MDP Algorithms for Portfolio Optimization Problems in pure Jump Markets

Nicole Bäuerle [*], Ulrich Rieder [†]

Abstract

We consider the problem of maximizing the expected utility of the terminal wealth of a portfolio in a continuous-time pure jump market with general utility function. This leads to an optimal control problem for *Piecewise Deterministic Markov Processes.* Using an embedding procedure we solve the problem by looking at a discrete-time contracting Markov Decision Process. Our aim is to show that this point of view has a number of advantages, in particular as far as computational aspects are concerned. We characterize the value function as the unique fixed point of the dynamic programming operator and prove the existence of optimal portfolios. Moreover, we show that value iteration as well as Howard's policy improvement algorithm work. Finally we give error bounds when the utility function is approximated and when we discretize the state space. A numerical example is presented and our approach is compared to the approximating Markov chain method.

AMS 2000 Subject Classification: Primary 91B28, 93E20 Secondary 90C39, 60G55

Keywords: Portfolio Optimization, Piecewise Deterministic Markov Processes, Markov Decision Process, Operator Fixed Points, Approximation Algorithms.

---

[*]Institut für Stochastik, Universität Karlsruhe (TH), D-76128 Karlsruhe, Germany, `email: baeuerle@stoch.uni-karlsruhe.de`

[†]Institut für Optimierung und Operations Research, Universität Ulm, D-89069 Ulm, Germany, `email: ulrich.rieder@uni-ulm.de`

1 Introduction

In this paper we consider the classical problem of maximizing the expected utility of the terminal wealth of a portfolio in a financial market. However, in contrast to most papers on this subject, we consider a continuous-time pure jump market. More precisely, we assume that jumps in the stock prices arrive according to a Poisson process and that the relative jump heights of the stocks are iid vectors. The behavior of the stock prices between jumps is purely deterministic. Thus, the price processes fall in the class of so-called *Piecewise deterministic Markov processes* (PDMP). It is well-known that the model parameters can be chosen such that this market model approximates a Lévy market arbitrarily well. We restrict our exposition to a fairly simple model but extensions are also discussed which do not complicate the analysis. Note that we do not assume a special form for the utility function. This model has various advantages over jump-diffusion markets, in particular as far as computational aspects are concerned. By considering the embedded jump process, the optimization problem can be formulated as a discrete-time Markov Decision Process (MDP). Our main aim is to show that this point of view implies a number of interesting results:

(i) The first one is that the value function can be characterized as the fixed point of a dynamic programming operator. This operator can be shown to be contracting which establishes uniqueness of the fixed point. Differentiability of the value function is not needed.

(ii) The existence of an optimal portfolio strategy is rather easy to prove.

(iii) Established MDP algorithms like value iteration or Howard's policy improvement algorithm can be used to solve the problem as an alternative to the solution of the Hamilton-Jacobi-Bellman equation. In our numerical example the implementation of the value iteration on a grid outperforms the result of the so-called approximating Markov chain method (see Kushner and Dupuis (2001)) which is a standard algorithm for solving stochastic control problems.

(iv) Due to the contraction property of the operator, the convergence of numerical approximation schemes can be shown easily.

The paper is organized as follows: In Section 2 we introduce the market model and the portfolio problem. This market model is nicely presented in Jacobsen (2006), however has been used before e.g. by Schäl (2005) and Kirch and Runggaldier (2005). A different pure jump market model can be found in Norberg (2003). In Section 3 we show how the optimization problem for the piecewise deterministic Markov process can be solved by a discrete-time Markov Decision Process. This method is rather standard and is e.g. explained in Davis (1993) or Almudevar (2001) for general PDMP. However in finance it has only been used by Kirch and Runggaldier (2005) for an efficient hedging problem with a very special stock price process. Section 4 uses the MDP formulation to show that the value function is the unique fixed point of a dynamic programming (DP) operator and that the *value iteration* as well as the *policy iteration* hold. Moreover, we prove the existence of an optimal stationary

portfolio strategy. Surprisingly, though the utility function is unbounded (in contrast to Kirch and Runggaldier (2005)) it is possible to introduce a certain weighted supremum norm on the function space such that the DP operator is contracting. This observation is crucial for most of the results which follow. We like to note here that, though existence results for MDPs are well-established (see e.g. Bertsekas and Shreve (1978)), it is rather profound to use the right topology on the action space in order to obtain the desired result. In this section also some model extensions are discussed like regime switching of model parameters and additional risk constraints. In Section 5 we present some computational methods and approximation procedures which arise from the MDP formulation and which are an interesting addition to the traditional HJB approach. More precisely, we show that *Howard's policy improvement algorithm* is valid. Howard's algorithm allows to find (if not run until termination which yields the optimal policy) a good policy after a reasonable number of steps. Then we use the properties of the DP operator to approximate the value function by approximating the utility function of the problem. In this case, also the optimal portfolio strategy is approximated. This complements results of Jouini and Napp (2004) who have investigated this question in diffusion markets. It turns out that the use of the DP operator simplifies the analysis considerably. Finally, we give an error bound on the distance to the true value function when we use the iterates of a DP operator which is concentrated on a grid of the state space only. A numerical example is presented in Section 6 and our approach is compared to the approximating Markov chain method.

2 The Model

We consider a special class of continuous-time financial markets where asset dynamics follow a so-called *piecewise deterministic Markov process* (PDMP). The reason is twofold: First it can be shown that general Lévy processes can be approximated arbitrarily close by such type of processes and second, optimization problems arising from these models can essentially be solved via the theory of discrete-time Markov Decision Processes (MDP).

The financial market is as follows: Suppose $T > 0$ is the time horizon of our model and we have a probability space $(\Omega, \mathcal{F}, P)$ and on this space there is a homogeneous càdlàg Poisson process $N = (N_t)$ with constant rate $\lambda > 0$ and a sequence of independent and identically distributed random vectors (Y_n) with values in $(-1, \infty)^d$. The $Y_n's$ are assumed to have a distribution Q and are itself independent of (N_t). Thus, we can define the $\mathbb{R}^d$-valued compound Poisson process

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

Let us denote by $(\mathcal{F}_t)$ the filtration generated by $C = (C_t)$. By (C_t^k) we denote the k-the 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

$$S_t^0 := e^{rt},$$

2

where $r \geq 0$ denotes the fixed 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 price is given by $S_0^k > 0$.

In this financial market, the price processes are so-called *piecewise deterministic Markov processes* (PDMP), i.e. they show a deterministic evolution between jumps and the jumps occur at random time points according to a Poisson process and have a random height. If we denote 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-the jump. Since $Y_n^k > -1$ a.s. our asset prices stay positive. At first glance it seems to be quite special that the asset prices can only have joint jumps at the time points $T_0 < T_1 < T_2 < \ldots$. However, note that the distribution Q of 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. In what follows we denote $S_t := (S_t^1, \ldots, S_t^d)$.

Now we want to invest into this financial market. To ensure that the wealth process stays positive we do not allow for short-sellings. Thus, the portfolio strategy can be given in terms of fractions of invested capital. In what follows, an admissible portfolio strategy is given by an $(\mathcal{F}_t)$-predictable stochastic process (π_t) 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 $\pi_t = (\pi_t^1, \ldots, \pi^d)$ gives the fractions of wealth invested in the stocks at time t. $1 - \pi_t \cdot e$ is the fraction invested in the bond, thus the portfolio is self-financing. Let us denote by Π the set of all admissible portfolio strategies. The equation for the wealth process (X_t^π) is then:

$$dX_t^\pi = X_{t-}^\pi \Big(rdt + \pi_t \cdot (\mu - re)dt + \pi_t dC_t \Big). \tag{2.1}$$

The wealth process is again a PDMP. We obtain the following explicit expression for the wealth process:

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

The aim of our investor is now to maximize her expected utility from terminal wealth. Thus, we denote by $U : (0, \infty) \to \mathbb{R}_+$ an increasing, concave utility function and define for $\pi \in \Pi$, $t \in [0, T], x > 0$:

$$V_\pi(t, x) := \mathbb{E}_{t,x} U(X_T^\pi) \tag{2.2}$$

3

the expected utility when we start at time t with wealth x and use the portfolio strategy π. As usual $\mathbb{E}_{t,x}$ is the expectation w.r.t. the conditional probability measure $\mathbb{P}_{t,x} = \mathbb{P}(\cdot \mid X_t = x)$. The maximal utility when we start at time t with wealth x is given by

$$V(t,x) := \sup_{\pi \in \Pi} V_\pi(t,x). \tag{2.3}$$

Obviously we have

$$V_\pi(T,x) := U(x) = V(T,x). \tag{2.4}$$

Throughout we assume that

$$\int \|y\| Q(dy) < \infty$$

where for $x \in \mathbb{R}^d$, $\|x\| = |x_1| + \ldots + |x_d|$. This implies that V is well-defined. In Section 4 we will discuss some extensions of this model.

3   Solution by a Discrete-Time MDP

In principle the preceding optimization problem is a continuous-time problem. However, we can show that the value function V can be obtained as the value function of a suitably defined discrete-time Markov Decision Process which is obtained by looking at the embedded jump process. The advantage of this fact is that we can identify V as the unique fixed point of the (contracting) dynamic programming operator and the optimal portfolio strategy as its maximizer without any differentiability assumptions. Moreover, several different computational approaches for V arise. We consider now the following *discrete-time Markov Decision Process*:

- The state space is $E = [0,T] \times (0,\infty)$ endowed with the Borel σ-algebra $\mathfrak{B}(E)$. A state (t,x) represents the jump time point t and the wealth x of the process directly after the jump. The embedded Markov chain which is the state process of the MDP is denoted by $Z_n = (T_n, X_{T_n})$. Since we are only interested in the state process up to time T, we fix some external state $\Delta \notin E$ (cemetery state) and set $Z_n := \Delta$ whenever $T_n > T$.

- The action space of the MDP is given by

$$A := \{\alpha : [0,T] \to \mathcal{U} \text{ measurable}\} \tag{3.1}$$

where $\alpha(t) = \alpha_t \in \mathcal{U}$ denotes the fractions of wealth invested in the stocks t time units after the last jump. For $\alpha \in A$ we define the movement of the wealth between jumps by

$$\phi_t^\alpha(x) := x \exp\left(\int_0^t r + \alpha_s \cdot (\mu - re) ds\right). \tag{3.2}$$

$\phi_t^\alpha(x)$ gives the wealth t time units after the last jump, when the state directly after the jump was x.

4

- For $(t,x) \in E$, $\alpha \in A$ and $B \in \mathfrak{B}(E)$ the transition probability q is given by

$$q(B \mid t, x, \alpha) := \lambda \int_0^{T-t} e^{-\lambda s} \int 1_B\Big(t+s, \phi_s^\alpha(x)\big(1 + \alpha_s \cdot y\big)\Big) Q(dy)ds$$

$$q(\{\Delta\} \mid t, x, \alpha) := e^{-\lambda(T-t)} = 1 - q(E \mid t, x, \alpha)$$

$$q(\{\Delta\} \mid \{\Delta\}, \alpha) := 1.$$

- The one-stage reward function is defined for $(t,x) \in E$ and $\alpha \in A$ by

$$r(t, x, \alpha) := e^{-\lambda(T-t)} U\big(\phi_{T-t}^\alpha(x)\big).$$

$$r(\{\Delta\}, \alpha) := 0.$$

A measurable function $f : E \cup \{\Delta\} \to A$ (with $f(\Delta) = \alpha_0 \in A$ arbitrary) is called *decision rule*. The decision rule determines the action which has to be applied until the next jump as a function of the state Z_n directly after a jump. By F we denote the set of all decision rules. A sequence of decision rules (f_n), with $f_n \in F$ is called *Markov policy*. The expected reward of such a Markov policy is given by

$$J_{(f_n)}(t, x) := \mathbb{E}_{t,x}^{(f_n)} \left[\sum_{k=0}^\infty r\Big(Z_k, f_k(Z_k)\Big) \right], \quad (t, x) \in E.$$

As usual in MDP theory we denote by $\mathbb{E}_{t,x}^{(f_n)}$ the expectation w.r.t. the probability measure $\mathbb{P}_{t,x}^{(f_n)}$ induced by (f_n). If $f_n \equiv f$ for all $n \in \mathbb{N}$, i.e. the policy is *stationary*, then we write J_f. The value function is defined by

$$J(t, x) = \sup_{(f_n)} J_{(f_n)}(t, x), \quad (t, x) \in E.$$

If we define for a sequence (f_n) with $f_n \in F$

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

then (π_t) is called a *Markovian portfolio strategy*. We are now able to state the main result of this section. The continuous-time optimization problem can be solved by the above discrete-time MDP. We define for $\pi \in \Pi$: $Z_n^\pi := (T_n, X_{T_n}^\pi)$ and $Z_n^\pi = \Delta$ whenever, $T_n > T$.

Theorem 3.1.

a) *For any Markovian portfolio strategy (π_t) we have $V_\pi(t, x) = J_{(f_n)}(t, x)$, $(t, x) \in E$.*

b) *It holds: $V(t, x) = J(t, x)$, $(t, x) \in E$.*

Proof.   a) For a Markovian portfolio strategy $\pi \in \Pi$ we get

$$X_s^\pi = \phi_{s-T_k}^{f_k(Z_k^\pi)}(X_{T_k}^\pi) \text{ for } s \in [T_k, T_{k+1})$$

5

and $Z_k = Z_k^\pi$ $\mathbb{P}_{t,x}^{(f_n)}-a.s..$ Next denote by ε_T the one-point measure in T. Then we obtain with monotone convergence for $(t,x) \in E$

$$
\begin{aligned}
V_\pi(t,x) &= \mathbb{E}_{t,x}\left[\int_t^\infty U(X_s^\pi)\varepsilon_T(ds)\right] = \sum_{k=0}^\infty \mathbb{E}_{t,x}^{(f_n)}\left[\int_{T_k}^{T_{k+1}} U\left(\phi_{s-T_k}^{f_k(Z_k)}(X_{T_k})\right)\varepsilon_T(ds)\right] \\
&= \sum_{k=0}^\infty \mathbb{E}_{t,x}^{(f_n)}\left[\mathbb{E}_{t,x}^{(f_n)}\left[\int_{T_k}^{T_{k+1}} U\left(\phi_{s-T_k}^{f_k(Z_k)}(X_{T_k})\right)\varepsilon_T(ds) \,\Big|\, Z_k\right]\right] \\
&= \mathbb{E}_{t,x}^{(f_n)}\left[\sum_{k=0}^\infty r\left(Z_k, f_k(Z_k)\right)\right]
\end{aligned}
$$

where we use

$$
\mathbb{E}_{t,x}^{(f_n)}\left[\int_{T_k}^{T_{k+1}} U\left(\phi_{s-T_k}^{f_k(Z_k)}(X_{T_k})\right)\varepsilon_T(ds) \,\Big|\, Z_k\right] = r\left(Z_k, f_k(Z_k)\right)
$$

and the transition kernel is given by

$$
\mathbb{P}_{t,x}^{(f_n)}\left(Z_{k+1} \in B \mid Z_k\right) = q\left(B \mid Z_k, f_k(Z_k)\right)
$$

for $B \in \mathfrak{B}(E)$.

b) For any $\pi \in \Pi$ there exists a sequence (f_n) of measurable functions $f_n : E^{n+1} \to A$ with $\pi_t = f_n(Z_0^\pi, \ldots, Z_n^\pi)(t - T_n)$ for $t \in [T_n, T_{n+1})$. Because of the Markovian structure of the state process we obtain (see Bertsekas and Shreve (1978), p.216)

$$
V(t,x) = \sup_{\pi \text{ Markovian}} V_\pi(t,x) = J(t,x), \quad (t,x) \in E
$$

where the last equation follows from part a).

$\square$

Thus, we have reduced the problem to a discrete-time MDP. Let us now define

$$
\mathbb{M} := \{v : E \to \mathbb{R}_+ \mid v \text{ is measurable}\}.
$$

Of particular interest in this model are the following operators $L, \mathcal{T}_f, \mathcal{T}$ which act on $\mathbb{M}$:

$$
\begin{aligned}
Lv(t,x,\alpha) &:= e^{-\lambda(T-t)}U\left(\phi_{T-t}^\alpha(x)\right) + \int v(s,y)q(ds,dy \mid t,x,\alpha). \\
\mathcal{T}_f v(t,x) &:= Lv(t,x,f(t,x)), \ (t,x) \in E. \\
\mathcal{T}v(t,x) &= \sup_{\alpha \in A} Lv(t,x,\alpha).
\end{aligned}
$$

From MDP theory (see Bertsekas and Shreve (1978), chapter 9), it follows that

$$
\begin{aligned}
J_{(f_n)} &= \lim_{n\to\infty} \mathcal{T}_{f_0}\ldots\mathcal{T}_{f_{n-1}}0 \\
J_f &= \lim_{n\to\infty} \mathcal{T}_f^n 0 \\
J_f &= \mathcal{T}_f J_f.
\end{aligned}
$$

It will be shown in the next section that J_f is the unique fixed point of $\mathcal{T}_f$ in $\mathbb{B}_b$ which is defined below.

4 Existence of Optimal Strategies and Characterization of the Value Function

The key observation to the solution of the MDP is the fact that it is possible to define a peculiar norm on a subset of the function space $I\!M$ such that the operator $\mathcal{T}$ is contracting and has a unique fixed point and the fixed point is equal to the value function $V(t,x)$. The norm is a special weighted supremum norm with a suitably defined upper bounding function.

4.1 A Peculiar Norm

In order to obtain this norm we first introduce an upper bounding function:

Proposition 4.1. *Let $b(t,x) := e^{\beta(T-t)}(1+x)$, $(t,x) \in E$ for $\beta \geq 0$. Then there exist constants $c, c_\beta \in \mathbb{R}_+$, such that*

(i) *$r(t,x,\alpha) \leq cb(t,x)$ for all $(t,x,\alpha) \in E \times A$.*

(ii) *For all $(t,x,\alpha) \in E \times A$ it holds that*

$$\int b(s,y)q(ds,dy|t,x,\alpha) \leq c_\beta b(t,x)$$

where

$$c_\beta := \frac{\lambda(1+\bar{y})}{\beta+\lambda-\bar{\mu}}\left(1 - e^{-T(\beta+\lambda-\bar{\mu})}\right) \tag{4.1}$$

and $\bar{\mu} := \max\{\mu_1, \ldots, \mu_d, r\}$ is the maximal appreciation rate of the stocks and $\bar{y} := \max\{\int y_1 Q(dy), \ldots \int y_d Q(dy)\}$ is the maximal expected relative jump height. Such a function is called upper bounding function *for the MDP.*

Proof. Note that since $\mathcal{U}$ is compact, there exists a constant $\hat{c}$ such that for all $\alpha \in A$ and $s > 0$ it holds: $\alpha_s \cdot (\mu - re) \leq \hat{c}$. Moreover, the concave utility function U can be bounded from above by a linear function $c_1(1+x)$ for some $c_1 > 0$. Thus we obtain

$$\begin{aligned}
r(t,x,\alpha) &= e^{-\lambda(T-t)}U(\phi^\alpha_{T-t}(x)) \leq e^{-\lambda(T-t)}U(xe^{T(r+\hat{c})}) \\
&\leq e^{-\lambda(T-t)}c_1(1+xe^{T(r+\hat{c})}) \leq cb(t,x)
\end{aligned}$$

for c large enough. Condition (ii) follows since

$$\sup_{(t,x,\alpha)} \frac{\int b(s,y)q(ds,dy|t,x,\alpha)}{b(t,x)} =$$

$$= \sup_{(t,x,\alpha)} \frac{\lambda \int_0^{T-t} e^{-\lambda s + \beta(T-s-t)}\left(1 + \phi^\alpha_s(x)\left(1 + \alpha_s \cdot \int y\, Q(dy)\right)\right)ds}{e^{\beta(T-t)}(1+x)}$$

$$\leq \sup_{(t,x)\in E} \frac{\lambda \int_0^{T-t} e^{-s(\lambda+\beta)}\left(1 + xe^{\bar{\mu}s}(1+\bar{y})\right)ds}{1+x} = \frac{\lambda(1+\bar{y})}{\beta+\lambda-\bar{\mu}}\left(1 - e^{-T(\beta+\lambda-\bar{\mu})}\right) = c_\beta.$$

$\square$

Remark 4.2. The following observation is crucial: For β large we obviously get $c_\beta < 1$. In Proposition 4.3 we will see that c_β is the contraction module of the operators $\mathcal{T}$ and $\mathcal{T}_f$ for $f \in F$.

Next we introduce the weighted supremum norm $\|\cdot\|_b$ on $I\!\!M$ by

$$\|v\|_b := \sup_{(t,x)\in E} \frac{v(t,x)}{b(t,x)}$$

and define

$$I\!\!B_b := \{v \in I\!\!M \mid \|v\|_b < \infty\}.$$

Note that convergence in $\|\cdot\|_b$ is equivalent to uniform convergence on compact sets. In the remaining part of the paper we assume that β is large enough to ensure $c_\beta < 1$.

4.2 Main Results

At first we define the set

$$I\!\!M_c := \{v \in I\!\!B_b \mid v \text{ is continuous, } v(t,x) \text{ is concave and increasing in } x, \text{ decreasing in } t\}.$$

In order to prove existence of optimal policies, we have to show certain continuity and compactness conditions. To achieve this, we have to enlarge the action space by so-called randomized controls (this is a standard procedure for deterministic control problems). Thus, we define by

$$\mathcal{R} := \{\alpha : [0,T] \to P(\mathcal{U}) \text{ measurable}\} \tag{4.2}$$

the set of *relaxed controls* where $P(\mathcal{U})$ is the set of all probability measures on $\mathcal{U}$ equipped with the σ-algebra of the Borel sets. 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*. The Young topology on $\mathcal{R}$ is the coarsest topology such that all mappings of the form

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

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

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

We denote this class by $L^1([0,T], C(\mathcal{U}))$. W.r.t. the Young-topology, $\mathcal{R}$ is separable, metric and also compact, thus a Borel space. 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

$$\lim_{n\to\infty} \int_0^T \int_{\mathcal{U}} \psi(t,u)\alpha_t^n(du)dt = \int_0^T \int_{\mathcal{U}} \psi(t,u)\alpha_t(du)dt \tag{4.3}$$

8

for all $\psi \in L^1([0,T], C(\mathcal{U}))$. For $\alpha \in \mathcal{R}$ we define $\bar{\alpha}_s := \int_{\mathcal{U}} u\alpha_s(du)$ which is in A and $\phi_t^\alpha(x) := \phi_t^{\bar\alpha}(x)$. The transition kernel q of the MDP is extended for $\alpha \in \mathcal{R}$ by

$$\int v(s,y) q(ds, dy \mid t, x, \alpha) = \lambda \int_0^{T-t} e^{-\lambda s} \int \int v\Big(t+s, \phi_s^\alpha(x)(1 + u \cdot y)\Big) \alpha_s(du) Q(dy) ds$$

for all measurable $v : E \to \mathbb{R}$ where the integral exists. The operator $\mathcal{T}$ has now the following interesting properties:

Proposition 4.3. *It holds that*

a) *For $v \in \mathbb{M}_c$ we have: $Tv(t,x) = \sup_{\alpha \in \mathcal{R}} Lv(t,x,\alpha)$, $(t,x) \in E$.*

b) *$\mathcal{T} : \mathbb{M}_c \to \mathbb{M}_c$.*

c) *For $v, w \in \mathbb{B}_b$ and $f \in F$ we have*

$$\begin{aligned}
\|\mathcal{T}_f v - \mathcal{T}_f w\|_b &\le c_\beta \|v - w\|_b \\
\|\mathcal{T} v - \mathcal{T} w\|_b &\le c_\beta \|v - w\|_b.
\end{aligned}$$

Thus the operators $\mathcal{T}_f, \mathcal{T}$ are contracting if $c_\beta < 1$.

Proof.   a) Suppose $v \in \mathbb{M}_c$ and $\alpha \in \mathcal{R}$. We obtain with Jensen's inequality

$$\int v\Big(t+s, \phi_s^\alpha(x)(1 + u \cdot y)\Big) \alpha_s(du) \le v\Big(t+s, \phi_s^{\bar\alpha}(x)(1 + \bar\alpha_s \cdot y)\Big)$$

and thus $Lv(t,x,\alpha) \le Lv(t,x,\bar\alpha)$ which implies part a).

b) Suppose $v \in \mathbb{M}_c$ and let us consider $\mathcal{T}v$. Obviously, $\mathcal{T}v(t,x)$ is non-negative, increasing in x and decreasing in t. This can be seen directly from the definition of $\mathcal{T}$. $\mathcal{T}v \in \mathbb{B}_b$ follows from Proposition 4.1. Let us next prove that $x \mapsto \mathcal{T}v(t,x)$ is concave. So far we have worked with portfolios in terms of fractions of invested wealth. Since in our model it is guaranteed that the wealth process never falls to zero (given $x_0 > 0$) we can equivalently work with invested amounts a_t. More precisely, the fraction $\alpha \in A$ gives the same wealth as the amount $a_t := \alpha_t \phi_t^\alpha(x)$. Under a, the deterministic evolution between jumps is given by

$$\phi_t^a(x) := e^{rt}\left(x + \int_0^t e^{-rs} a_s \cdot (\mu - re) ds\right).$$

The advantage is now that $(x, a) \mapsto \phi_t^a(x)$ is linear. We show first that $(x, a) \mapsto Lv(t,x,a)$ is concave on $E \times A$: Fix $t \in [0,T]$, wealths $x_1, x_2 > 0$, controls a_1, a_2 and $\kappa \in (0,1)$. Let $\hat{x} := \kappa x_1 + (1 - \kappa) x_2$ and $\hat{a} := \kappa a_1 + (1 - \kappa) a_2$. Note that $\hat{a}$ is again admissible. For the next equation note that

$$\phi_s^{\hat{a}}(\hat{x}) = \kappa \phi_s^{a_1}(x_1) + (1 - \kappa) \phi_s^{a_2}(x_2).$$

9

Then we obtain

$$
\begin{aligned}
Lv(t, \hat{x}, \hat{a}) \;&=\; e^{-\lambda(T-t)} U\big(\phi^{\hat{a}}_{T-t}(\hat{x})\big) + \lambda \int_0^{T-t} e^{-\lambda s} \int v\Big(t+s, \phi^{\hat{a}}_s(\hat{x}) + \hat{a}_s \cdot y\Big) Q(dy) ds \\
&\geq\; \kappa Lv(t, x_1, a_1) + (1-\kappa) Lv(t, x_2, a_2).
\end{aligned}
$$

Now choose a_1, a_2 such that for $\varepsilon > 0$:

$$
Lv(t, x_1, a_1) \geq \mathcal{T} v(t, x_1) - \varepsilon \quad \text{and} \quad Lv(t, x_2, a_2) \geq \mathcal{T} v(t, x_2) - \varepsilon.
$$

Then

$$
\begin{aligned}
\mathcal{T} v(t, \hat{x}) \;&\geq\; Lv\big(t, \hat{x}, \hat{a}\big) \\
&\geq\; \kappa Lv(t, x_1, a_1) + (1-\kappa) Lv(t, x_2, a_2) \\
&\geq\; \kappa \mathcal{T} v(t, x_1) + (1-\kappa) \mathcal{T} v(t, x_2) - \varepsilon.
\end{aligned}
$$

Letting $\varepsilon \downarrow 0$ the statement follows.

Next we show that $(t, x) \mapsto \mathcal{T} v(t, x)$ is continuous for $v \in \mathbb{M}_c$. This follows when we can show that the mappings

$$
E \times \mathcal{R} \ni (t, x, \alpha) \;\mapsto\; r(t, x, \alpha) \tag{4.4}
$$

$$
E \times \mathcal{R} \ni (t, x, \alpha) \;\mapsto\; \int v(s, y) q(ds, dy | t, x, \alpha) \tag{4.5}
$$

are both continuous for $v \in \mathbb{M}_c$. Note that we use the relaxed controls here and that $\mathcal{R}$ is compact w.r.t. the Young topology. First we show that the mapping $(t, x, \alpha) \mapsto \phi^\alpha_t(x)$ is continuous. To this end, let $(t_n, x_n) \subset E$ and $(\alpha_n) \subset \mathcal{R}$ be sequences with $(t_n, x_n) \to (t, x) \in E$ and $\alpha_n \to \alpha \in \mathcal{R}$. Since

$$
\phi^\alpha_t(x) = x e^{rt} \exp \Big(\int_0^t \int_{\mathcal{U}} u \cdot (\mu - re) \alpha_s(du) ds \Big)
$$

it suffices to show the continuity of the third factor. Since the exponential-function is continuous, the problem can be reduced to showing the continuity of the integral expression. To ease notation we define

$$
\mu^n_s \;:=\; \int_{\mathcal{U}} u \cdot (\mu - re) \alpha^n_s(du), \quad \mu_s := \int_{\mathcal{U}} u \cdot (\mu - re) \alpha_s(du).
$$

Thus we look at

$$
\left| \int_0^{t_n} \mu^n_s ds - \int_0^t \mu_s ds \right| \leq \left| \int_0^{t_n} \mu^n_s ds - \int_0^t \mu^n_s ds \right| + \left| \int_0^t \mu^n_s ds - \int_0^t \mu_s ds \right|
$$

Since $u \mapsto \psi(s, u) = u \cdot (\mu - re)$ for $0 \leq s \leq t$ and $u \mapsto \psi(s, u) = 0$ for $t \leq s \leq T$ are continuous, it follows from (4.3) that the second term tends to 0 for $n \to \infty$.

10

Obviously, the first term is bounded by $\hat{c}|t-t_n|$ which also tends to 0 and the continuity of $(t,x,\alpha) \mapsto \phi_t^\alpha(x)$ is shown.

Since U is concave on $(0,\infty)$, it is also continuous and it follows from the continuity of $\phi_t^\alpha(x)$ that the mapping (4.4) is continuous.

The continuity of the mapping (4.5) is first shown for bounded and continuous functions v. This can be done very similar as for (4.4). Next let $v \in \mathbb{M}_c$ be arbitrary. Since the sequence (v_n) defined by $v_n := v \wedge n$ is bounded and continuous and $v_n \uparrow v$ it follows with monotone convergence and with Lemma 7.14 in Bertsekas and Shreve (1978) that the mapping (4.5) is lower semicontinuous. On the other hand $v \in \mathbb{M}_c$ implies the existence of a constant $\tilde{c}$ such that $v - \tilde{c}b \le 0$ and $(t,x) \mapsto v(t,x) - \tilde{c}b(t,x)$ is continuous. Hence there exists a sequence (v_n) of bounded and continuous functions with $v_n \downarrow v - \tilde{c}b$ (see again Lemma 7.14 in Bertsekas and Shreve (1978)). Noting that the mapping (4.5) with v replaced by b is continuous and using monotone convergence it follows that (4.5) is upper semicontinuous and thus continuous.

c) This part is rather standard: For $v, w \in \mathbb{B}_b$ and $f \in F$ we obtain

$$\mathcal{T}_f v(t,x) - \mathcal{T}_f w(t,x) \;\le\; \sup_{\alpha \in A} \int \Big(v(s,y) - w(s,y) \Big) q(ds, dy | t, x, \alpha)$$

$$\le\; \|v - w\|_b \sup_{\alpha \in A} \int b(s,y) q(ds, dy | t, x, \alpha).$$

Interchanging v and w and completing the norm on the left-hand-side yields the result for $\mathcal{T}_f$. Obviously the preceding inequality holds also for $\mathcal{T}$ which implies the second statement.

$\square$

Proposition 4.3 c) implies that J_f is the unique fixed point of $\mathcal{T}_f$ in $\mathbb{B}_b$. Next we define the limit $J_\infty := \lim_{n \to \infty} \mathcal{T}^n 0$ which exists since the sequence $(\mathcal{T}^n 0)$ is increasing. The limit function J_∞ has the following properties:

Proposition 4.4. *It holds that*

a) *$J_\infty \in \mathbb{M}_c$ and $J_\infty = \mathcal{T} J_\infty$.*

b) *For all $v \in \mathbb{M}_c$ there exists an $f \in F$ with $\mathcal{T} v = \mathcal{T}_f v$.*

c) *$J_\infty = \sup_f J_f = J$.*

Proof. a) Let us first show that $J_\infty \in \mathbb{M}_c$. Obviously, $0 \in \mathbb{M}_c$. From Proposition 4.3 it follows that $J_n := \mathcal{T}^n 0 \in \mathbb{M}_c$ for all $n \in \mathbb{N}$. Thus, J_∞ is non-negative, concave and increasing in x and decreasing in t. This follows directly by taking the limit. Moreover, due to Proposition 4.3 we know that J_n converges uniformly in the weighted supremum norm to J_∞ which implies that the continuity of J_n carries over to J_∞. The statement $J_\infty = \mathcal{T} J_\infty$ follows from Banach's fixed point Theorem. Thus part a) is shown.

b) It follows from a classical measurable selection result (see e.g. Bertsekas and Shreve (1978), chapter 7.5) that there exists a randomized maximizer $f : E \to \mathcal{R}$ for each $v \in \mathbb{M}_c$. In view of the proof of Proposition 4.3 a) we also get a deterministic maximizer $\bar{f}$. Note that $\bar{f}$ satisfies the required measurability conditions.

c) Note that we always have $J \leq J_\infty$. On the other hand, we know by part a) and b) that there exists a maximizer $f \in F$ such that $J_\infty = \mathcal{T}_f J_\infty$. By Proposition 4.3 c) we get $J_\infty = J_f \leq J$. Thus, equality holds.

$\square$

The value function of our portfolio problem can now be characterized as the unique fixed point of the operator $\mathcal{T}$:

Theorem 4.5. *The following statements hold for our portfolio problem:*

a) *The value function V of the problem satisfies for all $(t, x) \in E$:*

$$V(t, x) = J(t, x)$$

and V is the unique fixed point of $\mathcal{T}$ in $\mathbb{M}_c$.

b) *It holds for $J_0 \in \mathbb{M}_c$ that*

$$\|V - \mathcal{T}^n J_0\|_b \leq \frac{c_\beta^n}{1 - c_\beta} \|\mathcal{T} J_0 - J_0\|_b.$$

c) *There exists an optimal stationary portfolio strategy $\pi \in \Pi$, i.e. there exists $f \in F$ such that*

$$\pi_t = f(Z_n^\pi)(t - T_n) \text{ for } t \in [T_n, T_{n+1}).$$

Proof. Part a) follows from Proposition 4.4 c), Theorem 3.1 and Banach's fixed point Theorem. Part b) is again an application of Banach's fixed point Theorem. c) follows from Proposition 4.4 c).

$\square$

Remark 4.6. A standard result form MDP theory implies that the *policy iteration* is also valid, i.e. an optimal policy can be obtained as a limit of maximizers of $\mathcal{T}^n J_0$ (see Bertsekas and Shreve (1978), chapter 9.5).

4.3   Model Extensions

In order to outline the MDP approach we have used a fairly simple model. This model can be extended in various ways without destroying the results in Section 4.2.

General Piecewise Deterministic Markov Processes
PDMP are more general than the model we have used. In particular we could replace the constant jump intensity λ by a state dependent intensity $\lambda(X_t)$. Moreover, the relative jump

height distribution Q may also depended on the state $Q(X_t; \cdot)$. In this case we need further continuity and regularity assumption on λ and Q to establish the main results of Section 4.2.

Regime-switching

Another interesting point is that we can introduce randomly varying coefficients (sometimes also called *regime switching* (see e.g. Bäuerle and Rieder (2004)): Suppose $I = (I_t)$ is a continuous-time Markov chain with finite state space S. We suppose that (I_t) is given in uniformized form, i.e. we have a constant jump intensity $\tilde{\lambda}$ and embedded transition probabilities $p_{ij}, i, j \in S$ with possibly $p_{ii} > 0$. Moreover, assume that the interest rate, the appreciation rates, the jump intensity and the jump distribution of the stocks depend on (I_t), i.e. the wealth equation for $\pi \in \Pi$ is given by

$$dX_t^\pi = X_t^\pi \Big(r(I_t) + \pi_t \cdot (\mu(I_t) - r(I_t)e)dt + \pi_t dC_t \Big) \tag{4.6}$$

with $C_t = \sum_{k=1}^{N_t} Y_k$ and (N_t) is a Cox-process with (stochastic) jump intensity $\lambda(I_t)$ at time t and also the distribution of Y_k might depend on the regime at the jump time. In this case the state space is given by $E = [0, T] \times (0, \infty) \times S$. All results of the previous section can be obtained for this model in an analogous way.

Partial Information

Problems with partial information for the investor can also be dealt with in this approach. For example in the previous extension, the Markov chain (I_t) might be "hidden". Using a result from filter theory it is again possible to reduce this problem to an MDP with now enlarged state space. Note that the filter process is again a PDMP. This has been done e.g. in Kirch and Runggaldier (2005), Bäuerle and Rieder (2007).

Dynamic Risk Constraints

We can also include dynamic risk constraints. This extension can be incorporated into the MDP framework without problems. The set $\mathcal{U}$ of admissible fractions is now constrained by considering the risk of the fractions. Suppose ρ is an arbitrary risk measure (e.g. Value-at-Risk, Average-Value-at-Risk) and we are at time point T_n with a wealth x, i.e. $X_{T_n}^\pi = x$. Next suppose we choose $u \in \mathcal{U}$ and fix this fraction until the next jump, i.e. $\pi_t \equiv u$ for all $t \in [T_n, T_{n+1})$. The admissible fractions $\mathcal{U}_c$ for the time interval $[T_n, T_{n+1})$ are then determined by the constraint

$$\mathcal{U}_c := \left\{ u \in \mathcal{U} \mid \rho\left(\frac{X_{T_{n+1}}^\pi - x}{x} \right) \leq b \right\}$$

which bounds the risk of the relative wealth difference after the next jump. Thus, the MDP is exactly the same as before with $\mathcal{U}$ replaced by $\mathcal{U}_c$.

5 Computational Methods

In what follows we want to solve the fixed point equation and thus find the value function, at least in some special cases or approximately. To this end, let us introduce the notation

$$\ell_v(t, x, u) := \lambda \int v\Big(t, x(1 + u \cdot y)\Big) Q(dy)$$

where $v \in I\!M_c$. Then we can write the fixed point equation in the following form:

$$V(t, x) = \sup_{\alpha \in A} \left\{ e^{-\lambda(T-t)} U(\phi_{T-t}^\alpha(x)) + \int_0^{T-t} e^{-\lambda s} \ell_V(t + s, \phi_s^\alpha(x), \alpha_s) ds \right\}. \tag{5.1}$$

This is a special deterministic control problem. The corresponding HJB equation is given by

$$\begin{aligned}
\lambda v(t, x) &= \sup_{u \in \mathcal{U}} \left\{ v_t + v_x x \big(r + u \cdot (\mu - re) \big) + \ell_v(t, x, u) \right\} \tag{5.2} \\
v(T, x) &= U(x).
\end{aligned}$$

In the case of an investor with power utility $U(x) = \frac{1}{\gamma} x^\gamma$, $0 < \gamma < 1$ it is well-known that the problem can be solved explicitly (see e.g. Øksendal and Sulem (2005) Example 3.2) and the solution is given by

$$V(t, x) = \frac{1}{\gamma} x^\gamma e^{\theta(T-t)}$$

where $\theta := \gamma r - \lambda + u^* \gamma (\mu - re) + \lambda \int (1 + u^* y)^\gamma Q(dy)$ and u^* is the solution of

$$\begin{cases} u\gamma(\mu - re) + \lambda \int (1 + uy)^\gamma Q(dy) \to \max \\ u \in \mathcal{U} \end{cases}$$

The optimal policy is to invest the constant fraction u^* of the wealth in the stock. However, the problem has to be solved numerically in general and some approaches are discussed in the next sections.

5.1 Howard's Policy Improvement Algorithm

Starting with an arbitrary portfolio strategy, Howard's policy improvement algorithm computes a new one which attains a higher value or, if no improvement is possible, returns the optimal portfolio strategy and value. In order to prove that this algorithm works here, let us introduce the following set. For a decision rule $f \in F$ denote

$$D(t, x, f) := \{\alpha \in A \mid LJ_f(t, x, \alpha) > J_f(t, x)\}$$

for $(t, x) \in E$. We obtain the following result:

Theorem 5.1. *Howard's policy improvement algorithm works for the portfolio problem. Let $f, g \in F$.*

a) If for some subset $E_0 \subset E$

$$g(t, x) \in D(t, x, f), \qquad (t, x) \in E_0$$
$$g(t, x) = f(t, x), \qquad (t, x) \notin E_0$$

then $J_g \geq J_f$ and $J_g(t, x) > J_f(t, x)$ for $(t, x) \in E_0$. In this case the decision rule g is called an improvement of f.

b) If $D(t, x, f) = \emptyset$ for all $(t, x) \in E$ then $J_f = V$, i.e. the decision rule f is optimal. More precisely, the stationary portfolio strategy $\pi \in \Pi$ defined by

$$\pi_t = f(Z_n^\pi)(t - T_n), \ t \in [T_n, T_{n+1})$$

is optimal for the portfolio problem.

Proof.   a) From the construction we get

$$\mathcal{T}_g J_f(t, x) > J_f(t, x)$$

if $(t, x) \in E$ and $\mathcal{T}_g J_f(t, x) = J_f(t, x)$ if $(t, x) \notin E_0$. Thus, we obtain by induction:

$$J_f \leq \mathcal{T}_g^n J_f$$

where the inequality is strict if we plug in $(t, x) \in E_0$. Letting $n \to \infty$ it follows that $J_f \leq J_g$ and in particular $J_f(t, x) < J_g(t, x)$ if $(t, x) \in E_0$.

b) $D(t, x, f) = \emptyset$ for all $(t, x) \in E$ means $\mathcal{T} J_f \leq J_f$. Since we always have $\mathcal{T} J_f \geq \mathcal{T}_f J_f = J_f$ we obtain $\mathcal{T} J_f = J_f$. Since the fixed point of $\mathcal{T}$ is unique, the statement follows with Theorem 4.5.

$\square$

Besides the algorithm, Theorem 5.1 can be used to prove the optimality of special portfolio strategies. Let us consider the following example: Assume that $r = \mu_i$ for all $i = 1, \ldots, d$ which means that the deterministic drift of all assets is equal to the drift of the bond. In this case we obtain $\phi_t(x) = xe^{rt}$ and the maximization in the fixed point equation reduces to a pointwise maximization:

$$V(t, x) \ = \ e^{-\lambda(T-t)} U(xe^{r(T-t)}) + \int_0^{T-t} e^{-\lambda s} \sup_{u \in \mathcal{U}} \left\{ \ell_V(t + s, xe^{rs}, u) \right\} ds$$

Using a measurable selection theorem, it can be shown that the maximum points define a decision rule which is the optimal one. Let us now consider the policy "invest all the money in the bond" and name this decision rule $f \equiv 0$. The corresponding value function is then

$$J_f(t, x) = U(xe^{r(T-t)}).$$

We try a policy improvement over this strategy:

$$\sup_{\alpha \in A} LJ_f(t, x, \alpha) = e^{-\lambda(T-t)} U(xe^{r(T-t)})$$

$$+ \lambda \int_0^{T-t} e^{-\lambda s} \sup_{u \in \mathcal{U}} \left\{ \int U\left(xe^{r(T-t+s)}(1 + u \cdot y)\right) Q(dy) ds \right\}$$

The maximizer of this equation is again f, if and only if for all time points $s \in [0, T - t]$ it holds that

$$\sup_{u \in \mathcal{U}} \int U\left(xe^{r(T-t+s)}(1 + u \cdot y)\right) Q(dy) \leq U\left(xe^{r(T-t+s)}\right).$$

Due to the concavity of U this condition is always fulfilled if $\mathbb{E} Y \leq 0$. Now suppose in addition that the utility function is differentiable. Hence we obtain:

Corollary 5.2. *If U is continuously differentiable and $U'(x + u \cdot Y)Y$ is integrable for all $x > 0$ and $\|u\|$ small, then "invest all the money in the bond" is optimal if and only if $\mathbb{E} Y \leq 0$.*

Proof. Since the mapping $u \mapsto \mathbb{E} U\left(xe^{r(T-t+s)}(1 + u \cdot Y)\right) =: g(u)$ is concave, $u \equiv 0$ is a maximum point if and only if $g'(0) \leq 0$. This holds if and only if $\mathbb{E} Y \leq 0$. $\qquad \square$

5.2   Approximation of the Utility Function

Another possibility to approximate the value function, if we cannot compute it exactly, is as follows: Suppose that instead of the utility function U we take $\tilde{U}$ which is somehow close to U and the problem with $\tilde{U}$ is simpler to solve. Then one would expect that also the corresponding value function and optimal policy are close to each other. In a Black-Scholes market this question has been considered in Jouini and Napp (2004), where the L^p-norm has been used. In order to formalized this thought let $U^{(n)}$ be a utility function and denote

$$\mathcal{T}^{(n)} v(t, x) = \sup_{\alpha \in \mathcal{R}} \left\{ e^{-\lambda(T-t)} U^{(n)}\left(\phi_{T-t}^\alpha(x)\right) + \int v(s, y) q(ds, dy \mid t, x, \alpha) \right\}$$

where we replace U by $U^{(n)}$ in $\mathcal{T}$ for $n \in \mathbb{N}$. $L^{(n)}$ is defined in an obvious way. Moreover, denote by $V^{(n)}$ the corresponding value function and by

$$A_n^*(t, x) := \{\alpha \in \mathcal{R} : \mathcal{T}^{(n)} V^{(n)}(t, x) = L^{(n)} V^{(n)}(t, x, \alpha)\}$$

$$A^*(t, x) := \{\alpha \in \mathcal{R} : \mathcal{T} V(t, x) = LV(t, x, \alpha)\}$$

the set of maximum points of $L^{(n)} V^{(n)}$ and LV respectively. Moreover, let us denote by

$$Ls A_n^*(t, x) := \{\alpha \in \mathcal{R} : \alpha \text{ is an accumulation point of } (\alpha_n) \text{ with } \alpha_n \in A_n^*(t, x) \ \forall n \in \mathbb{N}\}$$

the upper limit of the set sequence $\left(A_n^*(t, x)\right)$. Then we obtain:

Theorem 5.3.   *a) If U and $\tilde{U}$ are two utility functions with corresponding value functions V and $\tilde{V}$, then*

$$\|V - \tilde{V}\|_b \le \|U - \tilde{U}\|_b \frac{e^{T\bar{\mu}}}{1 - c_\beta}.$$

b) Let $\left(U^{(n)}\right)$ be a sequence of utility functions with $\lim_{n\to\infty} \|U^{(n)} - U\|_b = 0$. Then $\lim_{n\to\infty} \|V^{(n)} - V\|_b = 0$ and we get that $\emptyset \ne Ls A_n^*(t,x) \subset A^*(t,x)$ for all $(t,x) \in E$, i.e. in particular, the limit f^* of a sequence of decision rules (f_n^*) with $f_n^*(t,x) \in A_n^*(t,x)$ for all $(t,x) \in E$ defines an optimal portfolio strategy for the given model (with utility function U).

Proof.   a) We obtain with Theorem 4.5 (note that $\tilde{T}$ is defined as the operator T where U is replaced by $\tilde{U}$)

$$
\begin{aligned}
\|V - \tilde{V}\|_b &= \|J - \tilde{J}\|_b = \|\mathcal{T}J - \tilde{\mathcal{T}}\tilde{J}\|_b \\
&\le \|U - \tilde{U}\|_b \sup_{(t,x,\alpha)\in E\times A} e^{-\lambda(T-t)} \frac{b(t, \phi_{T-t}^\alpha(x))}{b(t,x)} + c_\beta \|V - \tilde{V}\|_b \\
&\le \|U - \tilde{U}\|_b e^{\bar{\mu}T} + c_\beta \|V - \tilde{V}\|_b.
\end{aligned}
$$

Solving for $\|V - \tilde{V}\|_b$ yields the stated result.

b) This part follows from standard arguments of stochastic dynamic programming (see Bertsekas and Shreve (1978), chapter 9.5).

$\square$

5.3   Value Iteration and State Space Discretization

We assume now that the state space is approximated by a grid $G \subset E$. We will now compute the value function only on G. Thus we define the grid operator T_G on $I\!\!M_c$ by

$$
\mathcal{T}_G v(t,x) := \begin{cases} \mathcal{T}v(t,x), & \text{for } (t,x) \in G \\ \sum_k \lambda_k \mathcal{T}v(t_k, x_k), & \text{for } (t,x) \notin G, (t,x) = \sum_k \lambda_k(t_k, x_k), \\ & \text{and } (t_k, x_k) \in G, \lambda_k \ge 0, \sum_k \lambda_k = 1. \end{cases}
$$

$\mathcal{T}_G$ coincides with $\mathcal{T}$ on G and is a linear interpolation elsewhere. It is easy to see that $T_G : I\!\!M_c \to I\!\!M_c$. Moreover, we define a different bounding function $b_G : E \to \mathbb{R}_+$ by

$$
b_G(t,x) := \begin{cases} b(t,x), & \text{for } (t,x) \in G \\ \sum_k \lambda_k b(t_k, x_k), & \text{for } (t,x) \notin G, (t,x) = \sum_k \lambda_k(t_k, x_k), \\ & \text{and } (t_k, x_k) \in G, \lambda_k \ge 0, \sum_k \lambda_k = 1. \end{cases}
$$

Since $b(t,x)$ is linear in x it is sufficient to interpolate the time component. Note that b_G is again a bounding function for our Markov Decision Model and that $b \le b_G$, thus for $v \in I\!\!M_c$:

$$\|v\|_G := \sup_{(t,x)\in E} \frac{v(t,x)}{b_G(t,x)} \le \|v\|_b.$$

17

If the mesh size in the time component is small, the operator $\mathcal{T}_G$ is again a contraction. To show this we compute

$$\sup_{(t,x,\alpha)} \frac{\int b_G(s,y)q(ds,dy|t,x,\alpha)}{b_G(t,x)}.$$

This can be done essentially as in Proposition 4.1. As an upper bound we obtain $c_G := c_\beta e^{\beta h}$ with the same constant c_β (see equation (4.1)) multiplied by a factor $e^{\beta h}$, where $h > 0$ is the maximal mesh size in the time component. Thus, when we choose β large and h small enough we obtain $c_G < 1$. The following theorem states that our value function V can be approximated arbitrary well by the iterates of the grid operator:

Theorem 5.4. *Suppose that $c_G < 1$. Then it holds for $J_0 \in I\!M_c$ arbitrary, that*

$$\|V - \mathcal{T}_G^n J_0\|_G \leq \frac{1}{1 - c_G}\Big(c_G^n\|\mathcal{T}_G J_0 - J_0\|_G + m(h)\Big)$$

where $m(h) := \|V - \mathcal{T}_G V\|_G \to 0$ if the mesh size tends to zero.

Proof. Since $c_G < 1$ the $\mathcal{T}_G$-operator is contracting and has a unique fixed point which we denote by J_G, i.e. $J_G = \mathcal{T}_G J_G$. An application of the triangle inequality yields

$$\|V - \mathcal{T}_G^n J_0\|_G \leq \|V - J_G\|_G + \|J_G - \mathcal{T}_G^n J_0\|_G$$

for $n \in \mathbb{N}$ arbitrary. By Banach's fixed point theorem we obtain

$$\|J_G - \mathcal{T}_G^n J_0\|_G \leq \frac{c_G^n}{1 - c_G}\|\mathcal{T}_G J_0 - J_0\|_G.$$

Moreover, again by applying the triangle inequality and since $J_G = \mathcal{T}_G J_G$ we obtain

$$\|V - J_G\|_G \leq \|V - \mathcal{T}_G V\|_G + \|\mathcal{T}_G V - \mathcal{T}_G J_G\|_G \leq m(h) + c_G\|V - J_G\|_G.$$

Solving this for
$$\|V - J_G\|_G$$

yields the desired result. Note that $\mathcal{T}_G V(t,x) = V(t,x)$ for $(t,x) \in G$ by definition of the $\mathcal{T}_G$-operator. Since V is continuous (see Proposition 4.4, Theorem 4.5) we obtain

$$m(h) = \|V - \mathcal{T}_G V\|_G \to 0, \text{ for } h \to 0.$$

$\square$

6   A Numerical Example

In this section we present a numerical example for the theoretical results of the preceding sections and compare the value iteration of Section 5.3 with the approximating Markov chain
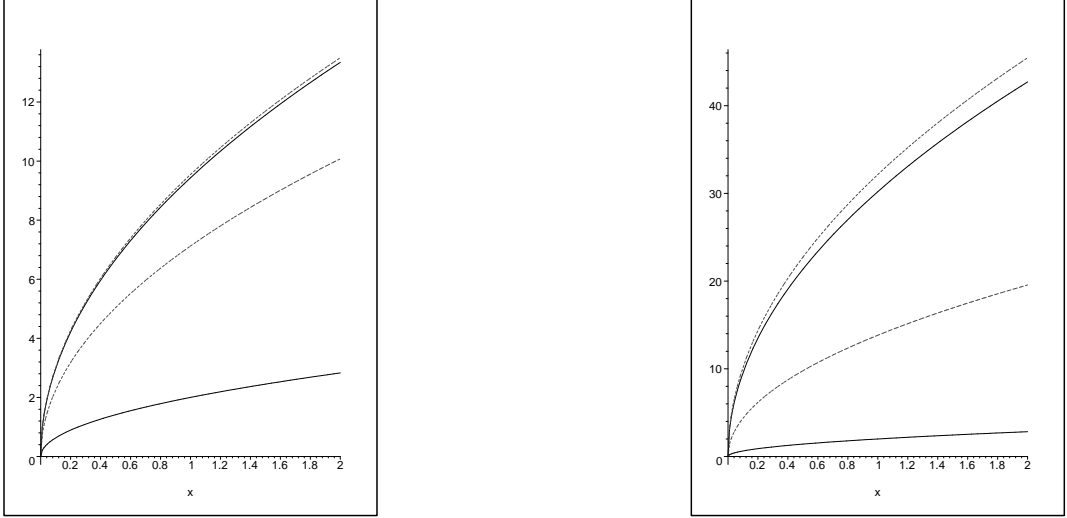
Figure 1: Functions V, U and $J_1, \tilde{V}_0$ for $\lambda = 40$ (left) and $\lambda = 70$ (right).

method. In the approximating Markov chain approach, the state process is approximated by a discrete time Markov chain.

Though our results hold for general utility functions, we have chosen a power utility $U(x) = \frac{1}{\gamma} x^\gamma$ with $\gamma = 0.5$ in the numerical example since this allows to compare the numerical value with the true value function computed as in Section 5. We take one stock, i.e. $d = 1$ and $\mu = r = 0$. The relative jump distribution is a transformed double exponential, i.e. the density is given by $p\lambda_+ e^{-\lambda_+ y} 1_{[y \geq 0]} + (1 - p)\lambda_- (y + 1)^{-\lambda_- - 1} 1_{[-1 < y < 0]}$ where we choose $p = 0.5$ and $\lambda_+ = \lambda_- = 1$. The time horizon is one year, i.e. $T = 1$. A straightforward implementation of the approximating Markov chain approach with grid size $h > 0$ for the time interval $[0, T]$ yields for $N := \frac{T}{h}$, $\tilde{V}_N(x) = U(x)$ and for $n = N - 1, \ldots, 0$:

$$\tilde{V}_n(x) = \sup_{u \in \mathcal{U}} \left\{ e^{-\lambda h} \tilde{V}_{n+1}(x) + (1 - e^{-\lambda h}) \int \tilde{V}_{n+1}\big(x(1 + yu)\big) Q(dy) \right\}$$

where $\tilde{V}_0(x)$ is an approximation for $V(0, x)$. The two figures show $\tilde{V}_0$ (lower dotted line) and the result of one step of value iteration with the grid operator (upper dotted line) $J_1 = \mathcal{T}_G J_0$ for $\lambda = 40$ and $\lambda = 70$ where we have started with $J_0(t, x) = U(x)$. The upper solid line shows the true value function $V(0, x)$, the lower solid line the utility function. For both values of λ and both algorithms we have chosen the mesh size $h = 0.01$. It is remarkable that one iteration step already yields such a good result. Obviously in both cases, the implementation of the grid operator outperforms the approximating Markov chain method using the same grid size. A key reason is that the approximating Markov chain approach is a crude approximation of the time integral which appears in the $\mathcal{T}$ operator. Moreover, a small mesh size h leads for small λ to a large probability $e^{-\lambda h}$ of staying in the state which in turn leads to weak contraction and slow convergence (see Fleming and Soner (1993) p.374 for a discussion). This problem is circumvented in our approach.

The same example also illustrates Howard's policy improvement algorithm when we start

with the decision rule $f \equiv 0$, "invest all the money in the bond". The corresponding value function is given by $J_f(t,x) = U(xe^{r(T-t)}) = U(x)$. In order to compute the first improvement, we have to do exactly the same step as for the value iteration before. The first improvement is given by $g \equiv 27/40$ which is already the optimal fraction.

References

ALMUDEVAR, A. (2001) A dynamic programming algorithm for the optimal control of piecewise deterministic Markov processes. *SIAM J. Control Optim.* **40**, 525–539.

BÄUERLE, N. and RIEDER, U. (2004) Portfolio optimization with Markov-modulated stock prices and interest rates. *IEEE Trans. Automat. Control* **49**, 442–447.

BÄUERLE, N. and RIEDER, U. (2007) Portfolio optimization with jumps and unobservable intensity process. *Math. Finance* **17**, 205–224.

BERTSEKAS, D. P. and SHREVE, S. E. (1978) *Stochastic optimal control: The discrete time case*, volume 139 of *Mathematics in Science and Engineering.* Academic Press Inc., New York.

DAVIS, M. H. A. (1993) *Markov models and optimization*, volume 49 of *Monographs on Statistics and Applied Probability.* Chapman & Hall, London.

FLEMING, W. H. and SONER, H. M. (1993) *Controlled Markov processes and viscosity solutions*, volume 25 of *Applications of Mathematics.* Springer-Verlag, New York.

JACOBSEN, M. (2006) *Point process theory and applications.* Probability and its Applications. Birkhäuser Boston Inc., Boston, MA. Marked point and piecewise deterministic processes.

JOUINI, E. and NAPP, C. (2004) Convergence of utility functions and convergence of optimal strategies. *Finance Stoch.* **8**, 133–144.

KIRCH, M. and RUNGGALDIER, W. J. (2005) Efficient hedging when asset prices follow a geometric Poisson process with unknown intensities. *SIAM J. Control Optim.* **43**, 1174–1195.

KUSHNER, H. J. and DUPUIS, P. (2001) *Numerical methods for stochastic control problems in continuous time*, volume 24 of *Applications of Mathematics (New York).* Springer-Verlag, New York, second edition. Stochastic Modelling and Applied Probability.

NORBERG, R. (2003) The Markov chain market. *Astin Bull.* **33**, 265–287.

ØKSENDAL, B. and SULEM, A. (2005) *Applied stochastic control of jump diffusions.* Universitext. Springer-Verlag, Berlin.

SCHÄL, M. (2005) Control of ruin probabilities by discrete-time investments. *Math. Methods Oper. Res.* **62**, 141–158.