Adaptive Policies for Stochastic Systems under a Randomized Discounted Cost Criterion[*]

Juan González-Hernández[†] Raquiel R. López-Martínez[‡]
J. Adolfo Minjárez-Sosa[§]

Abstract

The paper deals with a class of discrete-time stochastic control processes under a discounted optimality criterion with random discount rate, and possibly unbounded costs. The state process $\{x_t\}$ and the discount process $\{\alpha_t\}$ evolve according to the coupled difference equations $x_{t+1} = F(x_t, \alpha_t, a_t, \xi_t)$, $\alpha_{t+1} = G(\alpha_t, \eta_t)$ where the state and discount disturbance processes $\{\xi_t\}$ and $\{\eta_t\}$ are sequences of i.i.d. random variables with *unknown* distributions θ^{ξ} and θ^{η} respectively. Assuming observability of the process $\{(\xi_t, \eta_t)\}$, we use the empirical estimator of its distribution to construct asymptotically discounted optimal policies.

Key Words: Empirical distribution; discrete-time stochastic systems; discounted cost criterion; random rate; optimal adaptive policy.
AMS 2000 subject classifications: 93E10, 93E20, 90C40.

[*] Work supported partially by Consejo Nacional de Ciencia y Tecnología (CONACyT) under Grant 46633.

[†] Departamento de Probabilidad y estadística, IIMAS-UNAM, A.P. 20-726, 01000 MéxicoD.F., MEXICO.

[‡] Facultad de Matemáticas UV, A.P. 270, 91090 Xalapa Ver., MEXICO.

[§] Departamento de Matemáticas, Universidad de Sonora, Rosales s/n, Centro, 83000 Hermosillo, Sonora, MEXICO.

1 Introduction

Among the main motivations to study a discounted optimality criterion in stochastic control problems are 1) the mathematical convenience (the discounted criterion is the best understood of all performance indices), and 2) its natural economic or financial interpretation (see, for instance, [35]). In both cases, the discount factor is typically assumed to be fixed or constant on the course of the process, which simplifies the mathematical analysis. However, from the point of view of applications, this assumption might be too strong or unrealistic. Indeed, in economic and financial models (see e.g., [2, 9, 14, 21, 28, 32, 35]), the discount factor is typically a function of interest rates, which in turn are random variables. In these cases, we have a time-varying random discount factor that can be represented as a stochastic process.

In this paper we consider a class of discrete-time stochastic control processes under a discounted optimality criterion with random discount rate. The state and discount processes evolve according to the difference equations:

$$x_{t+1} = F(x_t, \alpha_t, a_t, \xi_t), \tag{1}$$

$$\alpha_{t+1} = G(\alpha_t, \eta_t), \tag{2}$$

for $t = 0, 1, \ldots,$ where F and G are known continuous functions, x_t, α_t, and a_t are the state, the discount rate, and the control at time t, respectively. Moreover, the state and discount disturbance process $\{\xi_t\}$ and $\{\eta_t\}$ are observable sequences of independent and identically distributed (i.i.d.) random variables with *unknown* distributions θ^ξ and θ^η, respectively.

The actions applied by the controller at the decision times are selected according to rules known as control policies. The role of such policies is to minimize a discounted performance index with possibly unbounded one-stage cost and a random discount rate that varies as in (2). Clearly, this performance index depends on the unknown distributions θ^ξ and θ^η. Thus, to construct "minimizing" policies, the controller must combine control tasks with suitable statistical estimation methods of the joint distribution θ of the random variables ξ and η. The resulting policy of this procedure is called *adaptive.*

Our approach consists in estimating θ by means of the empirical distribution θ_t of the process $\{(\xi_t, \eta_t)\}$. This method is very general in the sense that θ can be arbitrary. However, since the discounted cost criterion depends strongly on the decision selected at first stages (precisely when the information about θ is deficient) we can not ensure, in general, the existence of an optimal adaptive policy. Thus, the discounted optimality will be analyzed in an asymptotic sense (see [33, 16]).

The discounted cost criterion in stochastic control problems has been widely studied under different approaches: dynamic programming (see, e.g., [16, 18, 19, 20, 24, 27, 29]); convex analysis (see, e.g., [5, 6, 7, 26]); linear programming (see, e.g., [15]); Lagrange multipliers (see, e.g., [25]); adaptive procedures (see, e.g., [3, 12, 16, 17, 22, 23]); see also [5, 6, 7, 34] for other variants. In these references, a fixed (non-random) discount factor is assumed. Recently, in [13], the discount criterion with a random discount factor was studied under the assumption that the components of the corresponding control model are known by the controller. In contrast, the main feature of this paper is that the distribution of the state and the discount disturbances are unknown.

The paper is organized as follows. In Section 2 we introduce the Markov control model we are concerned with. Next, in Section 3, we present the discounted optimality criterion with random discount rates. Section 4 contains the basic assumptions and some preliminary results on the discounted criterion and the estimation process. The construction of adaptive control policies together with our main results are introduced in Section 5 and proved in Section 6. Finally, in Section 7 we present a consumption-investment example to illustrate our assumptions and results.

Notation. Given a Borel space X (that is, a Borel subset of a complete and separable metric space) its Borel sigma-algebra is denoted by $\mathcal{B}(X)$, and "measurable", for either sets or functions, means "Borel measurable". Given a Borel space X, we denote by $I\!P(X)$ the family of probability measures on X, endowed with the weak topology. Let X and Y be Borel spaces. Then a stochastic kernel $\gamma(dx \mid y)$ on X given Y is a function such that $\gamma(\cdot \mid y)$ is a probability measure on X for each fixed $y \in Y$, and $\gamma(B \mid \cdot)$ is a measurable function on Y for each fixed $B \in \mathcal{B}(X)$.

2 Markov control model

The Markov control processes associated to the system (1)-(2) is specified by the elements

$$\mathcal{M} := (X, \Gamma, A, S_1, S_2, P_1, P_2, c) \tag{3}$$

satisfying the following conditions. The state space X, the action space A, and the state and discount disturbance spaces S_1 and S_2, respectively, are Borel spaces. The set $\Gamma := [\alpha^*, \infty)$, $\alpha^* > 0$, is the discount rate space. For each pair $(x, \alpha) \in X \times \Gamma$, $A(x, \alpha)$ is a nonempty Borel subset of A denoting the set of admissible controls when the system is in state x and a discount rate α is imposed. The set

$$\mathbb{K} = \{(x, \alpha, a) : x \in X, \alpha \in \Gamma, a \in A(x, \alpha)\} \tag{4}$$

of admissible state-discount-action triplets is assumed to be a Borel subset of the Cartesian product of X, Γ, and A. In addition, the transition law P_1, corresponding to (1), is a stochastic kernel on X given $\mathbb{K}$, that is, for all $t \geq 0$, $(x, \alpha, a) \in \mathbb{K}$ and $B \in \mathcal{B}(X)$,

$$\begin{aligned} P_1(B|x, \alpha, a) : &= \text{Prob}\left[F(x_t, \alpha_t, a_t, \xi_t) \in B | x_t = x, \alpha_t = \alpha, a_t = a\right] \\ &= \int_{S_1} 1_B \left(F(x, \alpha, a, s)\right) \theta^\xi(ds), \end{aligned} \tag{5}$$

where $F : X \times \Gamma \times A \times S_1 \to X$, the function in (1), is continuous, $1_B(\cdot)$ denotes the indicator function of the set B, and $\{\xi_t\}$ is a sequence of i.i.d. random variables in S_1 and common *unknown* distribution $\theta^\xi \in \mathbb{P}(S_1)$. Similarly, for all $t \geq 0$, $\alpha \in \Gamma$ and $D \in \mathcal{B}(\Gamma)$, the transition law P_2, corresponding to (2), is defined as:

$$\begin{aligned} P_2(D|\alpha) : &= \text{Prob}\left[G(\alpha_t, \eta_t) \in D | \alpha_t = \alpha\right] \\ &= \int_{S_2} 1_D \left(G(\alpha, s)\right) \theta^\eta(ds), \end{aligned} \tag{6}$$

where $G : \Gamma \times S_2 \to \Gamma$, the function in (2), is continuous, and $\{\eta_t\}$ is a sequence of i.i.d. random variables in S_2 (independent of the process $\{\xi_t\}$) with unknown distribution $\theta^\eta \in \mathbb{P}(S_2)$. Finally, the cost-per-stage $c(x, \alpha, a)$ is a measurable real-valued function on $\mathbb{K}$, possibly unbounded.

The control model $\mathcal{M}$ has the following interpretation. At stage t, the system is in the state $x_t = x \in X$ and the discount factor $\alpha_t = \alpha \in \Gamma$

is imposed. Then, the controller gets estimates θ_t^ξ and θ_t^η of the unknown distributions θ^ξ and θ^η, respectively, and combines these estimates with the history of the system to select a control $a = a_t(\theta_t^\xi, \theta_t^\eta) \in A(x,\alpha)$. As a consequence of this the following happens: 1) a cost $c(x,\alpha,a)$ is incurred, and 2) the system moves to a new state $x_{t+1} = x'$ and a new discount factor $\alpha_{t+1} = \alpha'$ is imposed according to the transition laws (5) and (6). Once the transition to state x' occurs, the process is repeated.

Control policies. We define the space of admissible histories up to time t by $\mathbb{H}_0 := X \times \Gamma$ and $\mathbb{H}_t := (\mathbb{K} \times S_1 \times S_2)^t \times X \times \Gamma$, $t \geq 1$. A generic element of $\mathbb{H}_t$ is written as $h_t = (x_0, \alpha_0, a_0, \xi_0, \eta_0, ..., x_{t-1}, \alpha_{t-1}, a_{t-1}, \xi_{t-1}, \eta_{t-1}, x_t, \alpha_t)$. A (randomized, history-dependent) control policy is a sequence $\pi = \{\pi_t\}$ of stochastic kernels π_t on A given $\mathbb{H}_t$ such that $\pi_t(A(x_t,\alpha_t) \mid h_t) = 1$, for all $h_t \in \mathbb{H}_t$, $t \geq 0$. We denote by Π the set of all control policies and by $\mathbb{F} \subset \Pi$ the set of all (deterministic) stationary policies. As usual, every stationary policy $\pi \in \mathbb{F}$ is identified with some measurable function $f : X \times \Gamma \to A$ such that $f(x,\alpha) \in A(x,\alpha)$ for every $(x,\alpha) \in X \times \Gamma$, taking the form $\pi = \{f, f, f, ...\} =: f$. In this case we use the notation

$$c(x,\alpha,f) := c(x,\alpha,f(x,\alpha)) \quad \text{and} \quad F(x,\alpha,f,s) := F(x,\alpha,f(x,\alpha),s)$$

for all $x \in X$, $\alpha \in \Gamma$ and $s \in S$.

3 Discounted criterion

We assume that the costs are exponentially discounted with accumulative random discounted rates. That is, a cost C incurred at stage t is equivalent to a cost $C\exp(-S_t)$ at time 0, where $S_t = \sum_{i=0}^{t-1} \alpha_i$ if $t \geq 1$, $S_0 = 0$. In this sense, when using a policy $\pi \in \Pi$, given the initial state $x_0 = x$ and the initial discount factor $\alpha_0 = \alpha$, we define the total expected discounted cost (with random discount rates) as

$$V(\pi,x,\alpha) := E_{(x,\alpha)}^\pi \left[\sum_{t=0}^\infty \exp(-S_t) c(x_t,\alpha_t,a_t) \right], \tag{7}$$

where $E_{(x,\alpha)}^\pi$ denotes the expectation operator with respect to the probability measure $P_{(x,\alpha)}^\pi$ induced by the policy π, given $x_0 = x$ and $\alpha_0 = \alpha$. (see, e.g., [4] for the construction of $P_{(x,\alpha)}^\pi$)

The optimal control problem associated to the control model $\mathcal{M}$, is then to find an optimal policy $\pi^* \in \Pi$ such that $V(\pi^*, x, \alpha) = V^*(x, \alpha)$ for all $(x, \alpha) \in X \times \Gamma$, where

$$V^*(x, \alpha) := \inf_{\pi \in \Pi} V(\pi, x, \alpha) \tag{8}$$

is the optimal value function.

Remark 3.1 *From (2), observe that* $\{\exp(-S_t)\}$ *is a sequence of random variables (not necessarily independent) representing the rate of discount at each stage* t. *Moreover, if* $\alpha_t = \alpha$ *for all* $t \geq 0$ *and some* $\alpha \in (0, \infty)$, *the performance index (7) reduces to the usual* $\beta-$*discounted cost criterion with* $\beta = \exp(-\alpha)$.

In the context of our work (unknown distributions θ^ξ and θ^η) we must combine suitable statistical estimation methods of θ^ξ and θ^η with control procedures in order to construct optimal policies. However, as the performance index (7) depends heavily on the controls selected at the first stages (precisely when the information about the distributions θ^ξ and θ^η is deficient), we can not ensure, in general, the existence of such policies. Thus, the optimality of policies constructed in this paper will be studied in the following asymptotic sense.

Definition 3.2 *A policy* $\pi \in \Pi$ *is said to be asymptotically discounted optimal for the control model* $\mathcal{M}$ *if*

$$\left| V^{(n)}(\pi, x, \alpha) - E^{\pi}_{(x,\alpha)} \left[V^*(x_n, \alpha_n) \right] \right| \to 0 \ \ \textit{as} \ \ n \to \infty, \ \textit{for all} \ (x, \alpha) \in X \times \Gamma,$$

where

$$V^{(n)}(\pi, x, \alpha) := E^{\pi}_{(x,\alpha)} \left[\sum_{t=n}^{\infty} \exp(-S_{n,t}) c(x_t, \alpha_t, a_t) \right] \tag{9}$$

is the total expected discounted cost from stage n *onward, and*

$$S_{n,t} = \sum_{k=n}^{t-1} \alpha_k \ \ \textit{for} \ \ t > n, \quad S_{n,n} = 0. \tag{10}$$

Clearly, discounted optimality implies asymptotic discounted optimality. The notion of asymptotic optimality was introduced by Schall [33] to study a problem of estimation and control in dynamic programming (see also [12, 16, 22]).

4 Assumptions and preliminary results

Observe that we can write the system (1)-(2) as

$$y_{t+1} = H(y_t, a_t, \chi_t), \quad t = 0, 1, \ldots,$$

where, letting $Y := X \times \Gamma$, $S := S_1 \times S_2$, $y_t^T := (x_t, \alpha_t)$, and $\chi_t := (\xi_t, \eta_t)$, $H : Y \times A \times S \to Y$ is a continuous function defined as

$$H(y_t, a_t, \chi_t) := (F(x_t, \alpha_t, a_t, \xi_t), G(\alpha_t, \eta_t))^T,$$

and $\{\chi_t\}$ is a sequence of i.i.d. $S-$valued random variables, defined on an underlying probability space $(\Omega, \mathcal{F}, P)$, with unknown common distribution $\theta(\cdot) = \theta^{\xi}(\cdot)\theta^{\eta}(\cdot)$. Thus

$$\theta(B) = P\left(\chi_t \in B\right),\ t \geq 0,\ B \in \mathcal{B}(S).$$

In the remainder, the probability space $(\Omega, \mathcal{F}, P)$ is fixed and $a.s.$ means *almost surely with respect to* P.

Now, for notational convenience, we put the control model $\mathcal{M}$ in the form

$$\left(Y, A, \{A(y) \subset A | y \in Y\}, Q, c\right),$$

where Q is the stochastic kernel on Y given $\mathbb{K}= \{(y, a) : y \in Y, a \in A(y)\}$ (see (4)) defined as

$$\begin{aligned} Q(B|y, a) :&= \text{Prob}\left[y_{t+1} \in B | y_t = y, a_t = a\right] \\ &= \int_S 1_B\left(H(y, a, s)\right)\theta(ds) \\ &= \theta\left(\{s \in S\ :\ H(y, a, s) \in B\}\right), \quad B \in \mathcal{B}(Y). \end{aligned}$$

We shall require two sets of assumptions. In the first one, Assumption 4.1, we impose continuity and compactness conditions to ensure the existence of minimizers and a solution to the optimality equation, while Assumption 4.2 are technical requirements to get a suitable estimation process of the distribution θ (see Remark 4.4(c) below). Note that Assumption 4.1(a) allows a unbounded one-stage cost function $c(y, a)$ provided that it is majorized by a "bounding" function W.

Assumption 4.1 *a) For each $y \in Y$, the set $A(y)$ is compact.*
b) For all $y \in Y$ the function $a \rightarrow c(y,a)$ is lower semi-continuous (l.s.c.) on $A(y)$. Moreover, there exists a continuous function $W : Y \rightarrow [1, \infty)$ and a constant M such that

$$\sup_{a \in A(y)} c(y,a) \leq MW(y), \quad y \in Y.$$

c) The function

$$\bar{v}(y,a) := \int\limits_S v(H(y,a,s))\theta(ds)$$

is continuous and bounded on $I\!\!K$ for every measurable bounded function v on Y.
d) There exist constants $p > 1$ and $\beta_0 < \infty$ satisfying $1 \leq \beta_0 < \exp(\alpha^*)$ such that for all $y \in Y$ and $a \in A(y)$,

$$W^p\left(H(y,a,\chi_0)\right) \leq \beta_0 W^p(y) \quad a.s. \tag{11}$$

In addition, Assumption c) holds when v is replaced with W.

An equivalent condition to relation (11) is that for all $y \in Y$ and $a \in A(y)$

$$W\left(H(y,a,\chi_0)\right) \leq \beta_0' W(y) \quad \text{a.s.},$$

for some $1 \leq \beta_0' < \exp(\alpha^*)$. However, for convenience we use Assumption4.1(d).

Assumption 4.2 *a) The family of functions*

$$\mathcal{V}_W := \left\{ \frac{V^*\left(H(y,a,.)\right)}{W(y)} : (y,a) \in \boldsymbol{K} \right\}$$

is equicontinuous on S, where V^* is the optimal value function (see(8)).
b) The function

$$\varphi(s) := \sup_{(y,a)} \left[W(y)\right]^{-1} W\left(H(y,a,s)\right)$$

is continuous on S.

Remark 4.3 *Clearly Assumption 4.2(a) holds if S is a countable set. In addition, the function φ in Assumption 4.2 might be non continuous. In such case we replace Assumption 4.2(b) by supposing the existence of a continuous majorant $\bar{\varphi}$ of φ such that $E\left[\bar{\varphi}(\chi_0)\right]^p < \infty$ (see 4.4(c) below).*

To estimate θ we use the empirical distribution $\{\theta_t\} \subset \mathbb{P}(S)$ of the disturbance process $\{\chi_t\}$, defined as follows. Let $\nu \in \mathbb{P}(S)$ be a given arbitrary probability measure. Then

$$\begin{aligned} \theta_0 &:= \nu, \\ \theta_t(B) &:= \frac{1}{t}\sum_{i=0}^{t-1} 1_B(\chi_i), \quad \text{for all } t \geq 1 \text{ and } B \in \mathcal{B}(S). \end{aligned}$$

Remark 4.4 *a) Observe that the inequality (11) implies, for all $(y,a) \in \mathbb{K}$,*

$$\int_S W^p\left(H(y,a,s)\right)\theta_t(ds) = \frac{1}{t}\sum_{i=0}^{t-1} W^p\left(H(y,a,\chi_i)\right) \leq \beta_0 W^p(y) \quad a.s., \tag{12}$$

which in turn yields [see Lemma 6.1 below]

$$\int_S W^p\left(H(y,a,s)\right)\theta(ds) \leq \beta_0 W^p(y).$$

b) It is well-known (See, e.g., [8]) the fact that θ_t converges weakly to θ a.s., that is,

$$\int u d\theta_t \to \int u d\theta \quad a.s. \quad as\ t \to \infty,$$

for every real-valued, continuous and bounded function u on S.
c) Furthermore, from Assumption 4.1(d)

$$E\left[\varphi(\chi_0)\right]^p < \infty.$$

Thus, from Assumption 4.2, using the fact that $V^*(y) \leq CW(y)$ (see Proposition 4.5 below), and applying Theorem 6.4 in [30], we get

$$D_t \to 0 \quad a.s., \quad as \quad t \to \infty, \tag{13}$$

where

$$D_t := \sup_{(y,a)\in\mathbb{K}} \left| \int_S \frac{V^*(H(y,a,s))}{W(y)}\theta_t(ds) - \int_S \frac{V^*(H(y,a,s))}{W(y)}\theta(ds) \right|. \tag{14}$$

We denote by $\mathbb{B}_W$ the normed linear space of all measurable functions $u : Y \to \Re$ with a finite norm $\|u\|_W$ defined as

$$\|u\|_W := \sup_{y \in Y} \frac{|u(y)|}{W(y)}. \tag{15}$$

A first consequence of Assumption 4.1, which is stated in [13], is the following proposition.

Proposition 4.5 *Suppose that Assumption 4.1 holds. Then $V^* \in \mathbb{B}_W$, that is, there exists a constant $C > 0$ such that*

$$V^*(y) \leq CW(y) \quad \textit{for all} \;\; y \in Y. \tag{16}$$

In addition, V^* satisfies the optimality equation

$$V^*(y) = \inf_{a \in A(y)} \left(c(y,a) + \exp(-\alpha) \int_S V^*(H(y,a,s))\theta(ds) \right), \quad \forall y \in Y. \tag{17}$$

5 Main results

Let $\{V_t\}$ be a sequence of functions in $\mathbb{B}_W$ defined as $V_0 \equiv 0$, and for $t \geq 1$,

$$V_t(y) = \inf_{a \in A(y)} \left(c(y,a) + \exp(-\alpha) \int_S V_{t-1}(H(y,a,s))\theta_t(ds) \right), \quad y \in Y. \tag{18}$$

A straightforward calculation shows that for some constant C',

$$V_t(y) \leq C'W(y) \quad \text{a.s,} \quad y \in Y, \;\; t \geq 0. \tag{19}$$

Now, applying standard arguments on the existence of minimizers (see, e.g., [31]), under Assumption 4.1 and the continuity of H, for each $t > 0$ and $\delta_t > 0$, there exists $f_t \in \mathbb{F}$ such that

$$c(y, f_t) + \exp(-\alpha) \int\limits_S V_{t-1}(H(y, f_t, s))\theta_t(ds) \leq V_t(y) + \delta_t \quad \text{a.s.} \;\; y \in Y. \tag{20}$$

Definition 5.1 *Let $\{\delta_t\}$ be an arbitrary sequence of positive numbers such that $\delta_t \to 0$ as $t \to \infty$, and let $\{f_t\}$ be a sequence of functions in $\mathbb{F}$ satisfying (20). We define the policy $\hat{\pi} = \{\hat{\pi}_t\}$ as*

$$\hat{\pi}_t(h_t) = \hat{\pi}_t(h_t; \theta_t) := f_t(y_t), \;\; t > 0,$$

and $\hat{\pi}_0(y)$ is any fixed action in $A(y)$.

We can state our main results as follows:

Theorem 5.2 *Under Assumptions 4.1 and 4.2, we have*
a) $\|V_t - V^*\|_W \to 0$ *a.s., as* $t \to \infty$;
b) The policy $\hat{\pi}$ *is asymptotically discount optimal.*

6 Proofs

The proof of Theorem 5.2 is based on the following results.

Lemma 6.1 *Suppose that Assumption 4.1 holds. Then:*
a) For all $y \in Y$ *and* $a \in A(y)$,

$$\int_S W^p\left(H(y,a,s)\right)\theta(ds) \le \beta_0 W^p(y). \tag{21}$$

b) For all $y \in Y, a \in A(y)$*, and* $t > 0$,

$$\int_S W\left(H(y,a,s)\right)\theta_t(ds) \le \beta W(y) \quad a.s. \tag{22}$$

and

$$\int_S W\left(H(y,a,s)\right)\theta(ds) \le \beta W(y), \tag{23}$$

where $\beta := \beta_0^{1/p}$.
c) For all $y \in Y$ *and* $\pi \in \Pi$*, we have*

$$\sup_{t>0} E_y^{\pi}\left[W^p(y_t)\right] < \infty \quad and \quad \sup_{t>0} E_y^{\pi}\left[W(y_t)\right] < \infty.$$

Proof: It is clear that the part (a) follows from Assumption4.1(d). Next, the part (b) follows from the relations (12) and (21), and applying Jensen's inequality, while part (c) is a consequence of (21) and (23) (see details in [10, 11, 12, 19, 22]).■

We also need the following characterization of asymptotic optimality (see Definition 3.2) which is an adaptation of Theorem 4.6.2 in [18] (see also [33]) to our context (randomized discounted cost criterion).

Lemma 6.2 *A policy $\pi \in \Pi$ is asymptotically discount optimal for the control model $\mathcal{M}$ if, for all $y \in Y$,*

$$E_y^\pi \left[\Phi(y_t, a_t)\right] \to 0 \text{ as } t \to \infty,$$

where

$$\Phi(y,a) := c(y,a) + \exp(-\alpha) \int_S V^*\left(H(y,a,s)\right)\theta(ds) - V^*(y), \; (y,a) \in I\!\!K. \tag{24}$$

(Note that, by (17), Φ is nonnegative.)

Proof: Observe that for each $\pi \in \Pi$, $y \in Y$, and $t \geq 0$,

$$\Phi(y_t, a_t) = E_y^\pi \left[c(y_t, a_t) + \exp(-\alpha_t)V^*(y_{t+1}) - V^*(y_t)|h_t, a_t\right],$$

where h_t represent the history of the system up to time t (see definition of control policies). Hence, from definition (9) and (10), using the fact $\exp(-S_{n,t})\exp(-\alpha_t) = \exp(-S_{n,t+1})$, and applying the properties of conditional expectation, we have, for each $n \geq t$, $\pi \in \Pi$, and $y \in Y$

$$\begin{aligned}
& \sum_{t=n}^{\infty} E_y^\pi \left[\exp(-S_{n,t})\Phi(y_t, a_t)\right] \\
= \; & \sum_{t=n}^{\infty} E_y^\pi \left[\exp(-S_{n,t}) E_y^\pi \left[c(y_t, a_t) + \exp(-\alpha_t)V^*(y_{t+1}) - V^*(y_t)|h_t, a_t\right]\right] \\
= \; & \sum_{t=n}^{\infty} E_y^\pi \left[\exp(-S_{n,t})c(y_t, a_t)\right] + \sum_{t=n}^{\infty} E_y^\pi \left[\exp(-S_{n,t+1})V^*(y_{t+1}) - \exp(-S_{n,t})V^*(y_t)\right] \\
= \; & V^{(n)}(\pi, y) - E_y^\pi \left[V^*(y_n)\right] + \lim_{m\to\infty} E_y^\pi \left[\exp(-S_{n,m})V^*(y_m)\right] \qquad (25) \\
= \; & V^{(n)}(\pi, y) - E_y^\pi \left[V^*(y_n)\right],
\end{aligned}$$

where the last equality follows from Holder's inequality, Lemma 6.1(c), (16), the fact $\alpha^* \leq \alpha_t$, $t \geq 0$, and the following relation

$$\begin{aligned}
\lim_{m\to\infty} E_y^\pi \left[\exp(-S_{n,m})V^*(y_m)\right] & \leq \lim_{m\to\infty} \left(E_y^\pi \left[\exp(-p'S_{n,m})\right]\right)^{1/p'} \left(E_y^\pi [V^*(y_m)]^p\right)^{1/p} \\
& \leq \lim_{m\to\infty} C(E_y^\pi[W(y_m)]^p)^{1/p}(\exp(-p'\alpha^*(m-n)))^{1/p'} \\
& \leq CM \lim_{m\to\infty} (\exp(-p'\alpha^*(m-n)))^{1/p'} \\
& = 0.
\end{aligned}$$

(See Lemma 6.1 (c) for constant M).
Finally, since the limit

$$\lim_{t\to\infty} E_y^{\pi}\left[\Phi(y_t,a_t)\right]=0$$

implies

$$\lim_{n\to\infty}\sum_{t=n}^{\infty} E_y^{\pi}\left[\exp(-S_{n,t})\Phi(y_t,a_t)\right]=0$$

then, the relation (25) yields the desired result.■

We define the operators

$$\begin{aligned} Tu(y) &:= \inf_{a\in A(y)}\left\{c(y,a)+\exp(-\alpha)\int_S u\left(H(y,a,s)\right)\theta(ds)\right\}, \\ T_tu(y) &:= \inf_{a\in A(y)}\left\{c(y,a)+\exp(-\alpha)\int_S u\left(H(y,a,s)\right)\theta_t(ds)\right\}, \end{aligned}$$

for all $y\in Y$ and $u\in\mathbb{B}_W$. Observe that from Assumption 4.1 and Lemma 6.1, T and T_t map $\mathbb{B}_W$ to itself. In addition, following similar ideas of Proposition 8.3.9 in [19], we have that T and T_t are contraction operators with modulus $\gamma:=\beta_0\exp(-\alpha^*)<1$ (see Assumption 4.1(d)), respect to the norm $\|\cdot\|_W$. That is, for all $u,v\in\mathbb{B}_W$,

$$\|Tv-Tu\|_W\leq\gamma\,\|v-u\|_W$$

and

$$\|T_tv-T_tu\|_W\leq\gamma\,\|v-u\|_W \text{ a.s.}$$

6.1 Proof of Theorem 5.2

a) From (16)-(19), $V^*,V_t\in\mathbb{B}_W,\ t>0$,

$$TV^*=V^* \qquad \text{and} \qquad T_tV_{t-1}=V_t \text{ a.s. } \forall t>0. \tag{26}$$

Hence

$$\begin{aligned} \|V^*-V_t\|_W &\leq \|TV^*-T_tV^*\|_W+\|T_tV^*-T_tV_{t-1}\|_W \\ &\leq \|TV^*-T_tV^*\|_W+\gamma\,\|V^*-V_{t-1}\|_W \quad \text{a.s.} \end{aligned} \tag{27}$$

Now, from definition of T and T_t, and (14),

$$\begin{aligned}\|TV^* - T_tV^*\|_W &\leq \sup_{(y,a)\in IK} \left| \int_S \frac{V^*(H(y,a,s))}{W(y)} \theta_t(ds) - \int_S \frac{V^*(H(y,a,s))}{W(y)} \theta(ds) \right| \\ &= D_t \quad \text{a.s.} \qquad (28)\end{aligned}$$

Combining (27) and (28), we have,

$$\|V^* - V_t\|_W \leq D_t + \gamma \|V^* - V_{t-1}\|_W \quad a.s. \qquad (29)$$

Finally, denoting $l := \limsup_{t\to\infty} \|V^* - V_t\|_W < \infty$ (see (16) and (19)) and taking limsup on both sides of (29), from (13) we obtain $l \leq \gamma l$, which implies (since $0 < \gamma < 1$) that $l = 0$. This proves the part (a).

b) For each $t > 0$, we define the function $\Phi_t : \mathbb{K} \to \mathbb{R}$ (see(24)) as

$$\Phi_t(y,a) := c(y,a) + \exp(-\alpha) \int_S V_{t-1}(H(y,a,s))\, \theta_t(ds) - V_t(y).$$

We also define, for each $t > 0$,

$$\Psi_t := \sup_{y\in Y} [W(y)]^{-1} \sup_{a\in A(y)} |\Phi(y,a) - \Phi_t(y,a)|. \qquad (30)$$

Observe that from definition of the policy $\hat{\pi}$, we have (see (20)) $\Phi_t(., \hat{\pi}_t(.)) \leq \delta_t$, for each $t > 0$. Hence,

$$\begin{aligned}\Phi(y_t, \hat{\pi}_t(h_t)) &\leq |\Phi(y_t, \hat{\pi}_t(h_t)) - \Phi_t(y_t, \hat{\pi}_t(h_t)) + \delta_t| \\ &\leq \sup_{a\in A(y_t)} |\Phi(y_t,a) - \Phi_t(y_t,a)| + \delta_t \\ &\leq W(y_t)\Psi_t + \delta_t \quad \text{a.s.}\end{aligned}$$

Therefore, according to Lemma 6.2, to prove asymptotic optimality of $\hat{\pi}$, it is sufficient to show that

$$E_y^{\hat{\pi}}(W(y_t)\Psi_t) \to 0 \quad \text{as } t \to \infty. \qquad (31)$$

By adding and subtracting the term $\exp(-\alpha)\int_S V^*(H(y,a,s))\, \theta_t(ds)$, we have, for each $(y,a) \in \mathbb{K}$ and $t > 0$,

$$\begin{aligned}|\Phi_t(y,a) - \Phi(y,a)| &\leq |V^*(y) - V_t(y)| \\ &\quad + \exp(-\alpha) \int_S |V_{t-1}(H(y,a,s)) - V^*(H(y,a,s))|\, \theta_t(ds) \\ &\quad + \exp(-\alpha) \left| \int_S V^*(H(y,a,s))\, \theta_t(ds) - \int_S V^*(H(y,a,s))\, \theta(ds) \right|,\end{aligned}$$

which, from Lemma 6.1(a), (b), and definitions of the norm $\|\cdot\|_W$ and D_t (see (14)), implies

$$\frac{|\Phi_t(y,a,)-\phi(y,a)|}{W(y)} \leq \|V^* - V_t\|_W \\ + \beta \exp(-\alpha)\|V^* - V_{t-1}\|_W \; + D_t \quad \text{a.s.} \tag{32}$$

Thus, from Theorem 5.2(a) and (13),

$$\Psi_t \to 0 \;\; \text{a.s.,} \;\; \text{as } t \to \infty. \tag{33}$$

Now observe that from (32), (13), (16), and (19), $\sup_{t>0} \Psi_t \leq M_1 < \infty$ for some constant M_1. In addition, from (33) we have the convergence in probability

$$\Psi_t \xrightarrow{P_y^{\hat{\pi}}} 0 \quad \text{as } t \to \infty, \tag{34}$$

whereas from Lemma 6.1(c)

$$\sup_{t>0} E_y^{\hat{\pi}} \left(W(y_t)\Psi_t\right)^p \leq M_1^p \sup_{t>0} E_y^{\hat{\pi}} \left(W^p(y_t)\right) < \infty.$$

This implies (see, for instance, Lemma 7.6.9 in [1]) that $\{W(y_t)\Psi_t\}$ is $P_y^{\hat{\pi}}$-uniformly integrable.

On the other hand, for arbitrary positive numbers l_1 and l_2, we have, for $t > 0$,

$$P_y^{\hat{\pi}}\left(W(y_t)\Psi_t > l_1\right) \leq P_y^{\hat{\pi}}\left(\Psi_t > \frac{l_1}{l_2}\right) + P_y^{\hat{\pi}}\left(W(y_t) > l_2\right),$$

which, applying Chebyshev's inequality, yields

$$P_y^{\hat{\pi}}\left(W(y_t)\Psi_t > l_1\right) \leq P_y^{\hat{\pi}}\left(\Psi_t > \frac{l_1}{l_2}\right) + \frac{E_y^{\hat{\pi}}\left(W(y_t)\right)}{l_2}. \tag{35}$$

Now, (35) together with Lemma 4.4(c) and (34), implies the convergence in probability

$$W(y_t)\Psi_t \xrightarrow{P_y^{\hat{\pi}}} 0 \quad \text{as } t \to \infty. \tag{36}$$

Finally, (31) holds from (36) and the fact that $\{W(y_t)\Psi_t\}$ is $P_y^{\hat{\pi}}$-uniformly integrable.■

7 Example

We consider an infinite horizon consumption-investment problem where an investor musts allocate his/her current wealth x_t between investment (a_t) and consumption $(x_t - a_t)$, in each stage $t = 0, 1, 2, ...$ In addition, in each stage t, a discount factor $\exp(-\alpha_t)$ is imposed, which depends upon the current bank interest rate.

The state and action spaces are $X = A = [0, \infty)$, and assuming that borrowing is not allowed, the investment constraint set (i.e., the set of admissible controls) takes the form $A(x, \alpha) = [0, x]$. Moreover, we suppose that the bank receives at least an interest rate of $\exp(\alpha^*) - 1$, for some $\alpha^* > 0$. In this sense, the discount rate space is $\Gamma = [\alpha^*, \infty)$.

The state process $\{x_t\}$ and the discount process $\{\alpha_t\}$ evolve according to the coupled difference equations

$$x_{t+1} = a_t \rho(\xi_t), \qquad \alpha_{t+1} = h\alpha_t + \eta_t, \qquad t = 0, 1, 2, ...,$$

(x_0, α_0) given, where $h > 0$, $\{\xi_t\}$ and $\{\eta_t\}$ are independent sequences of i.i.d. random variables, and independent of (x_0, α_0), having a discrete distribution with values in S_1 and S_2, respectively. In addition, $\rho : S_1 \longrightarrow (0, \gamma]$ is a given measurable function with $1 \leq \gamma < \exp(\alpha^*)$.

The one-stage cost $c(x, \alpha, a)$ is an arbitrary nonnegative measurable function, which is l.s.c. in a, and satisfying

$$\sup_{a \in A(x,\alpha)} c(x, \alpha, a) \leq M \left(\bar{b}x + 1\right)^{1/p}, \quad (x, \alpha) \in X \times \Gamma, \tag{37}$$

for some $\bar{b} > 0$, $M > 0$, and $p > 1$.

Clearly, the Assumptions 4.1(a), (b) and 4.2 are satisfied, by taking $W(y) = W(x, \alpha) = \left(\bar{b}x + 1\right)^{1/p}$ and from Remark 4.3. We get Assumption 4.1(d) from the following relations: for all $y = (x, \alpha) \in Y = X \times \Gamma$, and $a \in A(y) = [0, x]$,

$$\begin{aligned} W^p[H(y, a, \chi_0)] &= \bar{b}a\rho(\xi_0) + 1 \\ &\leq \bar{b}x\gamma + 1 \leq \bar{b}x\gamma + \gamma \\ &= \gamma(\bar{b}x + 1) \\ &= \beta_0 W^p(y), \end{aligned}$$

where $\beta_0 := \gamma$.

Finally, Assumption 4.1(c) follows from Example C.6, Appendix C in [18].

Remark 7.1 *Usually, in a consumption-investment problem where the objective is to maximize a randomized discounted reward criterion, a utility function r is considered as the one-stage return. In particular, if we take*

$$r(x, \alpha, a) = b\sqrt{x-a}, \quad (x, \alpha) \in Y = X \times \Gamma, \ a \in A(x, \alpha),$$

the relation (37) is satisfied with the function r instead of c.

References

[1] Ash, R. B. (1972). *Real Analysis and Probability.* Academic Press, New York.

[2] Berument, H., Kilinc, Z.and Ozlale, U. (2004). The effects of different inflation risk premiums on interest rate spreads. *Physica A* **333,** 317–324.

[3] Cavazos-Cadena, R. (1990). Nonparametric adaptive control of discounted stochastic systems with compact space. *J. Optim. Theory Appl.,* **65,** 191-2007.

[4] Dynkin, E. B. and Yushkevich, A. A. (1979). *Controlled Markov Processes.* Springer-Verlag, New York.

[5] Feinberg, E. A.and Shwartz, A. (1994.) Markov decision models with weighted discounted criteria. *Math. Oper. Res.* **19**, 152–168.

[6] Feinberg, E. A.and Shwartz, A. (1995). Constrained Markov decision models with weighted discounted rewards. *Math. Oper. Res.* **20**, 302–320.

[7] Feinberg, E. A.and Shwartz, A. (1999). Constrained dynamic programming with two discount factors: applications and an algorithm. *IEEE Trans. Autom. Control* **44,** 628–631.

[8] Gaenssler, P. and Stute, W. (1979). Empirical processes: a survey for i.i.d. random variables. *Ann. Probab.* **7,** 193–243.

[9] Gil-Alana, L. A. (2004). Modelling the U. S. interest rate in terms of I(d) statistical model. *The Quarterly Review of Economics and Finance* **44,** 475–486.

[10] Gordienko, E. I. and Hernández-Lerma, O. (1995). Average cost Markov control processes with weighted norms: existence of canonical policies. *Appl. Math.* (Warsaw) **23,** 199-218.

[11] Gordienko, E. I. and Hernández-Lerma, O. (1995). Average cost Markov control processes with weighted norms: value iteration. *Appl. Math.*(Warsaw) **23,** 219-237.

[12] Gordienko, E. I. and Minjárez-Sosa, J. A. (1998). Adaptive control for discrete-time Markov processes with unbounded costs: discounted criterion. *Kybernetika* **34,** 217–234.

[13] González-Hernández, J. and López-Martínez, R. R. and Pérez-Hernández, R. (Markov control processes with randomized discounted cost in Borel space. *Math. Meth. Oper. Res.* to appear.

[14] Haberman, S. and Sung, J. (2005). Optimal pension funding dynamics over infinite control horizon when stochastic rates of return are stationary. *Insurance: Mathematics and Economics* **36,** 103–116.

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

[16] Hernández-Lerma, O. (1989). *Adaptive Markov Control Processes.* Springer-Verlag, New York.

[17] Hernández-Lerma, O. and Cavazos-Cadena, R. (1990). Density estimation and adaptive control of Markov processes: average and discounted criteria. *Acta Appl. Math.* **20,** 285–307.

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

[19] Hernández-Lerma, O.and Lasserre, J. B. (1999). *Further Topics on Discrete-Time Markov Control Processes.* Springer-Verlag, New York.

[20] Hernández-Lerma, O. and Muñoz-de-Ozak, M. (1992). Discrete-time Markov control processes with discounted unbounded cost: optimality criteria. *Kybernetika (Prague)* **28,** 191–212.

[21] Lee, P. and Rosenfield, D. B. (2005). When to refinance a mortgage: a dynamic programming approach. *European Journal of Operational Research* **166,** 266–277.

[22] Hilgert, N. and Minjárez-Sosa, J. A. (2001). Adaptive policies for time-varying stochastic systems under discounted criterion. *Math. Methods Oper. Res.* **54,** 491–505.

[23] Hilgert, N. and Minjárez-Sosa, J. A. (2006). Adaptive control of stochastic systems with unknown disturbance distribution: discounted criteria. *Math. Methods Oper. Res.* To appear.

[24] Kurano, M. (1998). Controlled Markov set-chains with discounting. *J. Appl. Prob.* **35,** 293–302.

[25] López-Martínez, R. R. and Hernández-Lerma, O. (2003). The Lagrange approach to constrained Markov processes: a survey and extension of results. *Morfismos* **7,** 1–26.

[26] Mao, X. and Piunovskiy, A. B. (2000). Strategic measure in optimal control problems for stochastic sequences. *Sthochastic Anal. Appl.* **18,** 755–776.

[27] Michael, K. Ng. (1999). A note on policy algorithms for discounted Markov decision problems. *Oper.Res.letters* **25,** 195–197.

[28] Newell, R. G.and Pize, W. A. (2003). Discounting the distant future: how much do uncertain rates increase valuation?. *Journal of Environmental Economic and Management* **46,** 52–71.

[29] Puterman, M. L. (1994. *Markov Decision Processes: Discrete Stochastic Dynamic Programming.* Wiley, New york.

[30] Ranga Rao, R. (1962). Relations between weak and uniform convergence of measures with applications. *Ann. Math. Statistics* **33,** 659–680.

[31] Rieder, U. (1978). Measurable selection theorems for optimization problems. *Manuscripta Math.* **24,** 115–131.

[32] Sack, B. and Wieland, V. (2000). Interest-rate smooothing and optimal monetary policy: A review of recent empirical evidence. *Journal of Economics and Business* **52,** 205–228.

[33] Schäl, M. (1987.) Estimation and control in discounted stochastic dynamic programming. *Stochastics* **20,** 51–71.

[34] Shwartz, A. (2001). Death and discounting. *IEEE Trans. Autom. Control* **46,** 628–631

[35] Stockey, N. L., Lucas, R. E. Jr. (1989). *Recursive Methods in Economic Dynamics.* Harvard University Press, Cambridge, MA.

[36] Van Nunen, J. A. E. E. and Wessels, J. (1978). A note on dynamic programming with unbounded rewards. *Manag. Sci.* **24,** 576–580.