J Optim Theory Appl (2014) 162:329–346
DOI 10.1007/s10957-012-0262-8

Markov Decision Processes on Borel Spaces with Total Cost and Random Horizon

Hugo Cruz-Suárez · Rocio Ilhuicatzi-Roldán · Raúl Montes-de-Oca

Received: 27 June 2012 / Accepted: 25 December 2012 / Published online: 24 January 2013

Abstract This paper deals with Markov Decision Processes (MDPs) on Borel spaces with possibly unbounded costs. The criterion to be optimized is the expected total cost with a random horizon of infinite support. In this paper, it is observed that this performance criterion is equivalent to the expected total discounted cost with an infinite horizon and a varying-time discount factor. Then, the optimal value function and the optimal policy are characterized through some suitable versions of the Dynamic Programming Equation. Moreover, it is proved that the optimal value function of the optimal control problem with a random horizon can be bounded from above by the optimal value function of a discounted optimal control problem with a fixed discount factor. In this case, the discount factor is defined in an adequate way by the parameters introduced for the study of the optimal control problem with a random horizon. To illustrate the theory developed, a version of the Linear-Quadratic model with a random horizon and a Logarithm Consumption-Investment model are presented.

Keywords Markov decision process · Total cost · Random horizon · Varying-time discount factor

H. Cruz-Suárez (✉) · R. Ilhuicatzi-Roldán
Facultad de Ciencias Físico-Matemáticas, Benemérita Universidad Autónoma de Puebla, Av. San Claudio y 18 Sur, Puebla, Mexico
e-mail: hcs@fcfm.buap.mx

R. Ilhuicatzi-Roldán
e-mail: rroldan@alumnos.fcfm.buap.mx

R. Montes-de-Oca
Departamento de Matemáticas, Universidad Autónoma Metropolitana-Iztapalapa, Av. San Rafael Atlixco 186, Col. Vicentina, Mexico D.F. 09340, Mexico
e-mail: momr@xanum.uam.mx

1 Introduction

This paper was initially motivated by the study of the discounted optimal control problem given in Puterman's book (see [1], Sect. 5.3). In this reference, it is proved that a discounted control problem can be treated as a control problem where the horizon is a random variable, which is supposed to follow a geometric distribution independent of the process.

Retaking the idea mentioned, this article presents the study of the problem with a random horizon independent of the process, considering in this case an arbitrary distribution. While researching the background of this problem, it was observed that it is natural for applications to have this kind of independence, for example, to analyze a bankruptcy in economic models, to model a failure of a system in economic models, to model a failure of a system in engineering, or to study the extinction of some natural resource in biology (see [1] p. 125, [2] p. 267, [3, 4]).

Another approach to study this kind of problems is assuming that the random horizon is measurable with respect to the filtration of the process. In this context, a class of problems is studied as an optimal stopping problem (see [1, 2, 5]). This problem has also been studied when the random horizon is the stopping time, see [6], for instance. However, there was not sufficient literature found for the case when the horizon is independent of the process, and the existing one presented limitations, which reinforce the intention to study the problem considering this independence.

The present work approaches the control problem in the cases where the support of the random horizon is either finite or infinite. In the case when the distribution has a finite support, the optimal control problem can be studied using the existing theory for MDPs (see [1, 3, 4, 7]). The same does not occur for the numerable case, on which this article is centered.

For the goal of this paper, firstly, the optimal control problem with the expected total cost and a random horizon as a performance criterion is presented. It is observed that this problem is equivalent to the one with the infinite-horizon expected total discounted cost as a performance criterion. The importance in this case is that the discount factor is varying over time (see (6), (7), and Remark 4.3(ii)). Usually, in the literature, the discount factor is a fixed number between 0 and 1 (see [8]). In the development presented in this work, even the case with a fixed discount factor (see Remark 4.3(iii)) is included.

Secondly, the optimal solution of the optimal control problem with a random horizon is characterized (the value function and the optimal policy). To do this, a standard dynamic programming approach is used (see [1, 8–11]). For this problem, it is assumed that the states and action spaces are Borel spaces, not necessarily compact, and with the cost function that is neither necessarily bounded. These situations have not been considered in previous papers (see [4, 7]). It is important to point out that in the optimal control problem with a random horizon there may not exist a stationary optimal policy (see [7]). In this case, using conditions of continuity and compactness on the components of the model, it is possible to guarantee that there exists a deterministic Markov optimal policy. This article also provides two fully worked examples.

Thirdly, in this paper it is proved that the optimal value function of the optimal control problem with a random horizon can be bounded from above by the optimal value

function of a discounted optimal control problem with a fixed discount factor. In this case, the discount factor is defined in an adequate way by the parameters introduced for the study of the optimal control problem with a random horizon. The previous description suggests that a problem of a random horizon can be approximated using a discounted optimal control problem. To do this, the paper presents additional conditions for the random variable (see Assumptions 5.1 and 5.4), which allow finding a bound for the difference of the value functions of the control problems.

The paper is organized as follows. Firstly, in Sect. 2, the basic theory of the Markov decision processes is presented. Afterwards, in Sect. 3, the optimal decision problem with a random horizon is described. Then, in Sect. 4, the optimal solution is characterized through the dynamic programming approach and the theory is verified in two examples. Finally, in Sect. 5, under certain assumptions, the problem with the random horizon is connected to a discounted problem with an infinite horizon.

2 Preliminaries

Let $(X, A, \{A(x) : x \in X\}, Q, c)$ be a Markov decision model, which consists of the state space X, the action set A (X and A are Borel spaces), a family $\{A(x) : x \in X\}$ of nonempty measurable subsets $A(x)$ of A, whose elements are the feasible actions when the system is in state $x \in X$. The set $\mathbb{K} := \{(x, a) : x \in X, \ \ a \in A(x)\}$ of the feasible state-action pairs is assumed to be a measurable subset of $X \times A$. The following component is the transition law Q, which is a stochastic kernel on X given $\mathbb{K}$. Finally, $c : \mathbb{K} \to \mathbb{R}$ is a measurable function called the cost per stage function.

A policy is a sequence $\pi = \{\pi_t : t = 0, 1, \ldots\}$ of stochastic kernels π_t on the control set A given the history $\mathbb{H}_t$ of the process up to time t ($\mathbb{H}_t = \mathbb{K} \times \mathbb{H}_{t-1}$, $t = 1, 2, \ldots, \mathbb{H}_0 = X$). The set of all policies is denoted by Π.

$\mathbb{F}$ denotes the set of measurable functions $f : X \to A$ such that $f(x) \in A(x)$, for all $x \in X$. A deterministic Markov policy is a sequence $\pi = \{f_t\}$ such that $f_t \in \mathbb{F}$, for $t = 0, 1, 2, \ldots$. A Markov policy $\pi = \{f_t\}$ is said to be stationary iff f_t is independent of t, i.e., $f_t = f \in \mathbb{F}$, for all $t = 0, 1, 2, \ldots$. In this case, π is denoted by f and $\mathbb{F}$ is the set of stationary policies.

Remark 2.1 In many cases, the evolution of a Markov control process is specified by a difference equation of the form $x_{t+1} = F(x_t, a_t, \xi_t)$, $t = 0, 1, 2, \ldots$, with x_0 given, where $\{\xi_t\}$ is a sequence of independent and identically distributed random variables with values in a Borel space S and a common distribution μ, independent of the initial state x_0. In this case, the transition law Q is given by $Q(B \mid x, a) = \int_S I_B(F(x, a, s))\mu(ds)$, $B \in \mathcal{B}(X)$, $(x, a) \in \mathbb{K}$, where $\mathcal{B}(X)$ is the Borel σ-algebra of X and $I_B(\cdot)$ denotes the indicator function of the set B.

Let $(\Omega, \mathcal{F})$ be the measurable space consisting of the canonical sample space $\Omega = \mathbb{H}_\infty := (X \times A)^\infty$ and $\mathcal{F}$ be the corresponding product σ-algebra. The elements of Ω are sequences of the form $\omega = (x_0, a_0, x_1, a_1, \ldots)$ with $x_t \in X$ and $a_t \in A$ for all $t = 0, 1, 2, \ldots$. The projections x_t and a_t from Ω to the sets X and A are called state and action variables, respectively.

Let $\pi = \{\pi_t\}$ be an arbitrary policy and μ be an arbitrary probability measure on X called the initial distribution. Then, by the theorem of C. Ionescu-Tulcea (see [8]), there is a unique probability measure P_μ^π on $(\Omega, \mathcal{F})$ which is supported on $\mathbb{H}_\infty$, i.e., $P_\mu^\pi(\mathbb{H}_\infty) = 1$. The stochastic process $(\Omega, \mathcal{F}, P_\mu^\pi, \{x_t\})$ is called a discrete-time Markov control process or a Markov decision process.

The expectation operator with respect to P_μ^π is denoted by E_μ^π. If μ is concentrated at the initial state $x \in X$, then P_μ^π and E_μ^π are written as P_x^π and E_x^π, respectively.

3 Statement of the Problem

Let $(\Omega', \mathcal{F}', P)$ be a probability space and let $(X, A, \{A(x) : x \in X\}, Q, c)$ be a Markov decision model with a planning horizon τ, where τ is considered as a random variable on $(\Omega', \mathcal{F}')$ with the probability distribution $\rho_t := P(\tau = t)$, $t = 0, 1, 2, \ldots, T$, where T is a positive integer or $T = \infty$. Define the performance criterion as

$$J^\tau(\pi, x) := E\left[\sum_{t=0}^{\tau} c(x_t, a_t)\right],$$

$\pi \in \Pi$, $x \in X$, where E denotes the expected value with respect to the joint distribution of the process $\{(x_t, a_t)\}$ and τ. Then the optimal value function is defined as

$$J^\tau(x) := \inf_{\pi \in \Pi} J^\tau(\pi, x), \tag{1}$$

$x \in X$. The optimal control problem with a random horizon is to find a policy $\pi^* \in \Pi$ such that $J^\tau(\pi^*, x) = J^\tau(x)$, $x \in X$, in which case, π^* is said to be optimal.

Assumption 3.1 For each $x \in X$ and $\pi \in \Pi$, the induced process $\{(x_t, a_t)\}$ is independent of τ.

Remark 3.1 Observe that under Assumption 3.1,

$$\begin{aligned} E\left[\sum_{t=0}^{\tau} c(x_t, a_t)\right] &= E\left[E\left[\sum_{t=0}^{\tau} c(x_t, a_t) \mid \tau\right]\right] \\ &= \sum_{n=0}^{T} E_x^\pi\left[\sum_{t=0}^{n} c(x_t, a_t)\right]\rho_n \\ &= \sum_{t=0}^{T}\sum_{n=t}^{T} E_x^\pi\left[c(x_t, a_t)\right]\rho_n \\ &= E_x^\pi\left[\sum_{t=0}^{T} P_t c(x_t, a_t)\right], \end{aligned}$$

$\pi \in \Pi$, $x \in X$, where $P_k := \sum_{n=k}^{T} \rho_n = P(\tau \geq k)$, $k = 0, 1, 2, \ldots, T$. Thus, the optimal control problem with a random horizon τ is equivalent to the optimal control problem with a planning horizon T and a nonhomogeneous cost $P_t c$.

4 Characterization of the Optimal Solution using Dynamic Programming Approach

Firstly, the finite case $T < +\infty$ is presented.

Assumption 4.1

(a) The one-stage cost c is lower semicontinuous, nonnegative and inf-compact on $\mathbb{K}$ (c is inf-compact iff the set $\{a \in A(x) : c(x, a) \leq \lambda\}$ is compact for every $x \in X$ and $\lambda \in \mathbb{R}$).

(b) Q is either strongly continuous or weakly continuous.

Remark 4.1 Assumption 4.1 is well-known in the literature of MDPs. A more detailed explanation can be found in [8], p. 28.

Theorem 4.1 *Let $J_0, J_1, \ldots, J_{T+1}$ be the functions on X defined by $J_{T+1}(x) := 0$ and for $t = T, T-1, \ldots, 0$,*

$$J_t(x) := \min_{a \in A(x)} \left[P_t c(x, a) + \int_X J_{t+1}(y) Q(dy \mid x, a) \right], \quad x \in X. \tag{2}$$

Under Assumption 4.1, these functions are measurable and for each $t = 0, 1, \ldots, T$, there is $f_t \in \mathbb{F}$ such that $f_t(x) \in A(x)$ attains the minimum in (2) for all $x \in X$. This implies that

$$J_t(x) = P_t c\big(x, f_t(x)\big) + \int_X J_{t+1}(y) Q\big(dy \mid x, f_t(x)\big),$$

$x \in X$ and $t = 0, 1, \ldots, T$. Then the deterministic Markov policy $\pi^* = \{f_0, \ldots, f_T\}$ is optimal and the optimal value function is given by $J^\tau(x) = J^\tau(\pi^*, x) = J_0(x)$, $x \in X$.

The proof of Theorem 4.1 is similar to the proof of Theorem 3.2.1 in [8].

Let $U_{T+1} := 0$ and

$$U_t := \frac{J_t}{P_t},$$

$t \in \{0, 1, 2, \ldots, T-1\}$. Then (2) is equivalent to

$$U_t(x) = \min_{a \in A(x)} \left[c(x, a) + \alpha_t \int_X U_{t+1}(y) Q(dy \mid x, a) \right], \tag{3}$$

where

$$\alpha_t := \frac{P_{t+1}}{P_t}, \quad t \in \{0, 1, 2, \ldots, T-1\}. \tag{4}$$

Remark 4.2 Observe that $\alpha_t = P(\tau \geq t+1 \mid \tau \geq t)$.

Now, let us analyze the case when $T = +\infty$. Under this condition, the performance criterion is the following (see Remark 3.1):

$$J^{\tau}(\pi, x) = E_x^{\pi}\left[\sum_{t=0}^{\infty} P_t c(x_t, a_t)\right], \tag{5}$$

$\pi \in \Pi$ and $x \in X$.

For every $n = 0, 1, 2, \ldots$, define

$$v_n^{\tau}(\pi, x) := E_x^{\pi}\left[\sum_{t=n}^{\infty}\prod_{k=n}^{t}\alpha_{k-1} c(x_t, a_t)\right], \tag{6}$$

$\pi \in \Pi$, $x \in X$, and

$$V_n^{\tau}(x) := \inf_{\pi\in\Pi} v_n^{\tau}(\pi, x), \quad x \in X. \tag{7}$$

$v_n^{\tau}(\pi, x)$ is the expected total cost from time n onwards, applied to (5), given the initial condition $x_n = x$, where x is a generic element of X.

Remark 4.3

(i) Note that $P_t = \prod_{k=0}^{t}\alpha_{k-1}$, where $t = 0, 1, 2, \ldots$, $\alpha_{-1} := P_0 = 1$ and α_k, $k = 0, 1, 2, \ldots$, is defined by (4).
(ii) Observe that $V_0^{\tau}(x) = J^{\tau}(x)$, $x \in X$ (see (1)).
(iii) In (6), if $n = 0$, $\alpha_k = \alpha$, $k \geq 0$ and $\alpha_{-1} = 1$, the performance criterion is reduced to an expected total discounted cost with a fixed discount factor.

For $N > n \geq 0$, we define

$$v_{n,N}^{\tau}(\pi, x) := E_x^{\pi}\left[\sum_{t=n}^{N}\prod_{k=n}^{t}\alpha_{k-1} c(x_t, a_t)\right], \tag{8}$$

with $\pi \in \Pi$, $x \in X$, and

$$V_{n,N}^{\tau}(x) := \inf_{\pi\in\Pi} v_{n,N}^{\tau}(\pi, x), \quad x \in X. \tag{9}$$

Assumption 4.2

(a) Same as Assumption 4.1.
(b) There exists a policy $\pi \in \Pi$ such that $J^{\tau}(\pi, x) < \infty$ for each $x \in X$.

Remark 4.4

(i) In Assumption 4.2, it is supposed that the cost function is nonnegative. This assumption can be changed without any loss of generality by taking a c which is bounded below. Namely, if $c \geq m$ for some constant m, then the problem with

a one-stage cost $c' := c - m$, which is nonnegative, is equivalent to the problem with a one-stage cost c.

(ii) Observe that if $c \geq m$, if Assumption 4.2(b) holds and, since

$$J^{\tau}(\pi, x) \geq m \sum_{t=0}^{\infty} P_t = m\big(1 + E[\tau]\big),$$

then $E[\tau] < \infty$. Conversely, in the case of c bounded, if $E[\tau] < \infty$, then Assumption 4.2(b) holds.

$M(X)^+$ denotes the cone of nonnegative measurable functions on X.

The proofs of Lemmas 4.1 and 4.2 below are similar to the proofs of Lemmas 4.2.4 and 4.2.6 in [8], respectively. This is why these proofs are omitted.

Lemma 4.1 *For every $N > n \geq 0$, let w_n and $w_{n,N}$ be functions on $\mathbb{K}$, which are nonnegative, lower semicontinuous and inf-compact on $\mathbb{K}$. If $w_{n,N} \uparrow w_n$ as $N \to \infty$, then*

$$\lim_{N\to\infty} \min_{a\in A(x)} w_{n,N}(x,a) = \min_{a\in A(x)} w_n(x,a), \quad x \in X.$$

Lemma 4.2 *Suppose that Assumption 4.1 holds. For every $u \in M(X)^+$, $\min_{a\in A(x)}[c(x,a) + \alpha_n \int_X u(y)Q(dy \mid x,a)] \in M(X)^+$. Moreover, there exists f_n in $\mathbb{F}$ such that*

$$\min_{a\in A(x)}\left[c(x,a) + \alpha_n \int_X u(y)Q(dy \mid x,a)\right] = c(x, f_n) + \alpha_n \int_X u(y)Q(dy \mid x, f_n),$$

$x \in X$.

Lemma 4.3 *Suppose that Assumption 4.2(a) holds and let $\{u_n\}$ be a sequence in $M(X)^+$. If $u_n \geq \min_{a\in A(x)}[c(x,a) + \alpha_n \int_X u_{n+1}(y)Q(dy \mid x,a)]$, $n = 0, 1, 2, \ldots$, then $u_n \geq V_n^{\tau}$, $n = 0, 1, 2, \ldots$.*

Proof Let $\{u_n\}$ be a sequence in $M(X)^+$ such that

$$u_n(x) \geq \min_{a\in A(x)}\left[c(x,a) + \alpha_n \int_X u_{n+1}(y)Q(dy \mid x,a)\right],$$

then, by Lemma 4.2,

$$u_n(x) \geq c\big(x, f_n(x)\big) + \alpha_n \int_X u_{n+1}(y)Q\big(dy \mid x, f_n(x)\big), \quad x \in X.$$

Iterating this inequality, one obtains

$$u_n(x) \geq E_x^{\pi}\left[c\big(x_n, f_n(x_n)\big) + \sum_{t=n+1}^{N-1}\prod_{j=n+1}^{t}\alpha_{j-1}c\big(x_t, f_t(x_t)\big)\right] + \prod_{j=n+1}^{N}\alpha_{j-1}E_x^{\pi}\big[u(x_N)\big], \quad x \in X. \tag{10}$$

Here,

$$E_x^{\pi}\big[u(x_N)\big] = \int_X u(y)Q^N\big(dy \mid x_n, f_n(x_n)\big),$$

where $Q^N(\cdot \mid x_n, f_n(x_n))$ denotes the N-step transition kernel of the Markov process $\{x_t\}$ when the policy $\pi = \{f_k\}$ is used, beginning at a stage n. Since u is nonnegative, $\alpha_k \leq 1$ and $x_n = x$, it is obtained from (10) that

$$u_n(x) \geq E_x^{\pi}\left[\alpha_{n-1}c\big(x_n, f_n(x_n)\big) + \sum_{t=n+1}^{N-1}\prod_{j=n}^{t}\alpha_{j-1}c\big(x_t, f_t(x_t)\big)\right].$$

Hence, letting $N \to \infty$ yields

$$u_n(x) \geq v_n^{\tau}(\pi, x) \geq V_n^{\tau}(x), \quad x \in X. \qquad \square$$

Lemma 4.4 *Suppose that Assumption 4.2 holds. Then, for every $n \geq 0$ and $x \in X$,*

$$V_{n,N}^{\tau}(x) \uparrow V_n^{\tau}(x) \quad \text{as } N \to \infty$$

and V_n^{τ} is lower semicontinuous.

Proof Using the dynamic programming equation given in (3), that is,

$$U_t(x) = \min_{a \in A(x)}\left[c(x,a) + \alpha_t \int_X U_{t+1}(y)Q(dy \mid x, a)\right], \tag{11}$$

for $t = N-1, N-2, \ldots, n$, with $U_N(x) = 0$, $x \in X$, it is obtained that $V_{n,N}^{\tau}(x) = U_n(x)$ and $V_{s,N}^{\tau}(x) = U_s(x)$, $n \leq s < N$. Furthermore, it is proved by backwards induction that U_s, $n \leq s < N$, is lower semicontinuous. For $t = n$, (11) is written as

$$V_{n,N}^{\tau}(x) = \min_{a \in A(x)}\left[c(x,a) + \alpha_n \int_X V_{n+1,N}^{\tau}(y)Q(dy \mid x, a)\right], \tag{12}$$

and $V_{n,N}^{\tau}(\cdot)$ is lower semicontinuous. Then, by the nonnegativity of c, for each $n = 0, 1, 2, \ldots$, the sequence $\{V_{n,N} : N = n, n+1, \ldots\}$ is nondecreasing. This implies that there exists a function $u_n \in M(X)^+$ such that for each $x \in X$, $V_{n,N}^{\tau}(x) \uparrow u_n(x)$, as $N \to \infty$. Moreover,

$$V_{n,N}^{\tau}(x) \leq v_{n,N}^{\tau}(\pi, x) \leq v_n^{\tau}(\pi, x),$$

$x \in X$ and $\pi \in \Pi$. Hence $V^{\tau}_{n,N}(x) \leq V^{\tau}_{n}(x)$, $N > n$, then $u_n \leq V^{\tau}_{n}$. Furthermore, u_n being the supremum of a sequence of lower semicontinuous functions, is lower semicontinuous. Using Lemma 4.1 and letting $N \to \infty$ in (12), it is obtained that

$$u_n(x) = \min_{a \in A(x)} \left[c(x,a) + \alpha_n \int_X u_{n+1}(y) Q(dy \mid x,a) \right], \tag{13}$$

$n = 0, 1, 2, \ldots$ and $x \in X$. Finally, by Lemma 4.3, $u_n \geq V^{\tau}_{n}$, we obtain that $u_n = V^{\tau}_{n}$ and conclude this way the proof of Lemma 4.4. □

Theorem 4.2 *Suppose that Assumption 4.2 holds, then*

(a) *The optimal value function V^{τ}_{n}, $n = 0, 1, 2, \ldots$, satisfies the optimality equation*

$$V^{\tau}_{n}(x) = \min_{a \in A(x)} \left[c(x,a) + \alpha_n \int_X V^{\tau}_{n+1}(y) Q(dy \mid x,a) \right], \quad x \in X, \tag{14}$$

and if $\{u_n\}$ is another sequence that satisfies the optimality equations in (14), then $u_n \geq V^{\tau}_{n}$.

(b) *There exists a policy $\pi^* = \{f_n \in \mathbb{F} \mid n \geq 0\}$ such that, for each $n = 0, 1, 2, \ldots$, the control $f_n(x) \in A(x)$ attains the minimum in (14), i.e.,*

$$V^{\tau}_{n}(x) = c\big(x, f_n(x)\big) + \alpha_n \int_X V^{\tau}_{n+1}(y) Q\big(dy \mid x, f_n(x)\big), \quad x \in X, \tag{15}$$

and the policy π^* is optimal.

Proof

(a) The proof of Lemma 4.4 guarantees that the sequence $\{V^{\tau}_{n}\}$ satisfies the optimality equations in (14), and by Lemma 4.3, if $\{u_n\}$ satisfies

$$u_n = \min_{a \in A(x)} \left[c(x,a) + \alpha_n \int_X u_{n+1}(y) Q(dy \mid x,a) \right],$$

it is concluded that $u_n \geq V^{\tau}_{n}$.

(b) The existence of $f_n \in \mathbb{F}$ that satisfies (15) is ensured by Lemma 4.2. Now, iterating (15) with $x_n = x \in X$, it is obtained that

$$\begin{aligned} V^{\tau}_{n}(x) &= E^{\pi}_{x} \left[c\big(x_n, f_n(x_n)\big) + \sum_{t=n+1}^{N-1} \prod_{j=n+1}^{t} \alpha_{j-1} c\big(x_t, f_t(x_t)\big) \right] \\ &\quad + \prod_{j=n+1}^{N} \alpha_{j-1} E^{\pi}_{x}\big[u(x_N)\big] \\ &\geq E^{\pi}_{x} \left[\sum_{t=n}^{N-1} \prod_{j=n}^{t} \alpha_{j-1} c\big(x_t, f_t(x_t)\big) \right], \end{aligned}$$

$n \geq 0$ and $N > n$. This implies that, letting $N \to \infty$, $V_n^\tau(x) \geq v_n^\tau(\pi^*, x)$, $x \in X$, and $\pi^* = \{f_k\}$. Moreover, in particular for π^*, $V_n^\tau(x) \leq v_n^\tau(\pi^*, x)$, $x \in X$. Therefore, π^* is optimal. □

Example 4.1 (A Linear–Quadratic Model (LQM) with a Random Horizon) Let $X = A = A(x) = \mathbb{R}$. The cost function is given by $c(x, a) = qx^2 + ra^2$, $(x, a) \in \mathbb{K}$, such that $q \geq 0$ and $r > 0$. The dynamics of the system is given by $x_{t+1} = \gamma x_t + \beta a_t + \xi_t$, $t = 0, 1, 2, \ldots, \tau$, with x_0 known. In this case, $\gamma, \beta \in \mathbb{R}$ and $\{\xi_t\}$ is a sequence of independent and identically distributed random variables taking values in $S = \mathbb{R}$, such that $E[\xi_0] = 0$ and $E[\xi_0^2] = \sigma^2$, where ξ_0 is a generic element of the sequence $\{\xi_t\}$.

Now, the LQM will be solved considering the following cases.
(a) It is assumed that the distribution of the horizon τ has a finite support, that is, $P(\tau = k) = \rho_k$, $k = 1, 2, 3, \ldots, T$, $T < \infty$.

Lemma 4.5 *The optimal policy π^* and the optimal value function J^τ for LQM are given by $\pi^* = (f_0, f_1, \ldots, f_T)$, where*

$$f_n(x) = -\frac{\alpha_n C_{n+1} \gamma \beta}{r + \alpha_n C_{n+1} \beta^2} x, \quad n = T, T-1, \ldots, 0,$$

and $J^\tau(x) = C_0 x^2 + D_0$, $x \in X$, where the constants C_n and D_n satisfy the following recurrence relations:

$$C_{T+1} = 0,$$
$$C_n = \frac{qr + \alpha_n C_{n+1}(q\beta^2 + r\gamma^2)}{r + \alpha_n C_{n+1} \beta^2}, \quad n = T, T-1, \ldots, 0,$$

and

$$D_{T+1} = 0,$$
$$D_n = \alpha_n \big(C_{n+1}\sigma^2 + D_{n+1}\big), \quad n = T, T-1, \ldots, 0.$$

Proof In this case, using (3), the dynamic programming equation is

$$U_t(x) = \min_{a \in \mathbb{R}} \big[qx^2 + ra^2 + \alpha_t E\big[U_{t+1}(\gamma x + \beta a + \xi)\big]\big], \tag{16}$$

where $\alpha_t = P(\tau > t+1)/P(\tau > t)$. For $t = T$ in (16), with $U_{T+1}(x) = 0$, it is obtained that $f_T(x) = 0$ and $U_T(x) = qx^2$, $x \in X$.

For $t = T - 1$, replacing U_T in (16), it is obtained that

$$\begin{aligned} U_{T-1}(x) &= \min_{a \in \mathbb{R}} \big[qx^2 + ra^2 + \alpha_{T-1} E\big[q(\gamma x + \beta a + \xi)^2\big]\big] \\ &= \min_{a \in \mathbb{R}} \big[qx^2 + ra^2 + \alpha_{T-1} q\big(\gamma^2 x^2 + \beta^2 a^2 + 2\gamma\beta a x + \sigma^2\big)\big], \quad x \in X. \end{aligned}$$

Then

$$f_{T-1}(x) = \frac{-\alpha_{T-1} C_T \gamma \beta}{r + \alpha_{T-1} C_T \beta^2} x, \quad x \in X,$$

where $C_T = q$. Hence $U_{T-1}(x) = C_{T-1}x^2 + D_{T-1}$, where

$$C_{T-1} = \frac{qr + \alpha_{T-1} C_T (q\beta^2 + r\gamma^2)}{r + \alpha_{T-1} C_T \beta^2}$$

and $D_{T-1} = C_T \alpha_{T-1} \sigma^2$.

Continuing with the procedure, it follows that

$$f_{T-2}(x) = \frac{-\alpha_{T-2} C_{T-1} \gamma \beta}{r + \alpha_{T-2} C_{T-1} \beta^2} x$$

and $U_{T-1}(x) = C_{T-2}x^2 + D_{T-2}$, $x \in X$, where

$$C_{T-2} = \frac{qr + \alpha_{T-2} C_{T-1} (q\beta^2 + r\gamma^2)}{r + \alpha_{T-2} C_{T-1} \beta^2}$$

and $D_{T-2} = \alpha_{T-2}(C_{T-1}\sigma^2 + D_{T-1})$.

Finally, in $t = 0$, it is obtained that

$$f_0(x) = \frac{-\alpha_0 C_1 \gamma \beta}{r + \alpha_0 C_1 \beta^2} x,$$

and $U_0(x) = C_0 x^2 + D_0$, $x \in X$, where

$$C_0 = \frac{qr + \alpha_0 C_1 (q\beta^2 + r\gamma^2)}{r + \alpha_0 C_1 \beta^2}$$

and $D_0 = \alpha_0(C_1\sigma^2 + D_1)$. Since Assumption 4.1 clearly holds, the result is obtained applying Theorem 4.1. □

(b) Now, it is supposed that the distribution of the horizon τ has an infinite support with $E[\tau] < \infty$.

Lemma 4.6 *LQM satisfies Assumption 4.2.*

Proof Clearly, Assumption 4.2(a) holds. Now, consider the stationary policy $h(x) = -\frac{\gamma}{\beta}x$, $x \in X$. In this case, it results in

$$x_t = \xi_{t-1}, \quad t \leq 1.$$

Then

$$J^\tau(h, x) = E_x^h \left[\sum_{t=0}^{\infty} P_t c(x_t, a_t) \right]$$

$$= E_x^h\left[\sum_{t=0}^{\infty} P_t\left(q + r\frac{\gamma^2}{\beta^2}\right)\xi_{t-1}^2\right]$$

$$= \left(q + r\frac{\gamma^2}{\beta^2}\right)\sigma^2\big(E[\tau] + 1\big).$$

Since $E[\tau] < \infty$, Assumption 4.2(b) holds. □

Example 4.2 (A Logarithm Consumption–Investment Model (LCIM)) Consider an investor who wishes to allocate his current wealth x_t between an investment a_t and consumption $x_t - a_t$, at each period $t = 0, 1, 2, \ldots, \tau$, where τ is a random variable with an infinite support. It is assumed that the investment constrain set is $A(x) = [0, x) \subseteq A = [0, 1]$, $x \in (0, 1]$ and $A(0) = 0$. This way the state space is $X = [0, 1]$. In this case, the utility function is defined by $u(x - a) := \ln(x - a)$ if $x \in (0, 1]$, and $u(x - a) := 0$ if $x = 0$. The relation between the investment and the accumulated capital is given by $x_{t+1} = a_t\xi_t$, $t = 0, 1, 2, \ldots, \tau$ with $x_0 = x \in X$, where $\{\xi_t\}$ is a sequence of random variables taking values in $(0, 1)$, independent and identically distributed with continuous density function Δ, such that $E[|\ln \xi_0|] = K < \infty$, where ξ_0 is a generic element of the sequence $\{\xi_t\}$.

In the case of maximization, equivalent results are obtained with adequate changes in the assumptions. For instance, in Assumption 4.2, the following change is necessary:

Assumption 4.3

(a) The one-stage reward g is upper semicontinuous, nonpositive and sup-compact on $\mathbb{K}$.
(b) Q is either strongly continuous or weakly continuous.
(c) There exists a policy $\pi \in \Pi$ such that $J^\tau(\pi, x) > -\infty$ for each $x \in X$.

Lemma 4.7 *The problem LCIM with a random horizon satisfies Assumption 4.3.*

Proof Observe that the set $A_\lambda(u) := \{a \in A(x) : u(x - a) \geq \lambda\}$, $\lambda \in \mathbb{R}$, is equal to $\{0\}$ if $\lambda \leq 0$ and equal to $\emptyset$ if $\lambda > 0$ for $x = 0$; equal to $[0, x - \exp(\lambda)]$ if $\lambda \leq \ln x$ and equal to $\emptyset$ if $\lambda > \ln x$ for $x \in (0, 1]$. Since $A_\lambda(u)$ is closed and compact for every $x \in X$, for Proposition A.1 in Appendix of [8], u is upper semicontinuous and sup-compact by definition. So Assumption 4.3(a) holds.

Now let $v : X \to \mathbb{R}$ be a continuous and bounded function and define

$$v'(x, a) := \int_X v(y)\,Q(dy \mid x, a)$$

$$= \int_0^1 v(as)\Delta(s)\,ds,$$

$x \in X$ and $a \in A(x)$. Observe that changing the variable u by as, it is obtained that

$$v'(x, a) = \int_{-\infty}^{\infty} v(u)I_{[0,a]}(u)\Delta\left(\frac{u}{a}\right)\frac{du}{a},$$

if $a \neq 0$ and $v'(x,0) = v(0)$. Since Δ is continuous, v' is continuous for $a \neq 0$. Let $\{a_n\}$ be an arbitrary sequence such that $a_n \to 0$, so

$$\lim_{n\to\infty} v'(x,a_n) = \lim_{n\to\infty} \int_0^1 v(a_n s)\Delta(s)\,ds = v(0),$$

then v' is continuous for $a = 0$. This way, it is verified that v' is continuous and bounded on $\mathbb{K}$ for every continuous and bounded function v on X, i.e., Q is weakly continuous. On the other hand, using $h(x) = 0$, $x \in X$, it is obtained that $J^\tau(h,x) = P_0 \ln x > -\infty$, $x \in X$. □

5 The Optimal Control Problem with a Random Horizon as a Discounted Problem

The optimal problem with a random horizon which is geometrically distributed with a parameter p $(0 < p < 1)$ coincides with a discounted problem with a discount factor p (see [4] and [1] p. 126). In this section, a condition is given that allows bounding the problem with a random horizon τ with an appropriate discounted problem, which is simpler in practice.

Recall that α_t, $t = 0, 1, 2\ldots$, is defined as

$$\alpha_t := \frac{P_{t+1}}{P_t}, \tag{17}$$

where $P_t = P(\tau \geq t)$, and consider the following assumption.

Assumption 5.1 $\{\alpha_t\}_{t=0}^\infty \subset (0,1)$ is a sequence such that $\bar{\alpha} := \lim_{t\to\infty} \alpha_t$ and $\alpha_t \leq \bar{\alpha}$.

Remark 5.1

(i) In the case of maximization, the assumption corresponding to Assumption 5.1 is the following: $\{\alpha_t\}_{t=0}^\infty \subset (0,1)$ is a sequence such that $\lim_{t\to\infty} \alpha_t = \bar{\alpha}$ and $\alpha_t \geq \bar{\alpha}$.

(ii) It is possible to find probability distributions that satisfy Assumption 5.1.

(iia) Consider τ with Logarithmic distribution, i.e., $\rho_k = -\frac{(1-p)^{k+1}}{(k+1)\ln p}$, $k = 0, 1, 2, \ldots$, where $0 < p < 1$. Since

$$\begin{aligned}
\alpha_t &= 1 - \frac{\rho_t}{\sum_{k=t}^\infty \rho_k} \\
&= 1 - \frac{(1-p)^{t+1}}{(t+1)\sum_{j=0}^\infty \frac{(1-p)^{t+j+1}}{t+j+1}} \\
&= 1 - \frac{1}{\sum_{j=0}^\infty \frac{(1-p)^j}{1+\frac{j}{t+1}}},
\end{aligned}$$

it is directly obtained that $\alpha_t \leq \alpha_{t+1}$ for all $t = 0, 1, 2, \ldots$, and

$$\bar{\alpha} = \lim_{t\to\infty} \alpha_t = 1 - p.$$

(iib) Take τ with a negative binomial distribution whose parameters are q and r, where $0 \leq q \leq 1$ and $r \in \mathbb{N}$. In this case,

$$\begin{aligned}\alpha_t &= 1 - \frac{\rho_t}{\sum_{k=t}^{\infty} \rho_k} \\ &= 1 - \frac{\frac{(r+t-1)!}{t!}}{\sum_{j=0}^{\infty} \frac{(r+t+j-1)!q^j}{(t+j)!}} \\ &= 1 - \frac{(1+\frac{1}{t})(1+\frac{2}{t})\cdots(1+\frac{r-1}{t})}{\sum_{j=0}^{\infty}(1+\frac{j+1}{t})(1+\frac{j+2}{t})\cdots(1+\frac{j+r-1}{t})q^j},\end{aligned}$$

then

$$\bar{\alpha} = \lim_{t\to\infty} \alpha_t = q.$$

Moreover, it is verified that $\alpha_{t+1} \leq \alpha_t$.

Now, consider the Markov decision model $(X, A, \{A(x) \mid x \in X\}, Q, c)$ and the performance criterion as

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

$\pi \in \Pi$, $x \in X$ and $\bar{\alpha} = \lim_{t\to\infty} \alpha_t$ (see Assumption 5.1). Let

$$V(x) := \inf_{\pi\in\Pi} v(\pi, x), \quad x \in X. \tag{18}$$

Consider the following assumption.

Assumption 5.2 There exists a policy $\pi \in \Pi$ such that $v(\pi, x) < \infty$, for each $x \in X$.

Lemma 5.1 *Suppose that Assumptions* 4.2(a), 5.1, *and* 5.2 *hold. Then* $J^{\tau}(x) \leq V(x)$, $x \in X$.

Proof By Theorem 4.2.3 in [8], there exists $f \in \mathbb{F}$ such that

$$V(x) = c\big(x, f(x)\big) + \bar{\alpha}\int_X V(y)Q\big(dy \mid x, f(x)\big), \quad x \in X. \tag{19}$$

Iterating (19), it is obtained that

$$V(x) = E_x^f\left[\sum_{t=0}^{n-1} \bar{\alpha}^t c\big(x_t, f(x_t)\big)\right] + \bar{\alpha}^n E_x^f\big[V(x_n)\big],$$

$n \geq 1$, $x \in X$, where

$$E_x^f\big[V(x_n)\big] = \int_X V(y)\, Q^n\big(dy \mid x, f(x)\big),$$

$Q^n(\cdot \mid x, f)$ denotes the n-step transition kernel of the Markov process $\{x_t\}$ when using $f \in \mathbb{F}$. By Assumption 5.1, it results in

$$\sum_{t=0}^{n-1} \bar{\alpha}^t c\big(x_t, f(x_t)\big) \geq \sum_{t=0}^{n-1} \prod_{k=0}^{t} \alpha_{k-1} c\big(x_t, f(x_t)\big) = \sum_{t=0}^{n-1} P_t c\big(x_t, f(x_t)\big),$$

hence

$$V(x) \geq E_x^f\left[\sum_{t=0}^{n-1} P_t c\big(x_t, f(x_t)\big)\right] + \bar{\alpha}^n E_x^f\big[V(x_n)\big],$$

$n \geq 1$, $x \in X$. Since V is nonnegative,

$$V(x) \geq E_x^f\left[\sum_{t=0}^{n-1} P_t c\big(x_t, f(x_t)\big)\right],$$

$n \geq 1$, $x \in X$. Letting $n \to \infty$ yields

$$V(x) \geq J^\tau(f, x) \geq J^\tau(x), \quad x \in X. \qquad \square$$

Let $f^* \in \mathbb{F}$ be the stationary optimal policy of the discounted problem and consider the following assumption:

Assumption 5.3 There exist positive numbers m and k, with $1 \leq k < 1/\bar{\alpha}$, and a function $w \in M(X)^+$ such that, for all $(x, a) \in \mathbb{K}$,

(a) $c(x, a) \leq m w(x)$, and
(b) $\int_X w(y) Q(dy \mid x, a) \leq k w(x)$.

Remark 5.2

(i) Assumption 5.3 implies Assumption 4.2(b).
(ii) Observe that

$$E_x^\pi\big[w(x_t)\big] \leq k^t w(x), \tag{20}$$

$t = 0, 1, 2, \ldots$ and $x \in X$.

For $t = 0$, (20) trivially holds, and for $t \geq 1$, using Assumption 5.3(b), it is obtained that

$$E_x^\pi\big[w(x_t) \mid h_{t-1}, a_{t-1}\big] = \int_X w(y) Q(dy \mid x_{t-1}, a_{t-1}) \leq k w(x_{t-1}).$$

Taking expectations into account results in

$$E_x^{\pi}\big[w(x_t)\big] \leq k E_x^{\pi}\big[w(x_{t-1})\big]. \tag{21}$$

Iterating (21), (20) is obtained.

Lemma 5.2 *Under Assumptions* 4.2(a), 5.1, *and* 5.3,

$$0 \leq V(x) - J^{\tau}\big(f^*, x\big) \leq m w(x) \sum_{t=0}^{\infty} \big(\bar{\alpha}^t - P_t\big) k^t.$$

Proof Firstly, observe that $\sum_{t=0}^{\infty}(\bar{\alpha}^t - P_t)k^t \leq \sum_{t=0}^{\infty}(\bar{\alpha}k)^t < \infty$, since $0 < \bar{\alpha}k < 1$ by Assumption 5.3. Then, under Assumption 5.3(a) and (20), for a stationary policy $f \in \mathbb{F}$, it is obtained that

$$\begin{aligned} v(f,x) - J^{\tau}(f,x) &= \sum_{t=0}^{\infty} \big(\bar{\alpha}^t - P_t\big) E_x^f\big[c\big(x_t, f(x_t)\big)\big] \\ &\leq m \sum_{t=0}^{\infty} \big(\bar{\alpha}^t - P_t\big) E_x^f\big[w(x_t)\big] \\ &\leq m w(x) \sum_{t=0}^{\infty} \big(\bar{\alpha}^t - P_t\big) k^t. \end{aligned}$$

Taking $f = f^*$, where f^* is the deterministic stationary optimal policy of the discounted problem, the proof is concluded. □

Let $D_n : \mathbb{K} \to \mathbb{R}$ be the discrepancy functions defined as

$$D_n(x,a) := c(x,a) + \alpha_n \int_X V_{n+1}^{\tau}(y) Q(dy \mid x, a) - V_n^{\tau}(x), \quad (x,a) \in \mathbb{K}.$$

Assumption 5.4 $P(\tau < +\infty) = 1$.

Theorem 5.1 *Under Assumptions* 4.2(a), 5.1, 5.3, *and* 5.4,

$$V(x) - J^{\tau}(x) \leq m w(x) \sum_{t=0}^{\infty} \big(\bar{\alpha}^t - P_t\big) k^t + \sum_{t=0}^{\infty} \prod_{k=0}^{t} \alpha_{k-1} E_x^{\pi}\big[D_t\big(x_t, f^*\big)\big], \tag{22}$$

$x \in X$ *and* $\pi \in \Pi$.

Proof For $x \in X$, by Lemma 5.2,

$$\begin{aligned} V(x) - J^{\tau}(x) &= V(x) - J^{\tau}\big(f^*, x\big) + J^{\tau}\big(f^*, x\big) - J^{\tau}(x) \\ &\leq m w(x) \sum_{t=0}^{\infty} \big(\bar{\alpha}^t - P_t\big) k^t + J^{\tau}\big(f^*, x\big) - J^{\tau}(x). \end{aligned}$$

Now, for $\pi \in \Pi$ and $x \in X$,

$$\begin{aligned} v_n^\tau(\pi, x) &= E_x^\pi\left[\sum_{t=n}^{\infty}\prod_{k=n}^{t}\alpha_{k-1}c(x_t, a_t)\right] \\ &= E_x^\pi\left[\sum_{t=n}^{\infty}\prod_{k=n}^{t}\alpha_{k-1}D_t(x_t, a_t)\right] \\ &\quad - E_x^\pi\left[\sum_{t=n}^{\infty}\prod_{k=n}^{t+1}\alpha_{k-1}\left(\int_X V_{t+1}^\tau(y)Q(dy \mid x_t, a_t) - V_t^\tau(x_t)\right)\right] \\ &= E_x^\pi\left[\sum_{t=n}^{\infty}\prod_{k=n}^{t}\alpha_{k-1}D_t(x_t, a_t)\right] \\ &\quad - E_x^\pi\left[\sum_{t=n}^{\infty}\left(\prod_{k=n}^{t+1}\alpha_{k-1}E_x^\pi\left[V_{t+1}^\tau(x_{t+1}) \mid x_t, a_t\right] - \prod_{k=n}^{t}\alpha_{k-1}V_t^\tau(x_t)\right)\right] \\ &= E_x^\pi\left[\sum_{t=n}^{\infty}\prod_{k=n}^{t}\alpha_{k-1}D_t(x_t, a_t)\right] \\ &\quad - \sum_{t=n}^{\infty}\left[\prod_{k=n}^{t+1}\alpha_{k-1}E_x^\pi\left[V_{t+1}^\tau(x_{t+1})\right] - \prod_{k=n}^{t}\alpha_{k-1}E_x^\pi\left[V_t^\tau(x_t)\right]\right]. \end{aligned} \tag{23}$$

Observe that, for some positive integer M,

$$\begin{aligned} &\sum_{t=n}^{\infty}\left[\prod_{k=n}^{t+1}\alpha_{k-1}E_x^\pi\left[V_{t+1}^\tau(x_{t+1})\right] - \prod_{k=n}^{t}\alpha_{k-1}E_x^\pi\left[V_t^\tau(x_t)\right]\right] \\ &\quad = \lim_{M\to\infty}\sum_{t=n}^{M}\left[\prod_{k=n}^{t+1}\alpha_{k-1}E_x^\pi\left[V_{t+1}^\tau(x_{t+1})\right] - \prod_{k=n}^{t}\alpha_{k-1}E_x^\pi\left[V_t^\tau(x_t)\right]\right] \\ &\quad = \lim_{M\to\infty}\left[\prod_{k=n}^{M+1}\alpha_{k-1}E_x^\pi\left[V_{M+1}^\tau(x_{M+1})\right] - \alpha_{n-1}V_n^\tau(x_n)\right] \\ &\quad = \lim_{M\to\infty}\left[\frac{P_{M+1}}{P_{n-1}}E_x^\pi\left[V_{M+1}^\tau(x_{M+1})\right] - \alpha_{n-1}V_n^\tau(x_n)\right], \end{aligned}$$

and by Assumption 5.4, $\lim_{M\to\infty}\frac{P_{M+1}}{P_{n-1}} = 0$. Then it is obtained in (23) that

$$v_n^\tau(\pi, x) = E_x^\pi\left[\sum_{t=n}^{\infty}\prod_{k=n}^{t}\alpha_{k-1}D_t(x_t, a_t)\right] + \alpha_{n-1}V_n^\tau(x_n).$$

Since $\alpha_{n-1} \leq 1$ and $x = x_n$, it is obtained that

$$v_n^\tau(\pi, x) - V_n^\tau(x) \leq \sum_{t=n}^{\infty} \prod_{k=n}^{t} \alpha_{k-1} E_x^\pi \big[D_t(x_t, a_t) \big].$$

Finally, since $v_0^\tau(\pi, x) = J^\tau(\pi, x)$, and taking $\pi = f^*$, (22) is obtained. □

6 Concluding Remarks

The results obtained in this paper permit working with discounted control problems with a varying-time discount factor, possibly depending on the state of the system and on the corresponding action as well. Besides, for MDPs taken into account in the article, i.e., MDPs with a total cost and a random horizon, it is possible to develop methods, as the rolling horizon procedure or the policy iteration algorithm, in order to approximate the optimal value function and the optimal policy.

References

1. Puterman, M.L.: Markov Decision Process: Discrete Stochastic Dynamic Programming. Wiley, New York (1994)
2. Baüerle, N., Rieder, U.: Markov Decision Processes with Applications to Finance. Springer, New York (2010)
3. Kozlowski, E.: The linear–quadratic stochastic optimal control problem with random horizon at the finite number of infinitesimal events. Ann. UMCS Inform. **1**, 103–115 (2010)
4. Levhari, D., Mirman, L.J.: Savings and consumption with uncertain horizon. J. Polit. Econ. **85**, 265–281 (1977)
5. Bather, J.: Decision Theory: An Introduction to Dynamic Programming and Sequential Decision. Wiley, New York (2000)
6. Chatterjee, D., Cinquemani, E., Chaloulos, G., Lygeros, J.: Stochastic control up to a hitting time: optimality and Rolling-Horizon implementation (2009). arXiv:0806.3008
7. Iida, T., Mori, M.: Markov decision processes with random horizon. J. Oper. Res. Soc. Jpn. **39**, 592–603 (1996)
8. Hernández-Lerma, O., Lasserre, J.B.: Discrete-Time Markov Control Processes: Basic Optimality Criteria. Springer, New York (1996)
9. Guo, X., Hernandez-del-Valle, A., Hernández-Lerma, O.: Nonstationary discrete-time deterministic and stochastic control systems with infinite horizon. Int. J. Control **83**, 1751–1757 (2010)
10. Hinderer, K.: Foundation of Non-Stationary Dynamic Programming with Discrete Time Parameter. Springer, New York (1970)
11. Bertsekas, D.P., Shreve, S.E.: Stochastic Optimal Control: the Discrete Time Case. Academic Press, Massachusetts (1978)