Math. Meth. Oper. Res. (2007) 65: 27–44
DOI 10.1007/s00186-006-0092-2

ORIGINAL ARTICLE

Juan González-Hernández
Raquiel R. López-Martínez
J. Rubén Pérez-Hernández

Markov control processes with randomized discounted cost

Received: 14 July 2005 / Accepted: 3 July 2006 / Published online: 7 November 2006

Abstract In this paper we consider Markov Decision Processes with discounted cost and a random rate in Borel spaces. We establish the dynamic programming algorithm in finite and infinity horizon cases. We provide conditions for the existence of measurable selectors. And we show an example of consumption-investment problem.

Mathematics Subject Classification (2000) 90C40 · 93E20

Keywords Markov decision processes · Discounted cost · Random rate · Dynamic programming

1 Introduction

In this paper we consider a Markov decision problem (MDP) with discounted cost, in which the discount rate is random in each stage. Generally, interest rates in economic and financial models are stochastic processes (see, for example, Berument

This research was partially supported by the PROMEP grant 103.5/05/40.

J. González-Hernández
Departamento de Probabilidad y Estadística, IIMAS-UNAM, A. Postal 20–726,
Mexico D.F. 01000, Mexico
E-mail: juan@sigma.iimas.unam.mx

R. R. López-Martínez (✉)
Facultad de Matemáticas UV, A. Postal 270 Xalapa, Ver.91090, Mexico
E-mail: ralopez@uv.mx

J. R. Pérez-Hernández
Departamento de Ciencias Básicas, UPIITA-IPN, Mexico, Mexico
E-mail: eriiva@hotmail.com

et al. 2004; Gil and Luis 2004; Haberman and Sung 2005; Lee and Rosenfield 2005; Newell and Pizer 2003; Ogaki and Santaella 2000; Sack and Wieland 2000; Stockey and Lucas 1989). However, until now, it has been considered just fixed rates in discounted cost MDPs. To the best of our knowledge, this is the first article on discounted cost MDPs with where a random rate is considered. Empirical models of interest rate have been modeled in Gil and Luis (2004), Ogaki and Santaella (2000), and Sack and Wieland (2000). The effects of uncertain rates have been analyzed in Berument et al. (2004) and Newell and Pizer (2003). Haberman and Sung (2005) and Lee and Rosenfield (2005) used dynamic programming with a constant discount rate even thought they consider random rates in some parts of the dynamic of their systems. See Stockey and Lucas (1989) for economic applications with fixed discount rate. See also Cai et al. (2004) and Lettau and Uhlig (1999).

Discounted MDPs with fixed discount rate can be seen in Abbad and Daoui (2003), Feinberg and Shwartz (1994, 1995, 1999), Hernández-Lerma and González-Hernández (2000), Hu (2003), Kurano et al. (1998), Liu (1999), López-Martínez and Hernández-Lerma (2003), Mao and Piunovskiy (2000), Michael (1999), Shwartz (2001), and in Altman (1999, Chaps. 3 and 10), Bertsekas (1995, Sect. 5.4), Borkar (1991, Chap. III), Hernández-Lerma and Lasserre (1996, pp. 32 and 33, Chap. 4), Piunovskiy (1997, Sect. 2.2), Puterman (1994, pp. 146–163, Chap. 6). The MDPs with discounted cost have been studied with different approaches as dynamic programming (Kurano et al. 1998; Michael 1999), convex analysis (Feinberg and Shwartz 1994, 1995, 1999; Mao and Piunovskiy 2000), linear programming (Hernández-Lerma and González-Hernández 2000), Lagrange multipliers (López-Martínez and Hernández-Lerma 2003). In recent years there is a growing interest to consider different discount factors, for example, Feinberg and Shwartz (1994, 1995, 1999) and Shwartz (2001) studied the case with two or more fixed discounted factors and described several applications.

We begin Sect. 2 by defining a Markov decision model with performance criterion the expected discounted cost and with a randomized rate of discount, where the state and action spaces are Borel spaces. We finish this section with the canonical construction. In Sect. 3 we study finite horizon MDPs and by using the dynamic programming algorithm we prove the existence of optimal policies, under the assumption of the existence of measurable selectors that satisfy optimality equations. In Sect. 4 we give conditions that assure the existence of such selectors. In Sect. 5 we present infinite horizon MDPs, we show the optimality equation and we prove the existence of optimal policies. In Sect. 6 we give an example of consumption–investment problem and we find the solution.

2 Markov control model

The classical model of discounted cost MDP considers a constant discount factor $\beta = (1+r)^{-1}$, where r is a fixed interest rate. We consider instead of $e^{-\alpha} = (1+r)^{-1}$ where r is a random rate, so $\alpha = \ln(1+r)$. Since r lies in $(0, \infty)$, then α lies in $(0, \infty)$. For this reason we consider state-space $X \times (0, \infty)$. The elements of a Markov control model (MCM) and notation we use throughout is five-tuple.

$$\big(X, A, \{A(x,\alpha) | (x,\alpha) \in X\}, Q, c\big), \tag{1}$$

where

(a) $X = X' \times (0, \infty)$ is state space and X' is Borel space,
(b) A is action space and Borel space,
(c) a family $\{A(x, \alpha)|(x, \alpha) \in X\}$ of nonempty measurable subsets $A(x, \alpha)$ of A, where $A(x, \alpha)$ denotes the set of feasible control or actions when the system is in state $(x, \alpha) \in X$, and with the property that the set

$$\mathbb{K} = \left\{(x, \alpha, a) : a \in A(x, \alpha),\ (x, \alpha) \in X\right\} \tag{2}$$

of feasible state-actions pairs is a measurable subset of $X \times A$,
(d) a stochastic kernel Q on X given $\mathbb{K}$ called *transition law*,
(e) a measurable function $c : \mathbb{K} \to \mathbb{R}$ called the *cost-per-stage function*. For our purposes, we consider that one stage-cost function c is constant with respect to the variable α.

We note that all results are true, with obvious changes, if we consider a general function cost $c(x, \alpha, a)$ on $\mathbb{K}$. But the examples that we are interested in correspond to model (1).

Throughout the following, we suppose that:

Assumption 1 The set $\mathbb{K}$ contains the graph of a measurable function from X to A.

This assumption ensures that the sets in Definition 1 are nonempty.

For each $t = 0, 1, \ldots$, define the space H_t of *admissible histories* up to time t as $H_0 = X_0$, and

$$H_t := \left[\prod_{i=0}^{t-1} \mathbb{K}\right] \times X \quad \text{for} \quad t = 1, 2, \ldots, \tag{3}$$

where $\mathbb{K}$ is the set in (2). A generic element h_t of H_t, which is called an admissible t-history, is a vector of the form

$$h_t = (x_0, \alpha_0, a_0, \ldots, x_{t-1}, \alpha_{t-1}, a_{t-1}, x_t, \alpha_t) \tag{4}$$

with $(x_i, \alpha_i, a_i) \in \mathbb{K}$ for $i = 0, \ldots, t-1$, and $(x_t, \alpha_t) \in X$. Observe that, for each t, H_t is a subset of

$$\bar{H}_t = \prod_{i=0}^{t-1} (X \times A) \times X \quad \text{for} \quad t = 1, 2, \ldots, \tag{5}$$

and $\bar{H}_0 = H_0 = X_0$.

A policy is a sequence of actions that is taken by the controller, that is.

Definition 1 (a) *A randomized control policy is a sequence $\pi = \{\pi_t, t = 0, 1, \ldots\}$ of stochastic kernels π_t on the control set A given H_t, satisfying the constraint*

$$\pi_t(A(x_t, \alpha_t)|h_t) = 1 \quad \forall\, h_t \in H_t, \quad t = 0, 1, \ldots. \tag{6}$$

(b) *The set* Φ *denotes the set of all stochastic kernels* φ *in* $P(A|X)$*, such that* $\varphi(A(x,\alpha)|x,\alpha) = 1$ *for all* $(x,\alpha) \in X$*, and* $\mathbb{F}$ *stands for the set of all measurable functions* $f : X \to A$ *satisfying the constraint* $f(x,\alpha) \in A(x,\alpha)$ *for all* $(x,\alpha) \in X$*. The functions in* $\mathbb{F}$ *are called selectors of the multifunction* $(x,\alpha) \mapsto A(x,\alpha)$.

The set of all policies is denoted by Π. As usual, we will identify Φ with the family of *randomized stationary* policies, and $\mathbb{F}$ with the subfamily of *deterministic stationary* policies. In this way, we have that $\mathbb{F} \subset \Phi \subset \Pi$.

With these elements we can define the next stochastic processes.

The canonical construction (see, Hinderer 1970, pp. 78–83)

Let $(\Omega, \mathcal{F})$ be the measurable space consisting of the (canonical) sample space $\Omega := \bar{H}_\infty = \prod_{t=1}^{\infty}(X \times A)$ and $\mathcal{F}$ is the corresponding product σ-algebra. The element of Ω are sequence of the form $w = (x_0, \alpha_0, a_0, x_1, \alpha_1, a_1, \ldots)$ with (x_t, α_t) in X and a_t in A for all $t = 0, 1 \ldots$; the projections (x_t, α_t) and a_t from Ω to the sets X and A are called state and control (or action) variables, respectively. Observe that Ω contains the space $H_\infty = \prod_{t=0}^{\infty} \mathbb{K}$ of admissible histories $(x_0, \alpha_0, a_0, x_1, \alpha_1, a_1, \ldots)$ with $(x_t, \alpha_t, a_t) \in \mathbb{K}$ for all $t = 0, 1, \ldots$

Let $\pi = \{\pi_t\}$ be an arbitrary control policy and ν an arbitrary probability measure on X, referred to as the *initial distribution*. Then, by the theorem of C. Ionescu-Tulcea (Hernández-Lerma and Lasserre 1996, Appendix C; Hinderer 1970, pp. 78–83) there exists a unique probability measure P_ν^π on $(\Omega, \mathcal{F})$ which, by (6), is supported on H_∞, namely $P_\nu^\pi(H_\infty) = 1$, and, moreover, for all $B \in \mathcal{B}(X)$, $C \in \mathcal{B}(A)$ and $h_t \in H_t$ as in (4), $t = 0, 1, \ldots$:

$$P_\nu^\pi\big((x_0, \alpha_0) \in B\big) = \nu(B), \tag{7}$$

$$P_\nu^\pi(a_t \in C|h_t) = \pi_t(C|h_t), \tag{8}$$

$$P_\nu^\pi\big((x_{t+1}, \alpha_{t+1}) \in B|h_t, a_t\big) = Q(B|x_t, \alpha_t, a_t). \tag{9}$$

The stochastic process $\big(\Omega, \mathcal{F}, P_\nu^\pi, \{(x_t, \alpha_t)\}\big)$ is called a discrete-time Markov control process (or Markov decision process). The expectation operator with respect to P_ν^π is denoted by E_ν^π. If ν is concentrated at the *initial state* $(x,\alpha) \in X$, then we write P_ν^π and E_ν^π as $P_{(x,\alpha)}^\pi$ and$E_{(x,\alpha)}^\pi$, respectively.

Interpretation We observe the system in discrete time (days, months, years,...). The system starts at the state (x_0, α_0) and we apply a policy $\pi = \{\pi_t\}$ in the following way: we choose an action a_0 with distribution law $\pi_0(\cdot|h_0)$, which incurs the an immediate cost $c(x_0, a_0)$. Then, the system evolves to a new state (x_1, α_1) according to the transition law $Q(\cdot|x_0, \alpha_0, a_0)$. Now, we choose an action a_1 with distribution law $\pi_1(\cdot|h_1)$, which generates a new cost $c(x_1, a_1)$ and the system moves to another state (x_2, α_2) according to transition law $Q(\cdot|x_1, \alpha_1, a_1)$. The process is repeated at each time t within the problem's planning horizon.

The following notation will be useful for us.

Let $\phi \in \Phi$, $g : X \times A \to \mathbb{R}$ a measurable function, and Q a stochastic Kernel on X given $\mathbb{K}$.

Then we define

$$g(x, \alpha, \varphi) = \int_A g(x, \alpha, a)\varphi(\mathrm{d}a|x, \alpha)$$

and

$$Q(\cdot|x, \alpha, \varphi) = \int_A Q(\cdot|x, \alpha, a)\varphi(\mathrm{d}a|x, \alpha).$$

In particular for a function $f \in \mathbb{F}$ $(\subseteq \Phi)$, we have $g(x, \alpha, f) = g(x, \alpha, f(x, \alpha))$ and $Q(B|x, \alpha, f) = Q(B|x, \alpha, f(x, \alpha))$. Note that each of these functions is measurable.

3 Finite-horizon problems

In this section we consider the Markov control model (1) with a finite planning horizon N. The present value of the current cost in stage t is given by

$$\mathrm{e}^{-S_t} c(x_t, a_t)$$

where $S_0 = 0$ and $S_t := \alpha_0 + \cdots + \alpha_{t-1}$ for $t = 1, \ldots, N-1$. Finally, we consider that at stage N there is a terminal cost $C(x_N)$. That is, The control problem we are interested in is to minimize the finite horizon performance criterion

$$J(\pi, x, \alpha) := E^{\pi}_{(x,\alpha)} \left[\sum_{t=0}^{N-1} \mathrm{e}^{-S_t} c(x_t, a_t) + \mathrm{e}^{-S_N} c_N(x_N) \right]. \tag{10}$$

Thus, denoting by J^* the value function, i.e.,

$$J^*(x, \alpha) := \inf_{\Pi} J(\pi, x, \alpha), \quad (x, \alpha) \in X_0, \tag{11}$$

the problem is to find a policy $\pi^* \in \Pi$ such that

$$J(\pi^*, x, \alpha) = J^*(x, \alpha) \quad \forall (x, \alpha) \in X. \tag{12}$$

Our main result in this section is the following *Dynamical Programming (DP)* theorem that gives the value function and a deterministic optimal policy.

Theorem 2 *For $t = 0, 1, \ldots, N$, let J_t be the function on X, defined (backward, from $t = N$ to $t = 0$) by*

$$J_N(x, \alpha) := c_N(x) \tag{13}$$

and for $t = N-1, N-2, \ldots, 0$,

$$J_t(x, \alpha) := \min_{A(x,\alpha)} \left[c(x, a) + \mathrm{e}^{-\alpha} \int_X J_{t+1}(y, \beta) Q(\mathrm{d}(y, \beta)|x, \alpha, a) \right]. \tag{14}$$

Suppose that these functions are measurable and that, for each $t = 0, \ldots, N-1$, *there is a selector* $f_t \in \mathbf{F}$ *such that* $f_t(x, \alpha) \in A(x, \alpha)$ *attains the minimum in* (14) *for all* $(x, \alpha) \in X$*; that is, for all* (x, α) *in* X *and* $t = 0, \ldots, N-1$,

$$J_t(x, \alpha) := c(x, f_t) + \mathrm{e}^{-\alpha} \int_X J_{t+1}(y, \beta) Q(\mathrm{d}(y, \beta)|x, \alpha, f_t). \tag{15}$$

Then the policy $\pi^* = \{f_0, \ldots, f_{N-1}\}$ *is optimal and the value function* J^* *equals* J_0*, i.e.,*

$$J^*(x, \alpha) = J_0(x, \alpha) = J(\pi^*, x, \alpha) \quad \forall (x, \alpha) \in X. \tag{16}$$

Proof Let $\pi = \{\pi_t\}$ be an arbitrary policy, and let $C_t(\pi, x, \alpha)$ be the corresponding expected total cost from time t to terminal time N, given the state $(x_t, \alpha_t) = (x, \alpha)$ at time t, i.e.,

$$C_t(\pi, x, \alpha) := E^{\pi}\left[\sum_{n=t}^{N-1} \mathrm{e}^{S_t - S_n} c(x_n, a_n) + \mathrm{e}^{S_t - S_N} c_N(x_N, \alpha_N)|x_t = x, \alpha_t = \alpha\right] \tag{17}$$

for $t = 0, \ldots, N-1$

$$C_N(\pi, x, \alpha) := E^{\pi}(c_N(x_N)|x_N = x, \alpha_N = \alpha) = c_N(x).$$

$C_t(\pi, x, \alpha)$ is called the 'cost-to-go' or cost from time t onwards when using the policy π and $(x_t, \alpha_t) = (x, \alpha)$. In particular note that, from (10) and (17)

$$J(\pi, x, \alpha) = C_0(\pi, x, \alpha) \tag{18}$$

To prove the theorem, we shall show that, for all $(x, \alpha) \in X$ and $t = 0, \ldots, N$,

$$C_t(\pi, x, \alpha) \geq J_t(x, \alpha) \tag{19}$$

with equality if $\pi = \pi^*$, i.e.,

$$C_t(\pi^*, x, \alpha) = J_t(x, \alpha). \tag{20}$$

In particular for $t = 0$,

$$J(\pi, x, \alpha) \geq J_0(x, \alpha) \quad \text{with} \quad J(\pi^*, x, \alpha) = J_0(x, \alpha) \quad \forall (x, \alpha),$$

which yields the desired conclusion (16), as $J(\pi, \cdot, \cdot) \geq J_0(\cdot, \cdot)$ for arbitrary π implies $J^*(\cdot, \cdot) \geq J_0(\cdot, \cdot)$.

The proof of (19) and (20) is by backward induction. Observe that (19) and (20) trivially hold for $t = N$, since, from (18) and (13),

$$C_N(\pi, x, \alpha) = J_N(x, \alpha) = c_N(x).$$

Let us now assume (the induction hypothesis) that for some $t = N-1, \ldots, 0$,

$$C_{t+1}(\pi, x, \alpha) \geq J_{t+1}(x, \alpha) \quad \forall (x, \alpha) \in X. \tag{21}$$

Then

$$C_t(\pi, x, \alpha) = E^{\pi}\left[\sum_{i=t}^{N-1} \mathrm{e}^{S_t - S_i} c(x_i, a_i) + \mathrm{e}^{S_t - S_N} c_N(x_N, \alpha_N) | x_t = x, \alpha_t = \alpha\right]$$
$$= \int_A \left[c(x, a) + \mathrm{e}^{-\alpha} \int_X C_{t+1}(\pi, y, \beta) Q(\mathrm{d}y, \mathrm{d}\beta | x, \alpha, a)\right] \pi_t(\mathrm{d}a | x, \alpha)$$

hence,

$$C_t(\pi, x, \alpha) \geq \min_{A(x,\alpha)} \left[c(x, a) + \mathrm{e}^{-\alpha} \int_X J_{t+1}(y, \beta) Q(\mathrm{d}y, \mathrm{d}\beta | x, \alpha, a)\right]$$
$$= J_t(x, \alpha).$$

This proves (19). On the other hand, if equality holds in (21) with $\pi = \pi^*$ so that $\pi_t(\cdot|h_t)$ is the Dirac measure concentrated at $f(x_t, \alpha_t)$, then equality holds throughout the previous calculations which yields (20). □

Remark 1 (a) The nonstationary case can be reduced to former case with an adequate extension of spaces (see, for instance, Hinderer 1970, p. 78).
(b) We can see that when $\beta = \mathrm{e}^{-\alpha}$ is constant, we have the classical dynamic programming equation for fixed discounted factor (see, for instance, Hernández-Lerma and Lasserre 1996, pp. 32 and 33; Borkar 1991, Sect. III.2; Piunovskiy 1997, Subsect. 1.2.2.2).
(c) If we consider that observed interest rate in stage n is used to calculate the present value of current cost in stage n. Then the corresponding performance criterion would be

$$J^1(\pi, x, \alpha) := E^{\pi}_{(x,\alpha)}\left[\sum_{t=0}^{N-1} \mathrm{e}^{-S_{t+1}} c(x_t, a_t) + \mathrm{e}^{-S_N} c_N(x_N)\right]. \tag{22}$$

In this case (14) becomes

$$J^1_t(x, \alpha) := \min_{A(x,\alpha)} \left[\mathrm{e}^{-\alpha}\left(c(x, a) + \int_X J^1_{t+1}(y, \beta) Q(\mathrm{d}(y, \beta) | x, \alpha, a)\right)\right], \tag{23}$$

and (13) remains unchanged.

4 The measurable selection condition

In Theorem 2 we supposed the existence of measurable selectors, that satisfy the Eqs. (10)–(15). In this section we give conditions on the MCM (1) that assure the existence of selectors.

Assumption 3 For a given measurable function $u : X \to \mathbb{R}$, the function u^* from X to $\mathbb{R}$ defined,

$$u^*(x,\alpha) := \inf_{A(x,\alpha)} \left[c(x,a) + \mathrm{e}^{-\alpha} \int_X u(y,\beta) Q(\mathrm{d}y, \mathrm{d}\beta | x, \alpha, a) \right] \tag{24}$$

is measurable and there exists a selector $f \in \mathbf{F}$ such that the function within brackets attains its minimum at $f(x,\alpha) \in A(x,\alpha)$ for all $(x,\alpha) \in X$, i.e.,

$$u^*(x,\alpha) = c(x,f) + \mathrm{e}^{-\alpha} \int_X u(y,\beta) Q(\mathrm{d}y, \mathrm{d}\beta | x, \alpha, f) \quad \forall\, (x,\alpha) \in X.$$

We recall some definitions that will be used in the conditions below.

Let Y be a metric space and v a function from Y to $\mathbb{R} \cup \{\infty\}$ such that $v(y) < \infty$ for at least one point $y \in Y$. The function v is said to be *lower semicontinuous* (l.s.c.) at $y \in Y$, if $\liminf v(y_n) \geq v(y)$ for any sequence $\{y_n\}$ in Y that converges to y. The function v is called lower semicontinuous (l.s.c.) if it is l.s.c. at every point of Y. A function $v : \mathbb{K} \to \mathbb{R}$ is said to be *inf-compact* on $\mathbb{K}$ if, for every $(x,\alpha) \in X$ and $r \in \mathbb{R}$, the set $\{a \in A(x,\alpha) | v(x,\alpha,a) \leq r\}$ is compact. A multifunction ψ from X to A is said to be *upper semicontinuous* (u.s.c) if $\psi^{-1}[F]$ is closed in X for every closed set $F \subset A$. Let $\mathbb{B}(X)$ be the family of measurable bounded functions on X, and $\mathbb{C}(X) \subset \mathbb{B}(\mathbb{X})$ the subfamily of continuous functions.

We now consider the three conditions under which, in particular, Assumption 4.1 is satisfied

Condition 4 (a) *The control sets $A(x,\alpha)$ are compact for all $(x,\alpha) \in X$.*
(b) *The one-stage cost c is such that $c(x,\cdot)$ is l.s.c on $A(x,\alpha)$ for every $(x,\alpha) \in X$.*
(c) *The function*

$$v'(x,\alpha,a) := \int_X v(y,\beta) Q(\mathrm{d}y, \mathrm{d}\beta | x, \alpha, a) \tag{25}$$

on $\mathbb{K}$ satisfies one of the two following conditions:
(c_1) *$v'(x,\alpha,\cdot)$ is l.s.c. on $A(x,\alpha)$ for every $(x,\alpha) \in X$ and every $v \in \mathbb{C}(X)$,*
(c_2) *$v'(x,\alpha,\cdot)$ is l.s.c. on $A(x,\alpha)$ for every $(x,\alpha) \in X$ and every $v \in \mathbb{B}(X)$.*

Condition 5 (a) *$A(x,\alpha)$ is compact for all $(x,\alpha) \in X$ and the multifunction $(x,\alpha) \mapsto A(x,\alpha)$ is u.s.c.*
(b) *The one-stage cost c is l.s.c and bounded below.*
(c) *The transition law Q is either:*
(c_1) *weakly continuous, i.e., for every function $v \in \mathbb{C}(X)$, the function v' in* (25) *is continuous and bounded on $\mathbb{K}$*
(c_2) *strongly continuous, i.e., v' is continuous and bounded on $\mathbb{K}$ for every $v \in \mathbb{B}(X)$.*

Condition 6 (a) *The one-stage cost c is l.s.c bounded below and inf-compact on $\mathbb{K}$;*
(b) *Same as* 5(c)*, i.e., Q is either*

(b_1) *weakly continuous, or*
(b_2) *strongly continuous.*

We next show how the last three conditions relate to Assumption 3.

Theorem 7 (i) Each of Conditions 4 and 5 implies Assumption 3 for any nonnegative measurable function $u : X \mapsto \mathbb{R}$. Moreover, under 4(c_1) or 5(c_1), it suffices to take u nonnegative and $l.s.c.$ in which case, under 5(a,b,c_1) the function u^* in (24) is $l.s.c.$
(ii) Condition 6 implies Assumption 3 if, under b_1, u is nonnegative and $l.s.c$ or, under b_2, if u is a nonnegative measurable function. If in addition the multifunction $(x, \alpha) \mapsto A^*(x, \alpha)$ with $A^*(x, \alpha)$ equal to

$$\left\{a \in A(x, \alpha) | u * (x, \alpha) := c(x, a) + \mathrm{e}^{-\alpha} \int u(y, \beta) Q(\mathrm{d}y,\ \mathrm{d}\beta | x, \alpha, a)\right\}$$

is $l.s.c.$, then u^* is $l.s.c.$.

Remark 2 In Theorem 7, we suppose that u is nonnegative, but it is easily seen that it suffices to take a bounded below.

Proof (i) Let $u \geq 0$ be a measurable function on X.

To prove the first statement in (i), it clearly suffices to consider Conditions 4 (a), (b) and (c_2). Moreover, note that given *l.s.c.* functions $v_1,\ v_2$, then $v_1 + \mathrm{e}^{-\alpha} v_2$ is also *l.s.c.*. Hence the desired conclusion follows from Proposition D.5 in Hernández-Lerma and Lasserre (1996) provided that the function

$$a \mapsto \int_X u(y, \beta) Q(\mathrm{d}y, \mathrm{d}\beta \mid x, \alpha, a) \text{ is } l.s.c. \text{ on } A(x, \alpha) \text{ for every } (x, \alpha) \in X. \tag{26}$$

To prove this, let $\{u^n\}$ be a sequence in $\mathbb{B}(X)$ such that $u^n \uparrow u$, and let $\{a^l\}$ be a sequence in $A(x, \alpha)$. Converging to $a \in A(x, \alpha)$. Then, for each n we have

$$\begin{aligned}\liminf_{l\to\infty} \int u(y, \beta) Q(\mathrm{d}y, \mathrm{d}\beta | x, \alpha, a^l) &\geq \liminf_{l\to\infty} \int u^n(y, \beta) Q(\mathrm{d}y, \mathrm{d}\beta | x, \alpha, a^l) \\ &\geq \int u^n(y, \beta) Q(\mathrm{d}y, \mathrm{d}\beta | x, \alpha, a).\end{aligned}$$

Letting n tend to infinity we obtain (by the Monotone Convergence Theorem)

$$\liminf_{l\to\infty} \int u(y, \beta) Q(\mathrm{d}y, \mathrm{d}\beta | x, \alpha, a^l) \geq \int u(y, \beta) Q(\mathrm{d}y, \mathrm{d}\beta | x, \alpha)$$

which proves (26). Thus, as was already mentioned, we obtain Assumption 3 from Proposition D.5 in Hernández-Lerma and Lasserre (1996).

Let us now suppose that Conditions 4(c_1) or 5(c_1) hold. Then the second statement in (i) follows from the same argument above, but now based on the fact that $u \geq 0$ is l.s.c., then it is the limit of an increasing sequence in $\mathbb{C}$(X) (see Proposition A.2 in Hernández-Lerma and Lasserre 1996).

The last statement in (i) follows from the above arguments and Proposition D.5(b) in Hernández-Lerma and Lasserre (1996).

(ii) Suppose that Condition 6 holds with (b_2), and that $u \geq 0$ is measurable. Then, as in the proof of part (i), but now approximating u from below by functions in $\mathbb{B}$(X), one can show that the function

$$u'(x,\alpha,a) = c(x,a) + e^{-\alpha} \int_X u(y,\beta) Q(dy, d\beta \mid x,\alpha,a)\}, \quad (x,\alpha,a) \in \mathbb{K}$$

is l.s.c and bounded below. Thus we may obtain Assumption 3 from Proposition D.6(a) in Hernández-Lerma and Lasserre (1996), if u' is intf-compact on $\mathbb{K}$, that is, if for every $(x,\alpha) \in X$ and $r \in \mathbb{R}$, the set $\{a \in A(x,\alpha) | u'(x,a) \leq r\} := D$ is compact. But this is obviously true, since (by lower semicontinuity – see Proposition A.1(c) in Hernández-Lerma and Lasserre 1996) D is closed and, since $u \geq 0$, it is contained in the set $\{a \in A(x,\alpha) | c(x,a) \leq r\}$, which, by the inf-compactness of c [see Condition 6(a)], is compact. The proof under (b_1) is similar.

The last statement in (ii) follows from Proposition D.6(b) in Hernández-Lerma and Lasserre (1996). □

5 Infinite-horizon cost problem

Now we study MDPs with random discounted cost and infinite-horizon. Given a stationary control model as (1) and the performance criterion to be minimized is

$$V(\pi,x,\alpha) := E^{\pi}_{(x,\alpha)} \left[\sum_{t=0}^{\infty} e^{-S_t} c(x_t,a_t) \right], \quad \pi \in \Pi, (x,\alpha) \in X \tag{27}$$

where S_t as in Sect. 3. A policy $\pi*$ satisfying

$$V(\pi^*,x,\alpha) = \inf_{\pi} V(\pi,x,\alpha) =: V^*(x,\alpha) \quad \forall (x,\alpha) \in X \tag{28}$$

is said to be *optimal* and V^* is called the *value function*.

Throughout the following, we suppose that the one-stage cost c is nonnegative (although, in fact, for virtually all of the results to be true it suffices to assume that c is bounded below). Moreover, we will use V_n to denote the n-stage cost

$$V_n(\pi,x,\alpha) := E^{\pi}_{(x,\alpha)} \left[\sum_{t=0}^{n-1} e^{-S_t} c(x_t,a_t) \right]. \tag{29}$$

Hence (by the Monotone Convergence Theorem) we may write $V(\pi,x,\alpha)$ in (27) as

$$V(\pi,x,\alpha) = \lim_{n\to\infty} V_n(\pi,x,\alpha). \tag{30}$$

A measurable function $v : X \to \mathbb{R}$ is said to be a *solution* of *optimality equation* (OE) if it satisfies

$$V(x,\alpha) = \min_{A(x,\alpha)} \left\{ c(x,\alpha) + \mathrm{e}^{-\alpha} \int_X V(y,\beta) Q(\mathrm{d}y, \mathrm{d}\beta | x, \alpha, a) \right\} \quad \forall (x,\alpha) \in X. \tag{31}$$

In Theorem 10, we prove that the value function V^* in (28) is solution to the OE. To this end, we begin with the DP Theorem 2 for *finite*-horizon problems and with suitable change of indices we obtain the *forward form of the dynamic programming algorithm*, that is, the value iteration functions defined as

$$v_n(x,\alpha) = \min_{A(x,\alpha)} \left\{ c(x,a) + \mathrm{e}^{-\alpha} \int_X v_{n-1}(y,\beta) Q(\mathrm{d}y, \mathrm{d}\beta | x, \alpha, a) \right\} \quad \forall (x,\alpha) \in X \tag{32}$$

and $n = 1, 2, \ldots$, with $v_0(\cdot) := 0$. The idea then is to show that

$$V^*(x,\alpha) = \lim_{n\to\infty} v_n(x,\alpha) \quad \forall (x,\alpha) \in X. \tag{33}$$

This result is to be expected since v_n is the value function of the n-stage cost V_n in (29) with zero terminal cost, namely

$$v_n(x,\alpha) = \inf_{\pi} V_n(\pi, x, \alpha), \quad (x,\alpha) \in X. \tag{34}$$

This, letting $n \to \infty$ in (32) we anticipate to obtain (35), if we can justify the interchange of limits and minima.

This approach, requires first of all, the measurable selection condition in Assumption 3 for (32) and (35) to be well defined. We also impose the follow requirements.

Assumption 8 (a) The one-stage cost c is l.s.c., nonnegative, and inf-compact on $\mathbb{K}$.
(b) Q is strongly continuous.

Assumption 9 There exists a policy π such that $V(\pi, x, \alpha) < \infty$ for each $(x,\alpha) \in X$.

We shall denote by Π^0 the family of policies for which Assumption 9 holds. We now state our main result in this section.

Theorem 10 Suppose that Assumptions 8 and 9 hold. Then

(a) The value function V^* is the minimal solution to the OE, i.e.,

$$V^*(x,\alpha) = \min_{A(x,\alpha)} \left\{ c(x,a) + \mathrm{e}^{-\alpha} \int_X V^*(y,\beta) Q(\mathrm{d}y, \mathrm{d}\beta | x, \alpha, a) \right\} \tag{35}$$

for all $(x,\alpha) \in X$ and if u is another solution to the OE, then $u(\cdot) \geq V^*(\cdot)$.

(b) There exists a selector $f_* \in \mathbb{F}$ such that $f_*(x,\alpha) \in A(x,\alpha)$ attains the minimum in (35), i.e.,

$$V^*(x,\alpha) = c(x, f_*) + e^{-\alpha} \int V^*(y,\beta) Q(dy, d\beta | x, \alpha, f_*) \quad \forall (x,\alpha) \in X \tag{36}$$

and the deterministic stationary policy f_*^∞ is optimal. Conversely, if f_*^∞ is a stationary deterministic optimal policy, then it satisfies (36).

(c) If π^* is a policy such that $V(\pi^*, \cdot)$ is a solution to the OE and satisfies

$$\lim_{n\to\infty} E^{\pi}_{(x,\alpha)} \left[e^{-S_n} V(\pi^*, x_n, \alpha_n) \right] = 0 \quad \forall \pi \in \Pi^0 \text{ and } (x,\alpha) \in X, \tag{37}$$

then $V(\pi^*, \cdot) = V^*(\cdot)$, and so π^* is α-discounted optimal. In other words, if (37) holds, then π^* is optimal if and only if $V(\pi^*, \cdot)$ satisfies the OE.

(d) If an optimal policy exists, then there exists one that is deterministic stationary.

The proof of this theorem requires several lemmas.

Lemma 1 *Let u and u_n $(n = 1, 2, \ldots)$ be l.s.c. functions, bounded below, and inf-compact on $\mathbb{K}$. If $u_n \uparrow u$. Then*

$$\lim_{n\to\infty} \min_{A(x,\alpha)} u_n(x,\alpha,a) = \min_{A(x,\alpha)} u(x,\alpha,a) \quad \forall (x,\alpha) \in X. \tag{38}$$

Proof The proof is similar to that of Lemma 4.2.4 in Hernández-Lerma and Lasserre (1996, p.47), and, therefore is omitted. □

We need also in this case the existence of measurable selectors that satisfy the DP equation. To do this we use Theorem 7 and the following definition.

Definition 2 *$M(X)^+$ denotes the cone of nonnegative measurable functions on X, and, for every $u \in M(X)^+$, Tu is the function on X defined as*

$$Tu(x,\alpha) := \min_{A(x,\alpha)} \left[c(x,a) + e^{-\alpha} \int_X u(y,\beta) Q(dy, d\beta | x, \alpha, a) \right]. \tag{39}$$

Lemma 2 *Under Assumption* 8, *T maps $M(X)^+$ into itself, i.e., for every u in $M(X)^+$, Tu is also in $M(X)^+$, and moreover, there exists a selector $f \in \mathbb{F}$ such that*

$$Tu(x,\alpha) = c(x, f) + e^{-\alpha} \int_X u(y,\beta) Q(dy, d\beta | x, \alpha, f) \quad \forall (x,\alpha) \in X.$$

Notice also that, using the operator T, we may rewrite the OE (35) *and the functions in* (32) *as*

$$V^* = TV^* \quad \textit{and} \quad v_n = Tv_{n-1} \quad \textit{for} \quad n \geq 1$$

$v_0 = 0$, respectively. We shall next relate V^* to the functions u that satisfy $u \geq Tu$ or $u \leq Tu$.

Lemma 3 *Suppose that Assumptions* 8 *and* 9 *hold:*

(a) *If $u \in M(X)^+$ is such that $u \geq Tu$, then $u \geq V^*$.*
(b) *If $u : X \to \mathbb{R}$ is a measurable function such that Tu is well defined and, in addition, $u \leq Tu$ and*

$$\lim_{n\to\infty} E^{\pi}_{(x,\alpha)}\left[e^{-S_n}u(x_n,\alpha_n)\right] = 0 \quad \forall\pi \in \Pi^0 \quad \textit{and} \quad (x,\alpha) \in X \tag{40}$$

then $u \leq V^*$.

Proof (a) Let $u \in M^+(X)$ such that $u \geq Tu$, then, by Lemma 2, we may choose $f \in \mathbb{F}$ such that

$$u(x,\alpha) \geq c(x,f) + e^{-\alpha}\int_X u(y,\beta)Q(dy,d\beta|x,\alpha,f).$$

Iterations of this inequality give us

$$u(x,\alpha) \geq E^{\pi}_{(x,\alpha)}\left[\sum_{t=0}^{N-1} e^{-S_t}c(x_t,a_t)\right] + E^{\pi}_{(x,\alpha)}\left[e^{-S_N}u(x_N,a_N)\right], \tag{41}$$

where $\pi = (f, f, \ldots) = f^{\infty}$ and $E^{\pi}_{(x,\alpha)}\left[e^{-S_N}u(x_N,\alpha_N)\right] = \int u(y,\beta)\, Q^n(dy,d\beta| x,\alpha,f)$. Since $u \geq 0$, we have that

$$u(x,\alpha) \geq E^{\pi}_{(x,\alpha)}\left[\sum_{t=0}^{N-1} e^{-S_t}c(x_t,a_t)\right].$$

Letting $N \to \infty$, we get

$$u(x,\alpha) \geq V(\pi,x,\alpha) \geq V^*(x,\alpha) \quad \forall(x,\alpha) \in X.$$

This proves (a).
(b) Let $\pi \in \Pi$ and $(x,\alpha) \in X$ be arbitrary. From the Markov-like property (2.9) and the assumption $Tu \geq u$,

$$\begin{aligned}
&E^{\pi}_{(x,\alpha)}\left[e^{-S_{t+1}}u(x_{t+1},\alpha_{t+1})|h_t,a_t\right]\\
&= e^{-S_{t+1}}E^{\pi}_{(x,\alpha)}\left[u(x_{t+1},\alpha_{t+1})|x_t,\alpha_t,a_t\right]\\
&= e^{-S_{t+1}}\left[\int_X u(y,\beta)Q(dy,d\beta|x_t,\alpha_t,a_t)\right]\\
&= e^{-S_t}\left[c(x_t,a_t) + e^{-\alpha_t}\int_X u(y,\beta)Q(dy,d\beta|x_t,\alpha_t,a_t) - c(x_t,a_t)\right]\\
&\geq e^{-S_t}[u(x_t,\alpha_t) - c(x_t,a_t)].
\end{aligned}$$

Hence

$$e^{-S_t}c(x_t,a_t) \geq -E^{\pi}_{(x,\alpha)}\left[e^{-S_{t+1}}u(x_{t+1},\alpha_{t+1})|h_t,a_t\right] + e^{-S_t}u(x_t,\alpha_t)$$

Thus, taking expectations $E^{\pi}_{(x,\alpha)}$ and summing over $t = 0, \ldots, N-1$, we have

$$\begin{aligned} &E^{\pi}_{(x,\alpha)}\left[\sum_{t=0}^{N-1} \mathrm{e}^{-S_t} c(x_t, a_t)\right] \\ &\geq \sum_{t=0}^{N-1} \left(E^{\pi}_{(x,\alpha)}\big[-\mathrm{e}^{-S_t} u(x_t, \alpha_t) + \mathrm{e}^{-S_{t-1}} u(x_{t-1}, \alpha_{t-1})\big]\right) \\ &= -E^{\pi}_{(x,\alpha)}\big[-\mathrm{e}^{-S_N} u(x_N, \alpha_N)\big] + u(x, \alpha) \quad \forall N. \end{aligned}$$

Finally, letting $N \to \infty$ and using the hypothesis (40), it follows that $V(\pi, x, \alpha) \geq u(x)$, which implies $V^* \geq u$, as π and (x, α) were arbitrary. □

We shall now use Lemmas 1 and 3 to prove the limit (33).

Lemma 4 *Suppose that Assumptions* 8 *and* 9 *hold. Then* $v_n \uparrow V^*$ *and* V^* *satisfies the OE.*

Proof To begin, note that, from (35), (27) and the assumption that $c \geq 0$,

$$v_n(x, \alpha) \leq V_n(\pi, x, \alpha) \leq V(\pi, x, \alpha) \quad \forall n, \pi, (x, \alpha).$$

Therefore,

$$v_n(x, \alpha) \leq V^*(x, \alpha) \quad \forall n, (x, \alpha) \in X. \tag{42}$$

Now, the operator T in (39) is monotone. Therefore, since $v_0 := 0$ and $v_n := T v_{n-1}$ for $n \geq 1$, the functions form a nondecreasing sequence in $M(X)^+$, which implies that $v_n \uparrow v^*$ for some $v^* \in M(X)^+$. This, in turn (by the Monotone Convergence Theorem), implies $u_n \uparrow u$, where

$$u_n(x, \alpha) = c(x, a) + \mathrm{e}^{-\alpha} \int_X v_n(y, \beta) Q(\mathrm{d}y, \mathrm{d}\beta | x, \alpha, a),$$

$$u(x, \alpha) = c(x, a) + \mathrm{e}^{-\alpha} \int_X v^*(y, \beta) Q(\mathrm{d}y, \mathrm{d}\beta | x, \alpha, a).$$

On the other hand, as in the proof of Theorem 7(ii), one can show that the non-negative functions u and u_n $(n \geq 1)$ are l.s.c. and inf-compact on $\mathbb{K}$. Thus, from Lemma 1,

$$v^* = \lim_n v_n = \lim_n T v_{n-1} = T v^*;$$

that is, v^* satisfies the OE $v^* = T v^*$.

Hence, to complete the proof of the lemma, it only remains to show that $v^* = V^*$. But this is immediate because, by Lemma 3(a), $v^* = T v^*$ implies $v^* \geq V^*$, and the reverse inequality follows from (42) and the already established fact that $v_n \uparrow v^*$. □

Finally, we prove Theorem 10.

Proof of Theorem 10 (a) From Lemma 4, V^* is a solution to the OE, and the fact that V^* is the minimal solution follows from Lemma 3(a) – for if $u = Tu$, then $u \geq V^*$.

(b) The existence of a selector $f_* \in \mathbb{F}$ satisfying (36) is ensured by Lemma 2. Now, iteration of (36) shows [as in (41)] that

$$\begin{aligned} V^* &= E_{(x,\alpha)}^{f_*^\infty}\left[\sum_{t=0}^{n-1} \mathrm{e}^{-S_t} c(x_t, f_*)\right] + E_{(x,\alpha)}^{f_*^\infty}\left[\mathrm{e}^{-S_n} V^*(x_n, \alpha_n)\right] \\ &\geq E_{(x,\alpha)}^{f_*^\infty}\left[\sum_{t=0}^{n-1} \mathrm{e}^{-S_t} c(x_t, f_*)\right] \quad n \geq 1 \quad \forall (x, \alpha) \in X. \end{aligned}$$

This implies, letting $n \to \infty$, $V^*(x, \alpha) \geq V(f_*^\infty, x, \alpha)$ $\forall (x, \alpha) \in X$, whereas, from (28), $V^*(\cdot) \leq V(f_*^\infty, \cdot)$. That is, $V^*(\cdot) = V(f_*^\infty, \cdot)$ and, therefore, f_*^∞ is optimal.

To prove the converse, we note first the important fact that for any deterministic stationary policy f^∞, the cost $V(f^\infty, \cdot)$ satisfies

$$V(f^\infty, x, \alpha) = c(x, f) + \mathrm{e}^{-\alpha} \int_X V(f^\infty, y, \beta) Q(\mathrm{d}y, \mathrm{d}\beta | x, \alpha, f) \quad \forall (x, \alpha) \in X. \tag{43}$$

Indeed,

$$\begin{aligned} V(f^\infty, x, \alpha) &= E_{(x,\alpha)}^{f^\infty}\left[\sum_{t=0}^{\infty} \mathrm{e}^{-S_t} c(x_t, f)\right] \\ &= E_{(x,\alpha)}^{f^\infty}\left[c(x_0, f) + \sum_{t=1}^{\infty} \mathrm{e}^{-S_t} c(x_t, f)\right] \\ &= c(x, f) + \mathrm{e}^{-\alpha} E_{(x,\alpha)}^{f^\infty}\left[\sum_{t=1}^{\infty} \mathrm{e}^{-S_{t-1}} c(x_t, f)\right] \\ &= c(x, f) + \mathrm{e}^{-\alpha} E_{(x,\alpha)}^{f^\infty} E_{(x_1,\alpha_1)}^{f^\infty}\left[\sum_{t=1}^{\infty} \mathrm{e}^{-S_{t-1}} c(x_t, f)\right] \\ &= c(x, f) + \mathrm{e}^{-\alpha} E_{(x,\alpha)}^{f^\infty}\left[V(f^\infty, x, \alpha)\right] \\ &= c(x, f) + \mathrm{e}^{-\alpha} \int_X V(f^\infty, y, \beta) Q(\mathrm{d}y, \mathrm{d}\beta | x, \alpha, f). \end{aligned}$$

In particular, if f_* is stationary deterministic optimal, then $V(f_*^\infty, \cdot) = V^*(\cdot)$, in which case (43), with $f = f_*$, reduces to (36).

(c) If $V(\pi^*,\cdot)$ satisfies the OE, then from part (a) or Lemma 3(a) we get $V(\pi^*,\cdot)\geq V^*(\cdot)$. The reverse inequality follows from (37) and Lemma 3(b).

Finally, (d) is a consequence of (a) and (b). □

6 An example

In this section we consider a consumption–investment problem, consisting of an investor who wishes to allocate his/her current wealth x_t between investment (a_t) and consumption ($x_t - a_t$), in each period $t = 0, 1, \ldots, N$. Since we assume that borrowing is not allowed, the investment (or control) constraint set is $A(x,\alpha) = [0, x]$. The connection between investment decisions and accumulated capital is given by

$$\begin{aligned} x_{t+1} &= a_t\xi_t, \\ \alpha_{t+1} &= d\alpha_t + \eta_t, \quad t = 0, 1, \ldots, N, \end{aligned}$$

where $\{\xi_t\}$ and $\{\eta_t\}$ are independent sequences of i.i.d. random variables, and independent of the initial state (x_0, α_0). Let $E(\xi_0) = \mathrm{e}^m$ with $m > 0$, and $0 < d < 1$; the one-stage return $r(x, a)$ is supposed to be a *utility consumption*, say $r(x, a) := b(x-a)$ with $b > 0$, and we wish to maximize the expected total discounted utility.

We suppose that parameters d, m, α and expected value of $E\big(\mathrm{e}^{-\frac{\eta_0}{1-d}}\big)$ satisfy

$$E\big(\mathrm{e}^{-\frac{\eta_0}{1-d}}\big) \geq 1, \quad \alpha_0 + \frac{M}{1-d} \leq m \tag{44}$$

where $|\eta_0| \leq M$.

With obvious changes of min by max in the DP algorithm (13), (14) and $r_{N+1}(x) = 0$ we have

$$J_N(x,\alpha) = \max_{a\in[0,x]} \Big\{ b(x-a) + \mathrm{e}^{-\alpha} E\big(J_{N+1}(y)\big) \Big\} = bx,$$

and $f_N(x,\alpha) = 0$. Similarly

$$\begin{aligned} J_{N-1}(x,\alpha) &= \max_{a\in[0,x]} \big\{ b(x-a) + \mathrm{e}^{-\alpha} E\big(J_N(y,\beta)\big)\big\} \\ &= b \max_{a\in[0,x]} \big\{ x + a\big(-1 + \mathrm{e}^{m-\alpha}\big)\big\} \\ &= b e^{m-\alpha} x \end{aligned}$$

where the last equality is consequence of (44). Indeed, since

$$|\alpha| \leq |d^{N-2}\alpha_0 + \sum_{i=0}^{N-2} d^i \eta_i| \leq \alpha_0 + \frac{M}{1-d} \leq m.$$

we obtain, $\mathrm{e}^{m-\alpha} \geq 1$. And $f_{N-1}(x,\alpha) = x$. In general

$$J_{N-t}(x,\alpha) = \max_{a\in[0,x]} \big\{ b(x-a) + \mathrm{e}^{-\alpha} E\big(J_{N-t+1}(y,\beta)\big)\big\}$$

for each $t = 2, \ldots, N$.

The inequalities (44) yield that $m \geq \alpha$, and

$$m + \ln E(\mathrm{e}^{-\eta_0}) \geq \mathrm{d}\alpha, \ldots, m + \ln E\left(\mathrm{e}^{-(1+\cdots+d^{t-2})\eta_0}\right) \geq d^{t-1}\alpha.$$

Hence

$$\mathrm{e}^{mt-(1+d+\cdots+d^{t-1})\alpha} E(\mathrm{e}^{-\eta_0}) E\left(\mathrm{e}^{-(1+d)\eta_0}\right) \cdots E\left(\mathrm{e}^{-(1+d+\cdots+d^{t-2})\eta_0}\right) - 1 \geq 0.$$

Therefore

$$J_{N-t}(x, \alpha) = b\mathrm{e}^{mt-(1+\cdots+d^{t-1})\alpha} E(\mathrm{e}^{-\eta_0}) E\left(\mathrm{e}^{-(1+d)\eta_0}\right) \cdots E\left(\mathrm{e}^{-(1+\cdots+d^{t-2})\eta_0}\right) x$$

with $f_{N-t}(x, \alpha) = x$. Continuing this process we obtain the optimal value

$$J_0(x, \alpha) = b\mathrm{e}^{mN-(1+\cdots+d^{N-1})\alpha} E(\mathrm{e}^{-\eta_0}) E\left(\mathrm{e}^{-(1+d)\eta_0}\right) \cdots E\left(\mathrm{e}^{-(1+\cdots+d^{N-2})\eta_0}\right) x$$

with $f_0(x, \alpha) = x$. Hence the optimal policy is to invest the whole wealth in each stage $0 \leq t \leq N$ except at the last stage, when everything is consumed.

References

Abbad M, Daoui C (2003) Hierarchical algorithms for discounted Markov decision processes. Math Methods Oper Res 58:237–245

Altman E (1999) Constrained Markov decision processes. Chapman & Hall/CRC, Boca Raton

Bertsekas DP (1995) Dynamic programming and optimal control. Athena Scientific, Belmont

Berument H, Kilinc Z, Ozlale U (2004) The effects of different inflation risk prepius on interest rate spreads. Physica A 333:317–324

Borkar VS (1991) Topics in controlled Markov chains. Longman, New York

Cai J, Dickson D, David CM (2004) Ruin probabilities with a Markov chain interest model. Insur Math Econ 35:513–525

Feinberg EA, Shwartz A (1994) Markov decision models with weighted discounted criteria. Math Oper Res 19:152–168

Feinberg EA, Shwartz A (1995) Constrained Markov decision models with weighted discounted rewards. Math Oper Res 20:302–320

Feinberg EA, Shwartz A (1999) Constrained dynamic programming with two discount factors: applications and an algorithm. IEEE Trans Autom Control 42:628–631

Gil A, Luis A (2004) Modelling the U.S. interest rate in terms of I(d) statistical models. Q Rev Econ Finance 44:475–486

Haberman S, Sung J (2005) Optimal pension funding dynamics over infinite control horizon when stochastic rates of return are stationary. Insur Math Econ 36:103–116

Hernández-Lerma O, González-Hernández J (2000) Constrained Markov control processes in Borel spaces: the discounted case. Math Meth Oper Res, 52:271–285

Hernández-Lerma O, Lasserre JB (1996) Discrete-time Markov control processes: basic optimality criteria. Springer, Berlin Heidelberg New York

Hinderer K (1970) Foundations of non-stationary dynamical programming with discrete time parameter. Lecture notes in operation research, vol 33. Springer, Berlin Heidelberg New York

Hu Q (2003) Analysis for some properties of discrete time Markov decision processes. Optimization 52:495–505

Kurano M, Song J, Hosaka M, Huang Y (1998) Controlled Markov set-chains with discounting. J Appl Prob 35:293–302

Lee P, Rosenfield DB (2005) When to refinance a mortgage: a dynamic programming approach. Eur J Oper Res 166:266–277

Lettau M, Uhlig H (1999) Rules of thumb versus dynamic programming. Am Econ Rev 89(1):148–174

Liu K (1999) Weighted discounted Markov decision processes with perturbation. Acta Math Appl Sinica 15:183–189
López-Martínez RR, Hernández-Lerma O (2003) The Lagrange approach to constrained Markov processes: a survey and extension of results. Morfismos 7:1–26
Newell RG, Pizer WA (2003) Discounting the distant future: how much do uncertain rates increase valuation?. J Environ Econ Manage 46:52–71
Mao X, Piunovskiy AB (2000) Strategic measure in optimal control problems for stochastic sequences. Stoch Anal Appl 18:755–776
Michael K Ng (1999) A note on policy algorithms for discounted Markov decision problems. Oper Res Lett 25:195–197
Ogaki M, Santaella JA (2000) The exchange rate and the term structure of interest rates in México. J Dev Econ 63:135–155
Piunovskiy AB (1997) Optimal control of random sequences in problems with constraints. Kluwer, Boston
Puterman ML (1994) Markov decision processes: discrete stochastic dynamic programming. Wiley, New York
Sack B, Wieland V (2000) Interest-rate smoothing and optimal monetary policy: a review of recent empirical evidence. J Econ Bus 52:205–228
Shwartz A (2001) Death and discounting. IEEE Trans Autom Control 46:644–647
Stockey NL, Lucas RE Jr (1989) Recursive methods in economic dynamics. Harvard University Press, Cambridge