Morfismos, Vol. 4, No. 1, 2000, pp. 19–30

An stochastic consumption–investment problem with unbounded utility function

Heliodoro D. Cruz–Suárez [1]

Abstract

This paper is concerned with a discrete-time, infinite-horizon *consumption and investment* problem, which is formulated as a Markov Decision Process with unbounded utility of consumption, and the total discounted reward criterion. The conditions given in the paper permit to obtain explicit expressions for both the optimal policy and the optimal value function V^*. Moreover, we find a bound for the difference between V^* and a rolling–horizon reward function $\hat{V}$, that is, the discounted reward value when using a rolling-horizon policy.

1991 Mathematics Subject Clasification: 93E20, 90C40
Keywords and phrases: Consumption-investment problems; Markov decision processes; Discounted reward criterion; Value iteration.

1 Introduction

Consumption-investment problems appear in Mathematical Economics and Mathematical Finance, and they have been studied by many authors (see, for instance, [3, §3.6],[4], [5] and their references.)

This paper presents an *infinite-horizon* version of a discrete-time, finite-horizon consumption-investment problem that appears in [3]. Here, the problem is formulated as a Markov Decision Process (MDP) with unbounded reward function and total discounted reward criterion (see [1]). Our formulation allows us to obtain explicit expressions for both

[1]This paper is part of the author's M.Sc.Thesis presented at the Instituto de Ciencias Básicas, Universidad Veracruzana (October 1999).

the optimal value function and the optimal stationary policy. Furthermore, using results from [2] we find an error bound for the difference between the optimal value function and the rolling–horizon reward function, that is, the discounted reward value when using a rolling-horizon (or value iteration) policy. (See [4] for a review of the literature of rolling-horizon procedures).

The paper is organized as follows. Section 2 introduces the problem we are interested in. In Section 3 we present a summary of the theory of Markov Decision Processes (MDPs), which we use in the paper. Section 4 provides a formulation of our problem as an MDP, and presents our main results. We conclude in Section 5 with some general remarks.

2 The consumption-investment problem

In this section we state the problem we are concerned with. We closely follow [3, §3.6], except that here we are interested in a problem with an *infinite* time horizon. (In Section 4 we reformulate the problem as an MDP.)

An investor wishes to allocate his current wealth x_t between investment a_t and consumption $x_t - a_t$, at each time $t = 0, 1, \ldots$. We assume that borrowing is not allowed. Then the set

$$A(x) = [0, x]$$

consists of the admissible investment values given the wealth x. The relation between investment decision and capital is given by

$$x_{t+1} = a_t \cdot \xi_t, \tag{2.1}$$

so that the wealth at time $t + 1$ is proportional to the amount invested at time t, where we suppose that ξ_t are independent and identically distributed random variables independent of the initial wealth x_0.

We also have the utility function

$$u(x) := \frac{b}{\gamma} x^{\gamma} \qquad \text{for } \ x \geq 0, \tag{2.2}$$

where $b > 0$ and $0 < \gamma < 1$ are constants. Thus, the reward function r is given by

$$r(x, a) := u(x - a), \tag{2.3}$$

so that r is a "utility of consumption".

The objective is to maximize over all "rules" of investment π, the expected total discounted reward

$$V(\pi, x) := E_x^\pi \left[\sum_{t=0}^{\infty} \alpha^t r(x_t, a_t) \right],$$

where x is the initial wealth, and α $(0 < \alpha < 1)$ is a discount factor. To analize this problem we postulate the following assumption.

Assumption 2.1.

(a) The ξ_t are nonnegative random variables with a common density f;

(b) $m := E[\xi_0]$ and $m_\gamma := E[\xi_0^\gamma]$ are finite, $m > 1$, and

$$0 < \alpha \cdot m_\gamma < 1, \tag{2.4}$$

where γ is the constant in (2.2).

3 Discounted MDPs

In this section we present some results on the theory that we will use to solve our problem.

Let (X, A, Q, r) be a discrete-time, stationary, Markov control model (see, e.g., [3] for notation and terminology), which consists of the state space X, the control (or action) set A, the transition law Q, and the one-stage reward r. The sets X and A are assumed to be Borel spaces, with Borel σ-algebras $\mathcal{B}(X)$ and $\mathcal{B}(A)$, respectively. Moreover, for every $x \in X$ there is a nonempty Borel set $A(x) \subset A$ whose elements are the feasible control actions when the state of the system is x. Define $\mathbb{K} := \{(x, a) | x \in X,\ a \in A(x)\}$. The transition law $Q(B|x, a)$ is a stochastic kernel on X given $\mathbb{K}$ (that is, $Q(\cdot|x, a)$ is a probablility measure on X for every $(x, a) \in \mathbb{K}$, and $Q(B|\cdot)$ is measurable function on $\mathbb{K}$ for every $B \in \mathcal{B}(X)$). Finally, r is a measurable and (possibly) unbounded function, which represents the reward–per–stage.

We suppose:

Assumption 3.1. For each state $x \in X$,

(a) $A(x)$ is a (nonempty) compact subset of A;

(b) $r(x, a)$ is upper semicontinuous in $a \in A(x)$;

(c) The function u' on $\mathbb{K}$ defined as:

$$u'(x,a) := \int u(y)Q(dy|x,a) \tag{3.1}$$

is upper semicontinuous in $a \in A(x)$ for every measurable bounded function u on X.

Remark 3.2. Assumption 3.1 guarantees the existence of measurable maximizers or selectors for dynamic programming equations, such as (3.5) and (3.8), below.

Trivially, in our problem Assumptions 3.1(a) and 3.1(b) hold. On the other hand, by the continuity of the functions that appear in the dynamic programming equations of our problem, it suffices to verify (see Lemma 4.1) the following condition, which is weaker than 3.1(c):

3.1(c'). The function u' in (3.1) is continuous and bounded on $\mathbb{K}$ for every continuous bounded function u on X.

Using standard notation and definitions [3], we denote by Π the set of all policies and by $\mathbb{F}$ the subset of stationary policies. We identify each stationary policy $f \in \mathbb{F}$ with the measurable (Borel) function $f : X \to A$ such that $f(x) \in A(x)$ for every $x \in X$.

We focus here on the *expected total discounted reward* defined as

$$V(\pi, x) := E_x^\pi \left[\sum_{t=0}^{\infty} \alpha^t r(x_t, a_t) \right] \tag{3.2}$$

when the policy $\pi \in \Pi$ is used, and $x \in X$ is the initial state. In (3.2), $\alpha \in (0, 1)$ is a given discount factor, and E_x^π denotes the expectation with respect to the probability measure P_x^π induced by the pair (π, x) [3].

A policy π^* is said to be *optimal* if

$$V(\pi^*, x) = V^*(x) \qquad x \in X,$$

where

$$V^*(x) := \sup_{\pi \in \Pi} V(\pi, x) \tag{3.3}$$

is the so-called *optimal value function*.

Now, we list some definitions and results to be used in the next section.

Definition 3.3. For each nonnegative measurable function v on X, we define the function Hv by

$$(Hv)(x) := \sup_{a \in A(x)} \int v(y)Q(dy|x,a) \quad \text{for } x \in X.$$

Let $H^n v := H(H^{n-1}v)$ for $n = 1, 2, \ldots$, with $H^0 v = v$.

The following assumption ensures the result in Lemma 3.8, below.

Assumption 3.4.

$$R(x) := \sum_{t=0}^{\infty} \alpha^t H^t r_0(x) < \infty \tag{3.4}$$

for every $x \in X$, where

$$r_0(x) := \sup_{a \in A(x)} |r(x,a)|.$$

(Note that Assumption 3.4 holds if r is bounded).

Let $\mathcal{R} := \{V : X \to \mathbb{R},\ V \text{ measurable and } |V| \leq R\}$, with R as in (3.4).

Definition 3.5. The *value iteration (VI) functions* are defined as

$$V_n(x) := \sup_{a \in A(x)} \left[r(x,a) + \alpha \int V_{n-1}(y)Q(dy|x,a) \right] \tag{3.5}$$

for all $x \in X$ and $n = 1, 2, \ldots$, with $V_0(\cdot) \equiv 0$.

Remark 3.6. Under Assumptions 3.1 and 3.4, for each $n = 1, 2, \ldots$, there exists a stationary policy $f_n \in \mathbb{F}$ such that the supremum in (3.5) is attained, i.e.,

$$V_n(x) = r(x, f_n(x)) + \alpha \int V_{n-1}(y)Q(dy|x, f_n(x)) \tag{3.6}$$

for all $x \in X$ and $n = 1, 2, \ldots$, and f_n is said to be an n^{th} **VI decision function**.

Definition 3.7. Let N be a fixed positive integer. We define the (stationary) *rolling horizon* (RH) policy as $\hat{f} := f_N$, the N^{th} VI decision function. Furthermore, let

$$\hat{V}(x) := V(\hat{f}, x) \quad \text{for } x \in X \tag{3.7}$$

be the corresponding reward function when using the RH policy $\hat{f}$.

Finally, we have the following result.

Lemma 3.8. [3, Theorems 4.1 and 4.2]. Suppose that Assumptions 3.1 and 3.4 hold. Then:

(a) The optimal value function V^* in (3.3) is the unique function in $\mathcal{R}$ that satisfies the *Optimality Equation*

$$V^*(x) = \sup_{a \in A(x)} \left[r(x,a) + \alpha \int V^*(y) Q(dy|x,a) \right] \quad \forall x \in X. \quad (3.8)$$

(b) For every $x \in X$, $V_n(x) \to V^*(x)$ as $n \to +\infty$, with V_n as in (3.6).

(c) Let $\hat{V}$ be as in (3.7). Then

$$0 \le V^*(x) - \hat{V}(x) \le \alpha^N H^N R(x) + \sum_{t=N}^{\infty} \alpha^t H^t r_0(x) \quad (3.9)$$

$$\le 2 \sum_{t=N}^{\infty} \alpha^t H^t r_0(x) \quad \forall x \in X. \quad (3.10)$$

4 Results

This section gives the solution to the consumption-investment problem described in Section 2.

We begin with the following remark, in which the problem is formulated as an MDP.

Remark 4.1 The problem described in Section 2 can be stated as an MDP in the following manner.

(a) Take $X = A = [0, \infty)$, and $A(x) := [0, x]$ for $x \in X$;

(b) The transition law Q is defined via (2.1) as

$$Q(B|x,a) = \begin{cases} \displaystyle\int_B \frac{1}{a} f(z/a) dz & \text{if} \quad a \neq 0, \\ I_B(0) & \text{if} \quad a = 0, \end{cases} \quad (4.1)$$

where $(x,a) \in \mathbb{K}$, $B \in \mathcal{B}(X)$, f is the density of ξ_t (see Assumption 2.1), and I_B denotes the indicator function of $B \in \mathcal{B}(X)$;

(c) By (2.2) and (2.3), the reward function $r : \mathbb{K} \to \mathbb{R}$ is given by

$$r(x,a) = \begin{cases} \dfrac{b}{\gamma}(x-a)^{\gamma} & \text{if} \quad x \neq a, \\ \\ 0 & \text{if} \quad x = a. \end{cases} \tag{4.2}$$

Note that, for each $x \in X$, $A(x)$ is compact and $r(x,a)$ is continuous in $a \in A(x)$. Also, observe that r is unbounded.

In the remainder of the paper we concentrate on the Markov control model in Remark 4.1. The optimization criterion we are interested in is the expected total discounted reward in (3.2) and (3.3).

Before stating the solution to our problem, in Theorem 4.5, below, we present three preliminary results.

Lemma 4.2. Suppose that Assumption 2.1 holds. Then Assumption 3.1(a), (b) and (c') hold. Moreover, for each nonnegative continuous function $W : X \to \mathbb{R}$ such that

$$\int_X W(y)Q(dy|x,a) < \infty \qquad \forall a \in A(x),\ x \in X, \tag{4.3}$$

there exists a stationary policy $g \in \mathbb{F}$ such that

$$\sup_{a \in A(x)} \left[r(x,a) + \alpha \int W(y)Q(dy|x,a) \right]$$

$$= r(x,g(x)) + \alpha \int W(y)Q(dy|x,g(x)) \ \ \forall x. \tag{4.4}$$

Proof. Obviously, Assumption 3.1(a) and (b) hold in our problem (see Remark 4.1).

To verify Assumption 3.1(c'), choose an arbitrary state $x \in X$. Let $a \in A(x)$, and let $u : X \to \mathbb{R}$ be continuous and bounded. Assume that $a_n \in A(x)$ is such that $a_n \to a$. Note that, by (4.1),

$$\int_X u(y)Q(dy|x,a_n) = \int_0^{\infty} u(s \cdot a_n) f(s) ds. \tag{4.5}$$

Now, let M be a bound for u. Then

(a) $|u(s \cdot a_n) f(s)| \leq M f(s) \quad \forall s \in [0,\infty)$, and, furthermore, by the continuity of u,

(b) $u(s \cdot a_n) \to u(s \cdot a)$ for each $s \in [0,\infty)$.

Since f is integrable, the Dominated Convergence Theorem and (4.5) yield

$$\begin{aligned}\lim_{n\to\infty}\int_X u(y)Q(dy|x,a_n) &= \int_0^\infty u(s\cdot a)f(s)ds\\ &= \int_X u(y)Q(dy|x,a).\end{aligned}$$

Hence, as $x\in X$ and $a\in A(x)$ were arbitrary, we obtain Assumption 3.1(c').

Finally, to prove the last part of the lemma it sufficies to show that the function

$$v(x,a):=r(x,a)+\alpha\int W(y)Q(dy|x,a) \tag{4.6}$$

is upper semicontinuous (u.s.c.) on $A(x)$ for each $x\in X$, because then the existence of $g\in\mathbb{F}$ that satisfies (4.4) follows from well-known measurable selection theorems. Moreover, by (4.2), it is obvious that $r(x,\cdot)$ is u.s.c. (in fact, continuous) on $A(x)$. Thus, to prove that $v(x,\cdot)$ is u.s.c., it suffices to show that so is the integral in (4.3). To get this, choose an arbitrary $x\in X$, and let $\{a^k\}\subset A(x)$ be such that $a^k\to a\in A(x)$. Then

$$\int_X W(y)Q(dy|x,a^k)=\int_0^\infty W(sa^k)f(s)ds,$$

and $W(sa^k)\to W(sa)$, by the continuity of W. Hence, by (4.3) and Fatou's Lemma,

$$\begin{aligned}\limsup_{k\to\infty}\int_X W(y)Q(dy|x,a^k) &= \limsup_{k\to\infty}\int_0^\infty W(sa^k)f(s)ds\\ &\le \int_0^\infty W(sa)f(s)ds\\ &= \int_X W(y)Q(dy|x,a),\end{aligned}$$

which yields that the integral in (4.3) is u.s.c. on $A(x)$. Thus, as $x\in X$ was arbitrary, it follows that $v(x,\cdot)$ is u.s.c. on $A(x)$ for each $x\in X$. This completes the proof of the lemma. $\square$

Lemma 4.3. If Assumption 2.1 holds, then Assumption 3.4 also holds.

Proof. We want to prove that the series in (3.4), i.e.,

$$\sum_{t=0}^{\infty}\alpha^t(H^t r_0)(x), \tag{4.7}$$

converges for every $x \in X$. Let us first take $x = 0$. Then, by Remark 4.1(a), (c), we have $A(0) = \{0\}$ and $r_0(0) = 0$. Therefore,

$$(H^t r_0)(0) = 0 \qquad \forall t = 0, 1, \ldots$$

and so (4.7) converges.

Suppose now that $x > 0$. Then

$$\begin{aligned} r_0(x) &:= \sup_{a \in [0,x]} |r(x,a)| \\ &= \sup_{a \in [0,x)} |r(x,a)| \\ &= \sup_{a \in [0,x)} \frac{b}{\gamma}(x-a)^\gamma \\ &= \frac{b}{\gamma} x^\gamma. \end{aligned}$$

That is,

$$r_0(x) = \begin{cases} \dfrac{b}{\gamma} x^\gamma & \text{if} \quad x > 0, \\ 0 & \text{if} \quad x = 0. \end{cases}$$

Therefore, for $x > 0$ we have

$$\begin{aligned} (H r_0)(x) &= \sup_{a \in [0,x]} \int r_0(y) Q(dy|x,a) \\ &= \sup_{a \in [0,x]} \int_{(0,\infty)} \frac{b}{\gamma} y^\gamma Q(dy|x,a) \\ &= \frac{b}{\gamma} \sup_{a \in (0,x]} \int_{(0,\infty)} (s \cdot a)^\gamma f(s) ds \\ &= \frac{b}{\gamma} m_\gamma \sup_{a \in (0,x]} a^\gamma \\ &= m_\gamma \frac{b}{\gamma} x^\gamma. \end{aligned}$$

Note that we have omitted $a = 0$, since

$$\int r_0(y) Q(dy|x,0) = r_0(0) = 0.$$

Similarly, by induction, it is possible to show that

$$(H^t r_0)(x) = m_\gamma^t \frac{b}{\gamma} x^\gamma \qquad \forall t = 0, 1, \ldots, \text{ and } x > 0.$$

Hence, for $x > 0$,

$$\sum_{t=0}^{\infty} \alpha^t (H^t r_0)(x) = \frac{b}{\gamma} x^\gamma \sum_{t=0}^{\infty} (\alpha m_\gamma)^t,$$

which converges, by (2.4). This completes the proof of the lemma. □

The proof of the next lemma can be found in Section 3.6 of [3].

Lemma 4.4. Let $\delta := [\alpha m_\gamma]^{\frac{1}{\gamma-1}}$. Then the VI functions are

$$V_0(x) = 0 \qquad \forall x \in X,$$

$$V_n(x) = \begin{cases} \left[\delta^{n-1} \dfrac{1-\delta}{1-\delta^n}\right]^{\gamma-1} \dfrac{b}{\gamma} x^\gamma & \text{if} \quad x > 0, \\ 0 & \text{if} \quad x = 0, \end{cases} \tag{4.8}$$

for $n = 1, 2, \ldots$. Furthermore, the VI decision functions (in Remark 3.6) are given by

$$f_n(x) = \frac{1-\delta^{n-1}}{1-\delta^n} x \qquad \forall x \in X, \ \ n = 1, 2, \ldots \tag{4.9}$$

We now present our main results.

Theorem 4.5. Suppose that Assumption 2.1 holds. Then for the MDP formulated in Remark 4.1 we have:

(a) The optimal value function V^* (see (3.3)) is given by

$$V^*(x) = \exp\left[(\gamma - 1) \ln\left(\frac{1-\delta}{-\delta}\right)\right] \frac{b}{\gamma} x^\gamma \quad \text{for} \quad x > 0, \tag{4.10}$$

and $V^*(0) = 0$. Moreover, the optimal stationary policy f^* is given by

$$f^*(x) = \frac{x}{\delta} \quad \forall x \geq 0. \tag{4.11}$$

(b) Let N be a fixed positive integer. Let $\hat{f} = f_N$ be the RH policy (see Definition 3.7) given by (4.9), i.e.,

$$\hat{f}(x) := \frac{1-\delta^{N-1}}{1-\delta^N} x \quad \forall x \in X \tag{4.12}$$

Then

$$
\begin{aligned}
0 \le V^*(x) - \hat{V}(x) &\le 2\sum_{t=N}^{\infty} \alpha^t H^t r_0(x) \\
&= 2\frac{(\alpha m_\gamma)^N}{1-\alpha m_\gamma}\frac{b}{\gamma}x^\gamma \quad \forall x \ge 0,
\end{aligned}
\tag{4.13}
$$

where $\hat{V}(x) = V(\hat{f}, x),\ x \in X$.

Proof. First note that, by Lemmas 4.2 and 4.3, we have that Assumptions 3.1 and 3.4 hold. Hence, the results in Lemma 3.8 hold.

(a) From Lemma 3.8(b) and Lemma 4.4 we have, for $x > 0$,

$$
\begin{aligned}
V_n(x) &= \left[\delta^{n-1}\frac{1-\delta}{1-\delta^n}\right]^{\gamma-1}\frac{b}{\gamma}x^\gamma \\
&= \exp\left[(\gamma-1)\ln\left(\delta^{n-1}\frac{1-\delta}{1-\delta^n}\right)\right]\frac{b}{\gamma}x^\gamma \\
&= \exp\left[(\gamma-1)\ln\left(\frac{1-\delta}{1/\delta^{n-1}-\delta}\right)\right]\frac{b}{\gamma}x^\gamma.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
V^*(x) &= \lim_{n\to\infty} V_n(x) \\
&= \exp\left[(\gamma-1)\ln\left(\frac{1-\delta}{-\delta}\right)\right]\frac{b}{\gamma}x^\gamma,
\end{aligned}
$$

and (4.10) follows. Obviously, we also have $V^*(0) = 0$.

Similarly, to obtain f^*, we use (4.9) to get:

$$
\begin{aligned}
f_n(x) &= \frac{1-\delta^{n-1}}{1-\delta}x \\
&= \frac{1/\delta^{n-1}-1}{1/\delta^{n-1}-\delta}x,
\end{aligned}
$$

which yields (4.11) because

$$
f^*(x) = \lim_{n\to\infty} f_n(x) = \frac{x}{\delta}.
$$

(b) (4.13) follows from Lemma 3.8(c) and the calculations in the proof of Lemma 4.3. This completes the proof of the theorem. □

Remark. Straightforward calculations show that V^* and f^* given in (4.10) and (4.11), respectively , satisfy the Optimality Equation (3.8), i.e.,

$$V^*(x) = r(x, f^*(x)) + \alpha \int V^*(y) Q(dy|x, f^*(x)) \quad \forall x \in X.$$

5 Concluding remarks.

We have presented an infinite-horizon consumption-investment problem that, in particular, illustrates the theory of discrete-time MDPs with unbounded reward function and discounted criterion. Mild assumptions guarantee the existence of an optimal policy and the convergence of the value iteration (VI) procedure, which provides error bounds for RH policies.

Acknowledgement. This article is the result of a problem proposed and supervised by Dr. Raúl Montes de Oca Machorro of the Universidad Autónoma Metropolitana-Iztapalapa. I wish to thank him for his guidance, which made this paper possible. Also, thanks to M. C. Robert J. Flowers for the revision of this paper and helpful comments.

References

[1] H. D. Cruz-Suárez, *Rolling Horizon Policies in Markov control Processes with Total Discounted Reward Criterion*, Sc. Thesis, Instituto de Ciencias Básicas, Universidad Veracruzana. (In Spanish) 1999.

[2] O. Hernández-Lerma and J. B. Lasserre, *Value iteration and rolling plans for Markov control processes with unbounded rewards*, J. Math. Anal. Appl., 177 (1993), pp. 38-55.

[3] O. Hernández-Lerma and J. B. Lasserre, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, Springer–Verlag, New York, 1996.

[4] S. P. Sethi and G. Sorger, *A theory of rolling horizon decision making*, Ann. Oper. Res., 29 (1991), pp. 387-416.

[5] S. P. Sethi, *Optimal consumption-investment decisions allowing for bankruptcy: a survey*, Working paper, University of Toronto.