International Journal of Pure and Applied Mathematics

Volume 28 No. 2 2006, 265-280

POINTWISE APPROXIMATIONS OF DISCOUNTED MARKOV DECISION PROCESSES TO OPTIMAL POLICIES

Daniel Cruz-Suárez[1], Raúl Montes-de-Oca[2 §], Francisco Salem-Silva[3]

[1]División Académica de Ciencias Básicas
Universidad Juárez Autónoma de Tabasco
P.O. Box 5, Cunduacán, Tabasco, 86690, MEXICO
e-mail: daniel.cruz@basicas.ujat.mx

[2]Departamento de Matemáticas
Universidad Autónoma Metropolitana-Iztapalapa
186 San Rafael Atlixco Avenue
Vicentina, México D.F., 09340, MEXICO
e-mail: momr@xanum.uam.mx

[3]Facultad de Ciencias Físico Matemáticas
Benemérita Universidad Autónoma de Puebla
San Claudio y Rio Verde Avenue
San Manuel, CU, Puebla City, 72570, MEXICO
e-mail: fsalem@fcfm.buap.mx

Abstract: This paper deals with discrete-time Markov decision processes with Borel state and control spaces, with possibly unbounded costs and compact control constraint sets, and the expected total discounted cost criterion. Conditions that allow to detect a value iteration policy which is a pointwise approximation to the optimal policy are given. Besides, two illustrative examples are supplied.

AMS Subject Classification: 90C40
Key Words: discounted Markov decision process, optimality equation, value iteration algorithm, uniqueness of the optimal policy, approximation to the optimal policy

Received: April 6, 2006

§Correspondence author

1. Introduction

This paper deals with discrete-time Markov decision processes (MDPs) (see Bertsekas and Shreve [2], Dynkin and Yushkevich [6], Hernández-Lerma and Lasserre [9], Hinderer [10], and Yushkevich [12]).

MDPs for which the state space X and the control space A are both Borel spaces will be considered. Moreover, $A(x)$, the admissible control set, is assumed to be compact for each $x \in X$. The cost function c is nonnegative and (possibly) unbounded. The expected total discounted cost is used as the objective function.

For the MDPs studied in this article, the existence of a unique stationary policy f^* is assumed (see Cruz-Suárez et al [4] for conditions which ensure the uniqueness of optimal policies of discounted MDPs).

Consider a Markov decision process with the description from the previous paragraphs. Denote by V^* the optimal value function, and for each $n = 1, 2, \cdots$, let V_n and f_n, denote the minimum and the minimizer corresponding to the step n of the value iteration (or successive approximation) algorithm (see Hernández-Lerma and Lasserre [9]), respectively.

In the paper conditions (see Assumption 2.1, Assumption 2.2, 2.6 and 2.8 below) that imply the pointwise convergence at certain geometric rate of the sequence $\{V_n\}$ to V^* (see Hernández-Lerma and Lasserre [9]), and the pointwise convergence of the sequence $\{f_n\}$ to f^* (see Cruz-Suárez and Montes-de-Oca [3]) are assumed.

Here a new result is established (see Section 3 below) which consists in providing an algorithm and its corresponding stopping rule which allows to obtain, for each $x \in X$ and $\varepsilon > 0$, an $\varepsilon-$approximation to $f^*(x)$.

Specifically, given $x \in X$ and $\varepsilon > 0$, the existence and a way to detect a positive integer N (which may depend on x and ε) will be considered, such that the minimizer $f_N(x)$ differs from $f^*(x)$ in less than ε with respect to the metric of A. In this sense, $f_N(x)$ is an $\varepsilon-$approximation to $f^*(x)$.

This approach to approximate an optimal policy is different to the one usually considered in the literature of MDPs where the approximation to f^* is measured by means of the difference of the expected total discounted cost using f_N, denoted by $V(f_N, x)$, and $V^*(x), x \in X$ (see Hernández-Lerma and Lasserre [9]).

Two examples to illustrate the theory developed are provided below. One of them is an inventory/production model (see Example 8.2 in Yushkevich [12]), and the other example is the linear quadratic model (see Bertsekas [1]).

The paper is organized as follows. In Section 2, the basic theory of MDPs

and some assumptions will be provided. In Section 3, the main result which deals with obtaining a pointwise approximation to f^* will be established. In Section 4, the examples will be presented.

2. Preliminaries

Let $(X, A, \{A(x) : x \in X\}, Q, c)$ be a discrete-time, stationary Markov decision model (see Bertsekas and Shreve [2], Dynkin and Yushkevich [6], Hernández-Lerma and Lasserre [9], and Hinderer [10] for notation and terminology) where X is the state space, the space A is the control or action set. Both X and A are assumed to be Borel spaces (i.e. nonempty measurable subsets of complete and separable metric spaces). $A(x) \subset A$ is the measurable subset of admissible actions for each state $x \in X$, Q is the controlled transition law, i.e. $Q(\cdot \mid x, a)$ is a probability measure on X for every $(x, a) \in \mathbb{K} := \{(y, b) \mid y \in X, b \in A(y)\}$, and $Q(B \mid \cdot)$ is a measurable function on $\mathbb{K}$ for every measurable set $B \subset X$, and c is the cost function which is real-valued and defined on $\mathbb{K}$. This model has the following interpretation: at each epoch $t = 0, 1, 2, \cdots$, the state $x_t = x \in X$ of a dynamical system is observed and an action $a_t = a \in A(x)$ is chosen by the controller. As a consequence, a cost $c(x, a)$ is earned and the system moves to state $x_{t+1} = y \in X$ with probability $Q(\cdot \mid x, a)$.

A control policy π is a (measurable, possibly randomized) rule for choosing actions, and at each $t = 0, 1, \cdots$, the control prescribed by π may depend on the current state as well as on the history of previous states and actions. Denote the set of all policies by Π. Given the initial state $x_0 = x$, any policy π defines the unique probability distribution of the state-action process $\{(x_t, a_t)\}$; for details see, for instance, Hinderer [10]. Denote this probability distribution by P_x^π, and the corresponding expectation operator by E_x^π. Let $\mathbb{F}$ be the set of all measurable functions $f : X \to A$, such that $f(x) \in A(x)$ for every $x \in X$. A policy $\pi \in \Pi$ is *stationary* if there exists $f \in \mathbb{F}$ such that, under π, the control $f(x_t)$ is applied at each time $t = 0, 1, \cdots$. The set of all stationary policies is taken as $\mathbb{F}$.

The *expected total discounted cost* is defined as

$$V(\pi, x) = E_x^\pi[\sum_{t=0}^{\infty} \alpha^t c(x_t, a_t)], \tag{1}$$

when using the policy $\pi \in \Pi$, given the initial state $x \in X$. In (1), the number $\alpha \in (0, 1)$ is the discount factor.

The optimal control problem is to find a policy π^* that minimizes the given performance criterion. Thus, a policy π^* is said to be *optimal* if

$$V(\pi^*, x) = V^*(x),$$

$x \in X$, where

$$V^*(x) := \inf_{\pi \in \Pi} V(\pi, x), \tag{2}$$

$x \in X$, is the so-called *optimal value* function.

Now some assumptions and results to be used in the next sections will be listed.

Assumption 2.1. a. The one-stage cost c is nonnegative, and lower semicontinuous (l.s.c.) on $\mathbb{K}$.

b. The control-constraint set $A(x)$ is compact for every $x \in X$.

c. The transition law Q is strongly continuous, i.e.

$$\mu'(x,a) := \int \mu(y)\, Q(dy \mid x, a)$$

is continuous and bounded on $\mathbb{K}$, for every measurable bounded function $\mu : X \to \mathbb{R}$.

Assumption 2.2. There exist nonnegative constants $\overline{c}$ and β, with $1 \leq \beta < 1/\alpha$, and a weight function $w : X \to \mathbb{R}$, $w \geq 1$, such that for every state $x \in X$:

a. $\sup_{a \in A(x)} |c(x,a)| = \sup_{a \in A(x)} c(x,a) \leq \overline{c} w(x)$ (recall that in this paper c is assumed to be nonnegative).

b. $\sup_{a \in A(x)} \int w(y)\, Q(dy \mid x, a) \leq \beta w(x)$.

c. The function $w'(x,a) := \int w(y)\, Q(dy \mid x, a)$ is continuous in $a \in A(x)$.

Remark 2.3. Assumption 2.2 holds trivially when the cost function c is bounded. In fact, if $M > 0$ satisfies that

$$0 \leq c(x,a) \leq M, \tag{3}$$

for each $(x,a) \in \mathbb{K}$, then Assumption 2.2 holds taking $\beta = \overline{c} = 1$ and $w(x) = M, x \in X$ (observe that it is possible to assume, without loosing generality, that $M \geq 1$).

The *value iteration* functions are defined as

$$V_n(x) = \min_{a \in A(x)} \left[c(x,a) + \alpha \int V_{n-1}(y)\, Q(dy \mid x, a) \right], \tag{4}$$

for all $x \in X$ and $n = 1, 2, \cdots$, with $V_0(\cdot) = 0$. Using Assumption 2.1, it is possible to demonstrate (see Hernández-Lerma and Lasserre [9]) that, for each $n = 1, 2, \cdots$, there exists a stationary policy $f_n \in \mathbb{F}$ such that the minimum in (4) is attained, i.e.

$$V_n(x) = c(x, f_n(x)) + \alpha \int V_{n-1}(y) Q(dy \mid x, f_n(x)), \tag{5}$$

$x \in X$.

Lemma 2.4. (see [9], Theorem 8.3.6) *If Assumption 2.1 and Assumption 2.2 hold, then the optimal value function V^* defined in (2) satisfies the optimality equation, i.e. for all $x \in X$:*

$$V^*(x) = \min_{a \in A(x)} \left[c(x, a) + \alpha \int V^*(y) Q(dy \mid x, a) \right]. \tag{6}$$

There is also $f^* \in \mathbb{F}$ such that

$$V^*(x) = c(x, f^*(x)) + \alpha \int V^*(y) Q(dy \mid x, f^*(x)), \tag{7}$$

$x \in X$, and f^* is optimal; conversely, if $g \in \mathbb{F}$ is optimal, then it satisfies (7).

Moreover, for each $x \in X$ and $n = 1, 2, \cdots$,

$$|V_n(x) - V^*(x)| \leq O(n, x) := \overline{c}(\alpha\beta)^n w(x) / (1 - \alpha\beta). \tag{8}$$

Remark 2.5. a. If Assumption 2.1 holds and the cost function is bounded by a constant M (see Remark 2.3), then

$$O(n, x) = M\alpha^n/(1 - \alpha), \tag{9}$$

$x \in X, n = 1, 2, \cdots$ (observe that in this case $O(n, x), x \in X, n = 1, 2, \cdots$, is independent from $x \in X$).

b. Using (4) and the fact that $c \geq 0$, it is direct to prove that $V_n(\cdot) \leq V_{n+1}(\cdot)$ for all $n = 1, 2, \cdots$. Moreover, from (8) it follows that $V_n(x) \uparrow V^*(x)$ for every $x \in X$.

Assumption 2.6. Suppose that f^* given in (7) is unique.

Remark 2.7. See Cruz-Suárez et al [4] for conditions to ensure the uniqueness of optimal policies of discounted MDPs.

Assumption 2.8. a. For each $x \in X$, $c(x,\cdot)$ is a continuous function on $A(x)$.

b. For each $x \in X$,

$$\int V_n(y)\, Q(dy \mid x,\cdot)\, , n = 1,2,\cdots , \quad \text{and} \tag{10}$$

$$\int V^*(y)\, Q(dy \mid x,\cdot) \tag{11}$$

are finite and continuous functions on $A(x)$.

Lemma 2.9. *Suppose Assumptions 2.1, 2.6, and 2.8 hold. Then* $f_n(x) \to f^*(x)$, *for each* $x \in X$.

Proof. See Theorem 5.6 in Cruz-Suárez and Montes-de-Oca [3]. □

Lemma 2.10. *Let* $\{g_n\}$ *be a sequence of continuous real-valued functions on a metric space* (Y, d_Y). *Let* g *be a continuous real-valued function on* Y. *Then* $g_n(z_n) \to g(z)$, $n \to +\infty$ *for every* $z \in Y$ *and every sequence* $\{z_n\}$ *in* Y *which converges to* z *if and only if* $\{g_n\}$ *converges uniformly on compact sets to* g.

Proof. See the proof of Result 7.5 of Chapter XII in Dugundji [5] (see also the proof of Lemma 2.13 in Cruz-Suárez and Montes-de-Oca [3]). □

3. Main Result

Notation 3.1. a. Let d be the metric on the control space A. For each $\varepsilon > 0$, $x \in X$ and $n = 1,2,\cdots$, $B_\varepsilon(f_n(x))$ is the $\varepsilon-$neighborhood of $f_n(x)$ on $A(x)$, i.e.

$$B_\varepsilon(f_n(x)) = \{a \in A(x) \mid d(a, f_n(x)) < \varepsilon\}. \tag{12}$$

$B_\varepsilon^c(f_n(x))$ is the complement of $B_\varepsilon(f_n(x))$ with respect to $A(x)$, i.e.

$$B_\varepsilon^c(f_n(x)) = \{a \in A(x) \mid d(a, f_n(x)) \geq \varepsilon\}. \tag{13}$$

b. For each $n = 1,2,\cdots$,

$$G_n(x,a) := c(x,a) + \alpha \int V_{n-1}(y)\, Q(dy \mid x,a), \tag{14}$$

$(x,a) \in \mathbb{K}$, where V_{n-1} is the function given in (4).

c. $G(x,a) := c(x,a) + \alpha \int V^*(y)\, Q(dy \mid x,a)\,, (x,a) \in \mathbb{K}$.

d. Denote

$$D_n(x,a) = G_n(x,a) - V_n(x)\,, \tag{15}$$

$(x,a) \in \mathbb{K}, n = 1,2,\cdots$; and

$$D(x,a) = G(x,a) - V^*(x)\,, \tag{16}$$

$(x,a) \in \mathbb{K}$.

Remark 3.2. a. The functions $D_n, n = 1,2,\cdots$, and D are known in the literature of MDPs as the α-value iteration discrepancy functions and the α-discount discrepancy function, respectively (see Hernández-Lerma and Lasserre [9]).

b. Let $x \in X$ and $a \in A(x)$. From (7) and Assumption 2.6, it is evident that $D(x,a) = 0$ if and only if a is optimal.

Lemma 3.3. *Suppose that Assumptions 2.1, 2.2, and 2.8 hold. Then:*

a. for each $x \in X, \{G_n(x,\cdot)\}$ converges uniformly to $G(x,\cdot)$ on every compact subset of $A(x)$; moreover,

b. for each $x \in X, \{D_n(x,\cdot)\}$ converges uniformly to $D(x,\cdot)$ on every compact subset of $A(x)$.

Proof. a. This is a direct consequence of Lemma 2.4 (see Remark 2.5b) and of the Dini's Theorem (Kelley [11] p. 239). (See also the proof of Lemma 3.28 in Cruz-Suárez and Montes-de-Oca [3]).

b. Let x be an arbitrary, fixed state. Let ς be an arbitrary, fixed nonempty compact subset of $A(x)$. Since $V_n(x) \uparrow V^*(x)$ (see Remark 2.5b), then for each $\varepsilon > 0$ there exists a positive integer N_1 such that

$$|V_n(x) - V^*(x)| < \frac{\varepsilon}{2}, \tag{17}$$

for all $n \geq N_1$. From Lemma 3.3a, on the compact set ς, there is a positive integer N_2 for the same $\varepsilon > 0$, such that

$$|G_n(x,a) - G(x,a)| < \frac{\varepsilon}{2}, \tag{18}$$

for all $n \geq N_2$, and all $a \in \varsigma$. Let $\overline{N} = \max\{N_1, N_2\}$. Then, for all $n \geq \overline{N}$ and all $a \in \varsigma$,

$$\begin{aligned} |D_n(x,a) - D(x,a)\,| &= |G_n(x,a) - V_n(x) - G(x,a) + V^*(x)| \\ &\leq |G_n(x,a) - G(x,a)\,| + |V_n(x) - V^*(x)| < \varepsilon, \end{aligned} \tag{19}$$

where (19) is valid due to (17) and (18). Since x and ς are arbitrary, the desired result follows. □

Theorem 3.4. *Suppose Assumptions 2.1, 2.2, 2.6, and 2.8 hold. Then, for each $x \in X$ and $\varepsilon > 0$, there exists a positive integer N such that at least one of the conditions $B_{\varepsilon}^{c}(f_N(x)) = \emptyset$ or*

$$\inf_{a \in B_{\varepsilon}^{c}(f_N(x))} D_N(x,a) > 2O(N,x) \tag{20}$$

holds (observe that N may depend on x and ε).

Proof. The proof is obtained by contradiction. Thus, suppose that there exist $x \in X$ and $\varepsilon > 0$ which satisfy that $B_{\varepsilon}^{c}(f_n(x)) \neq \emptyset$ and

$$\inf_{a \in B_{\varepsilon}^{c}(f_n(x))} D_n(x,a) \leq 2O(n,x), \tag{21}$$

for all $n = 1, 2, \cdots$.

Since for each $n = 1, 2, \cdots$, $D_n(x,\cdot)$ is a continuous function, and $B_{\varepsilon}^{c}(f_n(x))$ is a compact set, then there is $a_n(x) \in B_{\varepsilon}^{c}(f_n(x))$ such that

$$\inf_{a \in B_{\varepsilon}^{c}(f_n(x))} D_n(x,a) = D_n(x, a_n(x)), \tag{22}$$

for all n. Let $y_n = a_n(x) \in A(x)$ for each $n = 1, 2, \cdots$. As $y_n \in B_{\varepsilon}^{c}(f_n(x))$ for all n and this set is compact, then there is a subsequence $\{y_{n_k}\}$ of $\{y_n\}$ and $y \in A(x)$ such that $y_{n_k} \to y, k \to \infty$. As $y_{n_k} \in B_{\varepsilon}^{c}(f_{n_k}(x)) \subset A(x)$, for all k, then

$$d(y_{n_k}, f_{n_k}(x)) \geq \varepsilon, \tag{23}$$

for all k. If $k \to \infty$ in (23), using Lemma 2.9, it follows that

$$d(y, f^*(x)) \geq \varepsilon; \tag{24}$$

then $y \neq f^*(x)$.

Now, Lemma 3.3b implies that $D_n(x,\cdot)$ converges to $D(x,\cdot)$ uniformly on every compact subset of $A(x)$. Consider

$$0 \leq D_{n_k}(x, y_{n_k}) \leq 2O(n_k, x), \tag{25}$$

$k = 1, 2, \cdots$. Then, by Lemma 2.10 and as $O(n_k, x) \to 0$, $k \to \infty$, from (25), it follows that:

$$D(x,y) = 0. \tag{26}$$

So, by uniqueness of f^* it follows that $y = f^*(x)$, which is a contradiction to (24). This is the end of the proof. □

Corollary 3.5. *Suppose that Assumptions 2.1, 2.2, 2.6, and 2.8 hold. Then, for each $x \in X$ and $\varepsilon > 0$, $f_N(x)$, where N is the positive integer whose existence is guaranteed in Theorem 3.4, satisfies the following inequality:*

$$d(f_N(x), f^*(x)) < \varepsilon \tag{27}$$

(in this case $f_N(x)$ is referred to as an $\varepsilon-$approximation to $f^*(x)$).

Proof. Let x be an arbitrary, fixed state. Let ε be an arbitrary, fixed positive number. If $B_\varepsilon^c(f_N(x)) = \emptyset$, then $B_\varepsilon(f_N(x)) = A(x)$. Hence

$$f^*(x) \in B_\varepsilon(f_N(x)),$$

i.e. $d(f_N(x), f^*(x)) < \varepsilon$.

On the other hand, suppose that (20) holds. Let $a \in B_\varepsilon^c(f_N(x))$. Then, using (8), (15), and Assumption 2.2b, it follows that

$$\begin{aligned}
D(x,a) &= c(x,a) + \alpha \int V^*(y)\, Q(dy \mid x,a) - V^*(x) \\
&= c(x,a) + \alpha \int V^*(y)\, Q(dy \mid x,a) - \alpha \int V_{N-1}(y)\, Q(dy \mid x,a) \\
&\quad + \alpha \int V_{N-1}(y)\, Q(dy \mid x,a) - V_N(x) + (V_N(x) - V^*(x)) \\
&= D_N(x,a) + \alpha \int [V^*(y) - V_{N-1}(y)]\, Q(dy \mid x,a) \\
&\quad + (V_N(x) - V^*(x)) \\
&\geq D_N(x,a) - \alpha \frac{\overline{c}(\alpha\beta)^{N-1}}{1-\alpha\beta} \int w(y) Q(dy \mid x,a) - O(N,x) \\
&\geq \inf_{a \in B_\varepsilon^c(f_N(x))} D_N(x,a) - \frac{\overline{c}(\alpha\beta)^N}{1-\alpha\beta} w(x) - O(N,x) \\
&= \inf_{a \in B_\varepsilon^c(f_N(x))} D_N(x,a) - 2O(N,x) \\
&> 0.
\end{aligned}$$

Then a is non-optimal in state x, (see Remark 3.2b). This implies that $f^*(x) \in B_\varepsilon(f_N(x))$, i.e. $d(f_N(x), f^*(x)) < \varepsilon$. Since x and ε are arbitrary, Corollary 3.5 follows. □

Remark 3.6. For each $x \in X$ and $\varepsilon > 0$, consider the set

$$Z(x,\varepsilon) = \{n \in \mathbb{N} \mid B_\varepsilon^c(f_n(x)) = \emptyset \text{ or } \inf_{a \in B_\varepsilon^c(f_n(x))} D_n(x,a) > 2O(n,x)\} \tag{28}$$

($\mathbb{N}$ denotes the set of all positive integers). Notice that Theorem 3.4 implies that $Z(x,\varepsilon) \neq \emptyset$ for each $x \in X$ and $\varepsilon > 0$. Besides, similar to the proof of Corollary 3.5, it is possible to prove that if $m \in Z(x,\varepsilon)$, then $d(f_m(x), f^*(x)) < \varepsilon$, i.e. $f_m(x)$ is an ε-approximation to $f^*(x)$. Hence, to find an ε-approximation to $f^*(x)$, the minimum of $Z(x,\varepsilon)$ will be sought, by means of the following algorithm.

Algorithm 3.7. Suppose that Assumptions 2.1, 2.2, 2.6, and 2.8 hold. For the steps below let x be a fixed state, and let $\varepsilon > 0$.

1. Let $n = 1$.
2. If $B_\varepsilon^c(f_n(x)) = \emptyset$, then stop and $f_n(x)$ is an ε-approximation to $f^*(x)$. Otherwise, go to the following step.
3. Compute
$$\inf_{a \in B_\varepsilon^c(f_n(x))} D_n(x,a). \tag{29}$$
4. If $\inf_{a \in B_\varepsilon^c(f_n(x))} D_n(x,a) > 2O(n,x)$, then stop and the corresponding minimizer $f_n(x)$ is an ε-approximation to $f^*(x)$. Otherwise, go to the following step.
5. Increase n in one unit, and then go to Step 2.

4. Examples

To illustrate Algorithm 3.7 two examples will be given. One of them is an inventory/production model (see Example 4.2 in Cruz-Suárez and Montes-de-Oca [3], and Example 8.2 in Yushkevich [12]). The other example is the linear-quadratic model (see Bertsekas [1]).

Example 4.1. (see Cruz-Suárez and Montes-de-Oca [3], Yushkevich [12]) Let $X = A = [0,1]$, $A(x) = [0, 1-x]$, $x \in X$, and consider

$$x_{t+1} = [x_t + a_t - \xi_t]^+, \tag{30}$$

$t = 0, 1, \cdots$, where $z^+ := \max\{0, z\}$, $z \in \mathbb{R}$. Here $\xi_0, \xi_1, \cdots$, are i.i.d. random variables taking values in $S = [0, \infty)$, and with common density Δ. It will be assumed that Δ is a continuous bounded function (notice that the

distribution function $\widehat{G}$ of ξ is a continuous function, where ξ is a generic element of the sequence $\{\xi_t\}$). The cost function is given by $c(x,a) = x^2 + x + \left[a - \frac{1}{2}(1-x)\right]^2, (x,a) \in \mathbb{K}$.

The discount factor is $\alpha = 1/75$.

Lemma 4.2. *Example 4.1 satisfies Assumption 2.1.*

Proof. Clearly Example 4.1 satisfies Assumption 2.1a and Assumption 2.1b. In reference to Assumption 2.1c, notice that if $\mu : X \to \mathbb{R}$ is a measurable and bounded function, then a simple computation permits to show that

$$\begin{aligned}\int \mu(y)\, Q(dy \mid x,a) &= \mu(0)\left[1 - \widehat{G}(x+a)\right] \\ &\quad + \int I_{[0,x+a]}(u)\, \mu(u)\, \Delta(x+a-u)\, du, \qquad (31)\end{aligned}$$

$(x,a) \in \mathbb{K}$,where $I_{[\cdot]}$ denotes the indicator function of the subset $[\cdot]$.

Since $\widehat{G}$ is a continuous function, it follows that $\mu(0)\left[1 - \widehat{G}(x+a)\right]$ is a continuous function on $\mathbb{K}$. Now, as μ is a bounded function and Δ is a continuous bounded function, it follows directly, using the Dominated Convergence Theorem, that

$$\int I_{[0,x+a]}(u)\, \mu(u)\, \Delta(x+a-u)\, du, \qquad (32)$$

is a continuous function on (x,a), for each $(x,a) \in \mathbb{K}$.

Therefore,

$$\int \mu(y)\, Q(dy \mid \cdot,\cdot) \qquad (33)$$

is a continuous function on $\mathbb{K}$. Hence, Example 4.1 satisfies Assumption 2.1. □

Remark 4.3. Using that $0 \le x \le 1$ and $0 \le a \le 1 - x$, it is easy to verify that

$$|c(x,a)| \le \frac{9}{4}, \qquad (34)$$

$(x,a) \in \mathbb{K}$. Hence, from Remark 2.3, Example 4.1 satisfies Assumption 2.2.

Lemma 4.4. *Example 4.1 satisfies Assumption 2.6.*

Proof. It will be shown that Example 4.1 satisfies Condition C1 of Cruz-Suárez et al [4] which implies Assumption 2.6.

Condition C1 of Cruz-Suárez et al [4] requires the five parts C1a-C1e.

C1a and C1b require X and A to be convex. Obviously, Example 4.1 satisfies these conditions.

An elementary computation permits to get that

$$(1-\lambda)\,a+\lambda a'\in A\left((1-\lambda)\,x+\lambda x'\right)=\left[0,1-\left((1-\lambda)\,x+\lambda x'\right)\right]$$

if $x,x'\in X$, $a\in A\left(x\right)=\left[0,1-x\right],a'\in A\left(x'\right)=\left[0,1-x'\right],\lambda\in\left[0,1\right]$. Furthermore, if $x,y\in X,x<y$, then $A\left(y\right)\subset A\left(x\right)$. These properties define Condition C1(c). Then Example 4.1 satisfies Condition C1c of Cruz-Suárez et al [4].

Condition C1d means that Q is induced by a difference equation $x_{t+1}=F(x_t,a_t,\xi_t),t=0,1,\cdots$, where $F:X\times A\times S\to X$ is a measurable function, and $\{\xi_t\}$ is a sequence of i.i.d. random elements with values in $S\subset\mathbb{R}^l$ for some positive integer l, and with common density Δ. In addition, Condition C1d of Cruz-Suárez et al [4] presents $F\left(\cdot,\cdot,s\right)$ as a convex function on $\mathbb{K}$, for each $s\in S$, and if $x,y\in X,x<y$, then $F\left(x,a,s\right)\leq F\left(y,a,s\right)$ for each $a\in A\left(y\right)$ and $s\in S$. Then, using (30) and the fact that $g(z):=z^+,z\in\mathbb{R}$ is a convex and nondecreasing function, it is trivial to see that Example 4.1 satisfies Condition C1d.

Finally, as $c\left(x,a\right)\leq c\left(y,a\right)$, for $x,y\in X,x<y$, and $a\in A\left(y\right)$ (observe that $\frac{\partial c}{\partial x}=\frac{5}{2}x+a+\frac{1}{2}\geq\frac{1}{2}>0$, for all $(x,a)\in\mathbb{K}$), and c is strictly convex on $\mathbb{K}$ (one way to prove the strict convexity of c on $\mathbb{K}$ is to verify the assumption in Theorem 3.6(a') in Fleming [7], p. 114), consequently, C1e holds. This completes the proof of Lemma 4.4. □

Remark 4.5. For Example 4.1, Assumption 2.8 has been proved in Lemma 4.5b in Cruz-Suárez and Montes-de-Oca [3].

Lemma 4.6. *Let $\varepsilon=0.25$ and $x=0$, then $f_1(0)$ is a 0.25-approximation to $f^*(0)$; i.e.*

$$\left|f_1\left(0\right)-f^*\left(0\right)\right|<\frac{1}{4}. \tag{35}$$

Proof. By means of straightforward computations,

$$V_1\left(0\right)=0, \tag{36}$$

with

$$f_1\left(0\right)=\frac{1}{2}, \tag{37}$$

then

$$B_\varepsilon^c(f_1(0))=[0,\frac{1}{4}]\cup[\frac{3}{4},1]. \tag{38}$$

Hence, it is obtained that

$$\min_{a\in B_\varepsilon^c(f_1(0))}D_1\left(0,a\right)=\frac{1}{16}>2O(1,0)=\frac{9}{148}.$$

Thus, $f_1(0) = \frac{1}{2}$ is a 0.25−approximation to $f^*(0)$. □

Example 4.7. (see Bertsekas [1]) Consider a simple linear system with

$$x_{t+1} = x_t + a_t + \xi_t, \quad t = 0, 1, \cdots, \tag{39}$$

with quadratic cost

$$c(x, a) = x^2 + a^2, \tag{40}$$

$x, a \in \mathbb{R}$. Here $X = A = A(x) = \mathbb{R}, x \in X$.

In addition, take $\alpha = 1/150$.

Assumption 4.8. The disturbances ξ_t, $t = 0, 1, \cdots$ are i.i.d. random variables with values in $S = \mathbb{R}$. Moreover, suppose that ξ_0 has a continuous density Δ, zero mean value and a finite variance $\sigma^2 = \frac{3}{2}$.

Remark 4.9. Example 4.7 has been studied in Cruz-Suárez and Montes-de-Oca [3] and Cruz-Suárez et al [4] under Assumption 4.8. In these references it has been proved (see Example 4.8 in Cruz-Suárez et al [4] and Example 4.9 in Cruz-Suárez and Montes-de-Oca [3]) that Example 4.7 satisfies Assumptions 2.1a, 2.1c, 2.6, 2.8, and instead of Assumption 2.1b, the inf-compactness of the cost function c on $\mathbb{K}$ is verified (c is inf-compact on $\mathbb{K}$ if the set $\{a \in A(x) | c(x, a) \leq \overline{s}\}$ is compact for every $x \in X$ and $\overline{s} \in \mathbb{R}$). Hence for this example there is a unique stationary optimal policy f^*, and the sequence of minimizers $\{f_n\}$ converges pointwise to f^*.

On the other hand, using results given in Hernández-Lerma and Lasserre [8] pp. 70-72, it is easy to verify that $f_1(x) = 0$, $f_n(x) = -\lambda_n x$, $n = 2, 3, \cdots$, $x \in X$, where $\lambda_n = \frac{1}{150} T_{n-1} / \left(1 + \frac{1}{150} T_{n-1}\right)$, and

$$T_n = 1 + \frac{\frac{1}{150} T_{n-1}}{1 + \frac{1}{150} T_{n-1}}, \tag{41}$$

with $T_0 = 0$, $T_1 = 1$, $n = 2, 3, \cdots$. Moreover, $V_1(x) = x^2$, $x \in X$ and

$$V_n(x) = T_n x^2 + \frac{1}{100} \sum_{i=1}^{n-1} \left(\frac{1}{150}\right)^{n-i-1} T_i, \tag{42}$$

$n = 2, 3, \cdots$, $x \in X$.

Besides, observe that it is easily seen that the policies f_n and f^* satisfy the constraint $f_n(x)$, $f^*(x) \in [-|x|, |x|]$ for all $x \in X$.

Therefore, in order to satisfy Assumption 2.1b and Assumption 2.2, Example 4.10 will be analyzed. Notice that in this equivalent example and in Example 4.7 the optimal value function V^*, the value iteration functions $V_n, n = 1, 2, \cdots$, the minimizers $f_n, n = 1, 2, \cdots$, and the optimal policy f^* coincide.

Example 4.10. Consider Example 4.7 with the restriction that $A(x) = [-|x|, |x|], x \in \mathbb{R}$ (suppose that Assumption 4.8 holds).

Lemma 4.11. *For Example 4.10, let* $w(x) = 2x^2 + 1$, $x \in X$, $\overline{c} = 1$, *and* $\beta = 4$. *Then, for each* $x \in X$*:*

a. $\sup_{a \in A(x)} c(x,a) \leq w(x)$, *and*

b. $\sup_{a \in A(x)} \int w(y) Q(dy \mid x,a) \leq 4w(x)$.

c. *For each* $x \in X$, $\int w(y) Q(dy \mid x,a)$ *is continuous in* $a \in A(x)$.

(Hence, since for each $x \in X$, $A(x)$ *is a compact set, it follows that Assumption 2.2 holds for Example 4.10.)*

Proof. a. For each $x \in X$, note that

$$\sup_{a \in [-|x|,|x|]} c(x,a) = 2x^2 \leq 2x^2 + 1 = w(x), \tag{43}$$

b. Then, for $x \in X$, consider

$$\sup_{a \in [-|x|,|x|]} \int w(y) Q(dy \mid x,a) = \sup_{a \in [-|x|,|x|]} \int \left(2y^2 + 1\right) Q(dy \mid x,a). \tag{44}$$

Therefore,

$$\begin{aligned} &\sup_{a \in [-|x|,|x|]} \int \left(2y^2 + 1\right) Q(dy \mid x,a) \\ &\qquad = 2 \sup_{a \in [-|x|,|x|]} \int (x + a + s)^2 \Delta(s)\, ds + 1 \\ &\qquad = 2 \sup_{a \in [-|x|,|x|]} \left[(x+a)^2\right] + 4 \leq 8x^2 + 4 = 4w(x). \end{aligned} \tag{45}$$

c. The proof follows directly from the fact that for each $x \in X$,

$$\int w(y) Q(dy \mid x,a) = 2(x+a)^2 + 4, \tag{46}$$

$a \in A(x)$. □

Lemma 4.12. *Let* $\varepsilon = 0.1$ *and* $x = 1$. *Then for Example 4.10, it is obtained that* $f_2(1)$ *is a* 0.1*-approximation to* $f^*(1)$.

Proof. Observe that $B_\varepsilon^c(f_i(1)) \neq \emptyset$, $i = 1, 2$, and it is straightforward to verify that $2O(1,1) > \inf_{a \in B_\varepsilon^c(f_1(1))} D_1(1,a)$.

For $n = 2$, using (41) and (42), it results that

$$\begin{aligned}
&\inf_{a \in B_\varepsilon^c(f_2(1))} D_2(1,a) \\
&\quad = \inf_{a \in B_\varepsilon^c(f_2(1))} \left[c(1,a) + \frac{1}{150} \int V_1(y)\, Q(dy \mid 1,a) - V_2(1) \right] \\
&\quad = \inf_{a \in B_\varepsilon^c(f_2(1))} \left[1 + a^2 + \frac{1}{150} \int y^2 Q(dy \mid 1,a) \right. \\
&\quad \left. - \left(\frac{152}{151}(1)^2 + \frac{1}{100} \right) \right] = \inf_{a \in B_\varepsilon^c(f_2(1))} \left[\frac{1}{22650} + \frac{1}{75}a + \frac{151}{150}a^2 \right], \quad (47)
\end{aligned}$$

where

$$\begin{aligned}
B_\varepsilon^c(f_2(1)) &= \{a \in [-1,1] \mid a \notin (f_2(1) - 0.1, f_2(1) + 0.1)\} \\
&= \left\{ a \in [-1,1] \mid a \notin \left(-\frac{1}{151} - 0.1, -\frac{1}{151} + 0.1 \right) \right\}.
\end{aligned}$$

To determine the infimum in (47), an elementary analysis of the behavior of $D_2(1,a)$ on $B_{0.1}^c(f_2(1))$ shows that

$$\begin{aligned}
\inf_{a \in B_\varepsilon^c(f_2(1))} D_2(1,a) &= D_2\left(1, -\frac{161}{1510}\right) \\
&= 0.0100667 > 0.00438356 = \frac{450}{73}\left(\frac{2}{75}\right)^2 = 2O(2,1). \quad (48)
\end{aligned}$$

Hence $f_2(1) = -(1/151)$ is a 0.1-approximation to $f^*(1)$. □

References

[1] D.P. Bertsekas, *Dynamic Programming: Deterministic and Stochastic Models*, Prentice-Hall, NJ (1987).

[2] D.P. Bertsekas, S.E. Shreve, *Stochastic Optimal Control: The Discrete Time Case*, Academic Press, New York (1978).

[3] D. Cruz-Suárez, R. Montes-de-Oca, Uniform convergence of the value iteration policies for discounted Markov decision processes, *Boletín de la Sociedad Matemática Mexicana*, To Appear.

[4] D. Cruz-Suárez, R. Montes-de-Oca, F. Salem-Silva, Conditions for the uniqueness of optimal policies of discounted Markov decision processes, *Mathematical Methods of Operations Research*, **60**, No. 3 (2004), 415-436.

[5] J. Dugundji, *Topology*, Allyn and Bacon Inc., USA (1966).

[6] E.B. Dynkin, A.A. Yushkevich, *Controlled Markov Processes*, Springer-Verlag, USA (1979).

[7] W. Fleming, *Functions of Several Variables*, Second Edition, Springer-Verlag, Berlin (1977).

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

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

[10] K. Hinderer, Foundations of non-stationary dynamic programming with discrete-time parameter, *Lecture Notes in Operations Research and Mathematical Systems*, Springer-Verlag, Berlin, **33** (1970),

[11] J.L. Kelley, *General Topology*, Springer-Verlag (1975).

[12] A.A. Yushkevich, Blackwell optimality in Borelian continuous-in-action Markov decision processes, *SIAM Journal on Control and Optimization*, **35**, No. 6 (1997), 2157-2182.