E') \mid v(t,x) \le C(x),\ v(t,0) = 0, v \text{ is increasing in } t, x \text{ and convex in } x \Big\}.$$

Then we obtain:

Theorem 5. *The following statements hold for the trade execution problem:*

a) The value function $V(t,x)$ is the unique fixed point of $\mathcal{T}$ in $\mathbb{M}_{cx}$.
b) There exists an optimal control process $\pi^* = (\pi_t^*)$ such that $\pi_t^* = f^*(t,X_t)$ and f^* satisfies $f^*(t,x) \le f^*(t,x+1) \le f^*(t,x)+1$, and (X_t) is the corresponding number of share process.

Proof. We show that $\mathcal{T} : \mathbb{M}_{cx} \to \mathbb{M}_{cx}$ and that for $v \in \mathbb{M}_{cx}$, the minimizer f^* as defined in (9) has the properties $f^*(t,x) \le f^*(t,x+1) \le f^*(t,x)+1$.

Let $v \in \mathbb{M}_{cx}$. Since $v(t,0) = 0$ we obtain $\mathcal{T}v \le C$. That $\mathcal{T}v(t,0) = 0$ is obvious. The continuity of $(t,x) \mapsto \mathcal{T}v(t,x)$ follows immediately from the definition of $\mathcal{T}$. We next prove that $\mathcal{T}v$ is increasing in x, i.e. $\mathcal{T}v(t,x) \le \mathcal{T}v(t,x+1)$, $x \in \mathbb{N}$. This can be seen since

$$\begin{aligned} C(a) + v(t+s,x-a) &\le C(a) + v(t+s,x+1-a), \text{ for } a = 0,\dots,x \\ C(x) + v(t+s,0) &\le C(x+1) + v(t+s,0). \end{aligned}$$

Next we show that $\mathcal{T}v$ is increasing in t. In what follows we write

$$G(t,x) := \min_{a \in \{0,\dots,x\}} \Big(C(a) + v(t,x-a) \Big).$$

Let $t \ge t'$ and consider

$$\begin{aligned} \mathcal{T}v(t,x) - \mathcal{T}v(t',x) = &\int_0^{T-t} \lambda e^{-\lambda s} \Big(G(t+s,x) - G(t'+s,x) \Big) ds + \\ &+ \int_{T-t}^{T-t'} \lambda e^{-\lambda s} \Big(C(x) - G(t'+s,x) \Big) ds. \end{aligned}$$

Let $a^* = f^*(t+s,x)$ then we obtain

$$G(t+s,x) - G(t'+s,x) \ge C(a^*) + v(t+s,x-a^*) - C(a^*) - v(t'+s,x-a^*) \ge 0$$

and we obviously have

$$C(x) - G(t'+s,x) \ge C(x) - C(x) - v(t'+s,0) = 0$$

which implies that $\mathcal{T}v$ is increasing in t.

Next we show that $f^*(t,x+1) \le f^*(t,x)+1$. If $f^*(t,x) = x$ the statement is clear, so suppose $a^* := f^*(t,x) \le x-1$. Now suppose there exists an $a > a^*+1$ with

$$C(a) + v(t,x+1-a) < C(a^*+1) + v(t,x+1-(a^*+1)).$$

This implies

$$C(a-1) - C(a^*) < C(a) - C(a^*+1) < v(t,x-a^*) - v(t,x+1-a)$$

and hence

$$C(a-1) + v(t,x-(a-1)) < C(a^*) + v(t,x-a^*)$$

which contradicts the definition of a^*.

The remaining two statements $\mathcal{T}v(t,y+1) - \mathcal{T}v(t,y) \ge \mathcal{T}v(t,y) - \mathcal{T}v(t,y-1)$ and $f^*(t,y+1) \ge f^*(t,y)$ for $y \in \mathbb{N}$ are simultaneously shown by induction on y. For $y = 1$ we have

$$\mathcal{T}v(t,2) - \mathcal{T}v(t,1) \ge \mathcal{T}v(t,1) - \mathcal{T}v(t,0)$$

and $f^*(t,2) = 1 = f^*(t,1)$. Suppose the statement is true for $y = 1,\ldots,x-1$. Let $a^* = f^*(t,x) \geq 1$. Suppose there exists an $0 < a < a^*$ (an easy argument gives us that $a = 0$ cannot be optimal) with

$$C(a) + v(t,x+1-a) \leq C(a^*) + v(t,x+1-a^*).$$

This implies that

$$C(a^*) - C(a) \geq v(t,x+1-a) - v(t,x+1-a^*) \geq v(t,x-a) - v(t,x-a^*)$$

where the last inequality follows from the induction hypothesis. Hence we conclude that

$$C(a^*) + v(t,x-a^*) \geq C(a) + v(t,x-a)$$

which is a contradiction to the definition of a^* and we obtain $f^*(t,x+1) \geq f^*(t,x)$. Now we have to show that $\mathcal{T}v(t,x+1) - \mathcal{T}v(t,x) \geq \mathcal{T}v(t,x) - \mathcal{T}v(t,x-1)$. Due to the convexity of C the statement is true when

$$G(t,x+1) - G(t,x) \geq G(t,x) - G(t,x-1).$$

Let us denote $f^*(t,x) =: a^* > 0$ and $b^* = f^*(t,x-1)$. Then $b^* \leq a^* \leq b^* + 1$, i.e. $b^* \geq a^* - 1$. We discern the following cases:

Case 1: $f^*(t,x+1) = a^*$.
Thus we have

$$\begin{aligned} G(t,x) - G(t,x-1) &\leq v(t,x-b^*) - v(t,x-1-b^*) \leq \\ &\leq v(t,x+1-a^*) - v(t,x-a^*) = G(t,x+1) - G(t,x). \end{aligned}$$

Case 2: $f^*(t,x+1) = a^* + 1$.
Here we have

$$\begin{aligned} G(t,x) - G(t,x-1) &\leq v(t,x-b^*) - v(t,x-1-b^*) \leq \\ &\leq v(t,x-a^*) - v(t,x-a^*-1) \leq G(t,x+1) - G(t,x). \end{aligned}$$

$\mathbb{M}_{cx}$ is a closed subset of the Banach space $\mathbb{B}_b$. Hence $V(t,x)$ is the unique fixed point of $\mathcal{T}$ in $\mathbb{M}_{cx}$. For $v = V$, the control function $f : E' \to A$ defined by

$$f(t,x)(s) := f^*(t+s,x) \ \text{ for } \ s \geq 0$$

is a minimizer of V. Then by Theorem 1 the control process $\pi^* = (\pi_t^*)$ is optimal where for $t \in (T_n, T_{n+1}]$

$$\pi_t^* = f(T_n, Z_n)(t - T_n) = f^*(t, Z_n) = f^*(t, X_t).$$

□

Remark 3. In [5] it is also shown that the optimal $f^*(t,x)$ is increasing in t and jumps only by sizes one, i.e. there are thresholds $0 \leq t_1(x) < t_2(x) < \ldots < t_x(x)$ such that if we have x shares we try to sell k between time $t_{k-1}(x)$ and $t_k(x)$.

References

1. Almudevar, A.: A dynamic programming algorithm for the optimal control of piecewise deterministic Markov processes. SIAM J. Control Optim. **40** (2001) 525-539
2. Bäuerle, N.: Discounted stochastic fluid programs. Math. Oper. Res., **26** (2001), 401-420
3. Bäuerle, N. and Rieder, U.: MDP Algorithms for Portfolio Optimization Problems in Pure Jump Markets. Finance Stoch. **13** (2009) 591-611
4. Bäuerle, N. and Rieder, U.: Markov Decision Processes with Applications to finance. Springer, Heidelberg, 2011
5. Bayraktar, E. and Ludkovski, M.: Optimal trade execution in illiquid financial markets. Math. Finance. To appear (2010)
6. Bertsekas, D.P. and Shreve, S.E. Stochastic optimal control. Academic Press, NY, 1978
7. Davis, M.H.A. Piecewise-deterministic Markov processes: a general class of non-diffusion stochastic models. J. Roy. Statist. Soc. Ser. B **46** (1984) 353–388
8. Davis, M.H.A. Markov models and optimization. Chapman & Hall, London, 1993
9. Dempster, M.A.H. and Ye, J.J.: Necessary and sufficient optimality conditions for control of piecewise deterministic Markov processes. Stochastics Stochastics Rep. **40** (1992) 125-145
10. Forwick, L., Schäl, M. and Schmitz, M.: Piecewise deterministic Markov control processes with feedback controls and unbounded costs. Acta Appl. Math. **82** (2004) 239-267
11. Guo, X. and Hernández-Lerma, O.: Continuous-time Markov Decision Processes. Springer, New York, 2009
12. Hernández-Lerma, O. and Lasserre, J.B. Discrete-time Markov control processes. Springer-Verlag, NY, 1996
13. Hordijk, A. and van der Duyn Schouten, F.A.: Discretization and weak convergence in Markov decision drift processes. Math. Oper. Res. **9** (1984) 112-141
14. Hordijk, A. and van der Duyn Schouten, F.A.: Markov decision drift processes: Conditions for optimality obtained by discretization. Math. Oper. Res. **10** (1985) 160-173
15. Jacobsen, M.: Point process theory and applications. Birkhäuser Boston Inc., Boston, MA, 2006
16. Kirch, M. and Runggaldier, W.J.: Efficient hedging when asset prices follow a geometric Poisson process with unknown intensities. SIAM J. Control Optim. **43** (2005)
17. Kitaev, M.Y. and Rykov, V.V.: Controlled queueing systems. CRC Press, Boca Raton, 1995
18. Rieder, U. and Winter, J.: Optimal control of Markovian jump processes with partial information and applications to a parallel queueing model. Math. Methods Oper. Res. **70** (2009) 567–596
19. Schäl, M.: On piecewise deterministic Markov control processes: control of jumps and of risk processes in insurance. Insurance Math. Econom. **22** (1998) 75–91
20. Schmidli, H.: Stochastic control in insurance. Springer, London, 2008
21. Yushkevich, A.A.: On reducing a jump controllable Markov model to a model with discrete time. Theory Probab. Appl. **25** (1980) 58-69
22. Yushkevich, A.A.: Bellman inequalities in Markov decision deterministic drift processes. Stochastics **23** (1987) 25-77

23. Yushkevich, A.A.: Verification theorems for Markov decision processes with controllable deterministic drift, gradual and impulse controls. Teor. Veroyatnost. i Primenen. **34** (1989) 528-551