International Journal of Pure and Applied Mathematics
Volume 40 No. 3 2007, 411-426

SUBOPTIMALITY OF THE VALUE ITERATION POLICIES IN DISCOUNTED LINEAR-QUADRATIC MODELS

Raúl Montes-de-Oca[1 §], Daniel Cruz-Suárez[2]

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

[2]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

Abstract: This paper deals with one-dimensional linear-quadratic (LQ) models which have been established as discounted Markov decision processes. These models have general and known coefficients. For each LQ-model taken into account, the existence of an optimal policy f^* is assumed. Conditions which permit to obtain, for each compact set of states ς, and for each $n = 1, 2, \cdots$, the suboptimality on ς of the value iteration policy f_n are given. This suboptimality consists in obtaining the range of the values of ε, such that f_n is a uniform on ς ε-approximation to f^*.

AMS Subject Classification: 90C40, 93E20
Key Words: discounted linear-quadratic model, optimal policy, value iteration policy, suboptimality of the value iteration policy

Received: August 31, 2007

[§]Correspondence author

1. Introduction

This paper deals with one-dimensional linear-quadratic (LQ) models established as discrete-time and infinite horizon discounted Markov decision processes (MDPs) (see [6]).

The classical one-dimensional LQ-model (see [1]) has the state space, the control space, and the restriction sets, all defined as the set of real numbers $\mathbb{R}$. This model has a linear transition probability law and a quadratic cost function.

For such a model it is well-known that, under appropriate assumptions (see [1]), the optimal policy f^* exists and is a linear function, that is, $f^*(x) = \theta x, x \in \mathbb{R}$, where θ is a real constant related to the Ricatti's equation (see [1]).

Even though the functional form of f^* is known, it is possible for θ to be an irrational number and it is necessary to provide approximations to it (see Remark 3.6 and Example 5.1 below).

In this paper, in order to approximate f^*, there is considered an LQ-model with convenient compact restriction sets which is equivalent to the classical model in the sense that for both models the optimal value function V^*, the value iteration functions V_n, the minimizers f_n of the value iteration equations, $n = 1, 2, \cdots$, and the optimal policy f^* coincide.

Moreover, the value iteration policies f_n (of the LQ-models with compact restriction sets mentioned above) will be used to approximate to f^*, and the main goal of the paper is to study the suboptimality of f_n using the new theory developed in [5].

Concretely, the suboptimality of f_n is understood as the specification, for each compact set of states ς , of the range of the values of ε such that f_n is a uniform on ς ε-approximation to f^*.

In the article, conditions to ensure the suboptimality of f_n are provided (see Theorem 4.4 below).

The paper is organized as follows. In Section 2, the basic theory of MDPs and some assumptions are provided. In Section 3, the LQ-models are supplied. In Section 4, the main result over suboptimality of the value iteration policies is established. Finally, a specific example is presented in Section 5.

2. Discounted Markov Decision Processes

In this section there are provided some results on the theory of Markov decision processes (see [6] and [7]) that will be used as the general context in which the linear-quadratic model will be established.

Let $(X,\, A,\, \{A(x) : x \in X\}, Q, c)$ be a discrete-time Markov control model, which consists of the state space X, the control (or action) set A, the transition law Q, and the cost-per-stage c. The sets X and A are assumed to be Borel spaces with Borel sigma-algebras $\mathbb{B}(X)$ and $\mathbb{B}(A)$, respectively. Moreover, for every $x \in X$ there is a nonempty set $A(x) \in \mathbb{B}(A)$ whose elements are the feasible actions when the state of the system is x. Define $\mathbb{K} := \{(y, b) \mid y \in X, b \in A(y)\}$. The transition law $Q(B \mid x, a), B \in \mathbb{B}(X), (x, a) \in \mathbb{K}$ is a stochastic kernel on X given $\mathbb{K}$. Finally, the cost-per-stage c is a nonnegative measurable function on $\mathbb{K}$.

Let Π be the set of all (possibly randomized, history-dependent) admissible policies. By standard convention, a *stationary* policy is taken as a measurable function $f : X \to A$ such that $f(x) \in A(x)$ for all $x \in X$. The set of the stationary policies is denoted by $\mathbb{F}$.

For every $\pi \in \Pi$ and initial state $x \in X$, let

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

be the *expected total discounted cost*. The number $\alpha \in (0, 1)$ is called the *discount factor*. Here, $\{x_t\}$ and $\{a_t\}$ denote the state and the control sequences, respectively, and E_x^{π} is the expectation operator with respect to the probability measure P_x^{π} defined on the space $\Omega := (X \times A)^{\infty}$ in a canonical way (see [6]). The *optimal control problem* is to find a policy π^* such that $V(\pi^*, x) = \inf_{\pi\in\Pi} V(\pi, x)$, for all $x \in X$, in which case π^* is said to be *optimal*.

The *optimal value function* V^* is defined for $x \in X$ as

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

Assumption 2.1. a. The one-stage cost $c(\cdot, \cdot)$ 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.

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

a. $\sup_{a\in A(x)} |c(x, a)| = \sup_{a\in A(x)} c(x, a) \leq \overline{c} w(x)$.

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

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

The *value iteration functions* are defined as

$$V_n(x) = \min_{a \in A(x)} \left[c(x,a) + \alpha \int V_{n-1}(y)\, Q(dy \mid x,a) \right], \tag{3}$$

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

Remark 2.3. Using Assumption 2.1, it is possible to demonstrate (see [7]) that, for each $n = 1, 2, \cdots$, there exists a stationary policy $f_n \in \mathbb{F}$ such that the minimum in (3) is attained, i.e.

$$V_n(x) = c(x, f_n(x)) + \alpha \int V_{n-1}(y)\, Q(dy \mid x, f_n(x)), \tag{4}$$

$x \in X$.

Lemma 2.4. (see [7], Theorem 8.3.6) *If Assumptions 2.1 and 2.2 hold, then the optimal value function V^* defined in (2) satisfies the optimality equation, i.e. for all $x \in X$:*

$$V^*(x) = \min_{a \in A(x)} \left[c(x,a) + \alpha \int V^*(y)\, Q(dy \mid x,a) \right]. \tag{5}$$

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

$$V^*(x) = c(x, f^*(x)) + \alpha \int V^*(y)\, Q(dy \mid x, f^*(x)), \tag{6}$$

$x \in X$, and f^* is optimal; conversely, if $g \in \mathbb{F}$ is optimal, then it satisfies (6).

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

$$|V_n(x) - V^*(x)| \leq O(n,x) := \overline{c}\,(\alpha\eta)^n\, w(x) \,/\, (1 - \alpha\eta). \tag{7}$$

Besides, for MDPs which satisfy Assumptions 2.1 and 2.2 the following two assumptions will be needed (see [2], [3], and [5]).

Assumption 2.5. Suppose that f^* given in (6) is unique. In addition, assume that for each $n = 1, 2, \cdots, f_n$ given in (4) is unique.

Assumption 2.6. a. The multifunction $x \longmapsto A(x)$ is upper semicontinuous.

b. $c(\cdot,\cdot)$ is a continuous function on $\mathbb{K}$.

c. $V_n(\cdot)$, $n = 1, 2, \cdots$, and $V^*(\cdot)$ are continuous functions on X.

d. The integrals

$$\int V_n(y)\, Q(dy \mid \cdot,\cdot) \tag{8}$$

and

$$\int V^{*}(y)\, Q(dy \mid \cdot, \cdot) \tag{9}$$

are finite and continuous functions on $\mathbb{K}$.

3. The Linear-Quadratic Model

3.1. One-Dimensional LQ-Model

Example 3.1 below provides the classical LQ-model (see [1], [6]).

Example 3.1. Let us consider the one-dimensional LQ-model, for which $X = A = \mathbb{R}$, and for $x \in X$

$$A(x) := \mathbb{R}. \tag{10}$$

The transition probability law is given by the linear system

$$x_{t+1} = \gamma x_t + \beta a_t + \xi_t, \quad t = 0, 1, \cdots, \tag{11}$$

with a quadratic one-stage cost

$$c(x, a) = qx^2 + ra^2, \tag{12}$$

$x \in \mathbb{R}$, $a \in A(x)$. Here β, γ, q, and r are constants that satisfy Assumption 3.2 below (see [1]).

Assumption 3.2. The disturbances ξ_t, $t = 0, 1, \cdots$, are i.i.d. random variables with values in $\widehat{S} = \mathbb{R}$. Suppose that ξ_0 has a continuous density Δ, zero mean value, and a finite variance σ^2. Moreover, $q \geq 0, r > 0$, and $\gamma\beta \neq 0$.

Now in Example 3.3 below, a version of the LQ-model with compact restriction sets $A(x), x \in X$ will be presented. In fact, it will be proved that Example 3.1 and Example 3.3 are equivalent, in the sense that for both examples 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 (see Lemma 3.5 and its proof below).

Example 3.3. Let $X = A = \mathbb{R}$, and for $x \in X$,

$$A(x) := \left[-\left|\gamma x/\beta\right|, \left|\gamma x/\beta\right|\right]. \tag{13}$$

The transition probability law is given by the linear system

$$x_{t+1} = \gamma x_t + \beta a_t + \xi_t, \quad t = 0, 1, \cdots, \tag{14}$$

with a quadratic one-stage cost

$$c(x,a) = qx^2 + ra^2, \tag{15}$$

$x \in \mathbb{R}$, $a \in A(x)$. Here β, γ, q, and r are constants that satisfy Assumption 3.4 below.

Assumption 3.4. Consider Assumption 3.2, but changing $\gamma\beta \neq 0$ by $\beta \neq 0$ and $\frac{1}{2} <\mid \gamma \mid < \frac{1}{2\sqrt{\alpha}}$.

3.2. Results of the LQ-model

Lemma 3.5. *Under Assumption 3.4, Example 3.3 satisfies:*
a. Assumptions 2.1, and 2.2.
b. For all $n \geq 1$ and $x \in \mathbb{R}$,

$$V_n(x) = U_n x^2 + H_n, \tag{16}$$

and

$$f_n(x) = -\frac{\alpha\beta\gamma U_{n-1} x}{R + SU_{n-1}}, \tag{17}$$

with $f_0 \equiv 0, V_o \equiv 0$,

$$U_n = \frac{P + \overline{Q} U_{n-1}}{R + SU_{n-1}}, \tag{18}$$

$H_n = \alpha\sigma^2 U_{n-1} + \alpha H_{n-1}$, *for all $n \geq 1$, where $P = qr, \overline{Q} = (r\gamma^2 + q\beta^2)\alpha, R = r, S = \alpha\beta^2$, $U_0 = 0$, and $H_0 = 0$.*

There exists U such that $U_n \to U$, when $n \to \infty$ and $H_n \to H$ when $n \to \infty$, where $H = (\alpha\sigma^2 U)/(1-\alpha)$.

Moreover, for each $x \in \mathbb{R}$

$$V^*(x) = Ux^2 + H, \tag{19}$$

and

$$f^*(x) = -\frac{\alpha\beta\gamma U}{R + SU} x. \tag{20}$$

Proof. a. Notice that Assumptions 2.1a and 2.1b trivially hold. Now, for Assumption 2.1c, let $B \in \mathbb{B}(\mathbb{R})$ and $(x,a) \in \mathbb{K}$. Then

$$Q(B \mid x,a) = \int_B \Delta(l - (\gamma x + \beta a)) dl, \tag{21}$$

where the last equality is due to the change of Variable Theorem. From Assumption 3.4 it follows that, for each $s \in \mathbb{R}$, $\Delta(s-(\gamma\cdot+\beta\ \cdot))$ is continuous. Therefore, Lemma 2.3 in [3] implies that Q is strongly continuous.

For each $x \in \mathbb{R}$, let

$$w(x) = \left(q + r\frac{\gamma^2}{\beta^2}\right)x^2 + \Theta, \tag{22}$$

where

$$\Theta = \left(q + r\frac{\gamma^2}{\beta^2}\right)\sigma^2/(4\gamma^2 - 1), \tag{23}$$

and take $\eta = 4\gamma^2$ and $\overline{c} = 1$ (note that Assumption 3.4 implies that $1 < \eta < \frac{1}{\alpha}$). Then, straightforward computations permit to obtain that

$$\sup_{a\in[-|\gamma x/\beta|,|\gamma x/\beta|]} \mid c(x,a) \mid = \sup_{a\in A(x)} (qx^2 + ra^2) \leq \overline{c}w(x), \tag{24}$$

and

$$\sup_{a\in[-|\gamma x/\beta|,|\gamma x/\beta|]} \int w(y)Q(dy \mid x,a) \leq \eta w(x). \tag{25}$$

Since

$$\begin{aligned} w'(x,a) &= \int w(y)Q(dy \mid x,a) \\ &= \left(q + r\frac{\gamma^2}{\beta^2}\right)(\gamma x + \beta a)^2 + \sigma^2 + \Theta, \end{aligned} \tag{26}$$

it follows that $w'(x,\cdot)$ is continuous. Then, Assumption 2.2 holds.

b. Proceeding by induction, it is easy to verify that

$$V_1(x) = U_1x^2 + H_1, \tag{27}$$

and $f_1(x) = 0$, $x \in \mathbb{R}$.

Now, suppose that (16) holds for V_{n-1}. Therefore, from (3) it is obtained that for all $x \in \mathbb{R}$,

$$\begin{aligned} V_n(x) &= \min_{a\in[-|\gamma x/\beta|,|\gamma x/\beta|]} \left[c(x,a) + \alpha\int V_{n-1}(y)\,Q\,(dy \mid x,a)\right] \\ &= \min_{a\in[-|\gamma x/\beta|,|\gamma x/\beta|]} \left[qx^2 + ra^2 + \alpha\int (U_{n-1}y^2 + H_{n-1})Q(dy \mid x,a)\right] \end{aligned}$$

$$
\begin{aligned}
&= \min_{a\in[-|\gamma x/\beta|,|\gamma x/\beta|]}[(q+\alpha\gamma^2U_{n-1})x^2+(r+\alpha\beta^2U_{n-1})a^2 \\
&\quad +2\alpha\beta\gamma U_{n-1}ax+\alpha\sigma^2U_{n-1}+\alpha H_{n-1}] \\
&= \min_{a\in[-|\gamma x/\beta|,|\gamma x/\beta|]}\left[(r+\alpha\beta^2U_{n-1})\left(a+\frac{\alpha\beta\gamma U_{n-1}}{r+\alpha\beta^2U_{n-1}}x\right)^2\right] \\
&\quad +\left(q+\alpha\gamma^2U_{n-1}-\frac{\alpha^2\beta^2\gamma^2U_{n-1}^2}{r+\alpha\beta^2U_{n-1}}\right)x^2+H_n \\
&= U_nx^2+H_n+\min_{a\in[-|\gamma x/\beta|,|\gamma x/\beta|]}\left(a+\frac{\alpha\beta\gamma U_{n-1}}{R+SU_{n-1}}x\right)^2 \qquad (28)\\
&= U_nx^2+H_n, \qquad (29)
\end{aligned}
$$

and from (28)

$$f_n(x)=-\frac{\alpha\beta\gamma U_{n-1}}{R+SU_{n-1}}x.$$

Also, it is direct to verify that for each $x\in\mathbb{R}$,

$$-\frac{\alpha\beta\gamma U_{n-1}}{R+SU_{n-1}}x\in\left[-\left|\gamma x/\beta\right|,\left|\gamma x/\beta\right|\right],$$

and obviously

$$\min_{a\in[-|\gamma x/\beta|,|\gamma x/\beta|]}\left(a+\frac{\alpha\beta\gamma U_{n-1}}{R+SU_{n-1}}x\right)^2=0. \qquad (30)$$

Note that V_n and f_n given by (16) and (17) are equal to the corresponding value iteration functions and the minimizers for the value iteration equation (3) for Example 3.1, respectively (see (4.7.2) and (4.7.3) in [6]). Consequently, from the theory developed in Section 4.7 in [6], there exists U such that $U_n\to U$, when $n\to\infty$ and $H_n\to H$ when $n\to\infty$, where $H=(\alpha\sigma^2U)/(1-\alpha)$. This yields that for each $x\in\mathbb{R}$,

$$V^*(x)=\lim_{n\to\infty}V_n(x)=\lim_{n\to\infty}\left(U_nx^2+H_n\right)=Ux^2+H. \qquad (31)$$

Finally, in Example 4.9 in [2] it has been proved that $\{f_n\}$ converges uniformly on compact sets to f^*. Hence, in particular, for each $x\in\mathbb{R}$,

$$f^*(x)=\lim_{n\to\infty}f_n(x)=\lim_{n\to\infty}-\frac{\alpha\beta\gamma U_{n-1}x}{R+SU_{n-1}}=-\frac{\alpha\beta\gamma Ux}{R+SU}. \qquad (32)$$

Observe that for each $x\in\mathbb{R}$,

$$-\frac{\alpha\beta\gamma U}{R+SU}x\in\left[-\left|\gamma x/\beta\right|,\left|\gamma x/\beta\right|\right], \qquad (33)$$

and that it is direct to prove that V^* and f^* given in (31) and (32), respectively, satisfy (6) for Example 3.3. Therefore, f^* is optimal for that example. □

Remark 3.6. Notice that f^* given by (20) (which is the optimal policy for both Examples 3.1 and 3.3) is a linear function with main coefficient

$$-\frac{\alpha\beta\gamma U}{R+SU}. \tag{34}$$

In (34) α, β, γ, R, and S are defined directly by the model, but U is the positive solution of the Ricatti's equation (see [1] or [6]), i.e.

$$U = \frac{P+\overline{Q}U}{R+SU} \tag{35}$$

(in fact, (35) is a consequence of (18) letting $n \to \infty$; observe that P and $\overline{Q}$ are also defined directly by the model).

For this article, the important case is when (34) is an irrational number (see, for instance, Example (5.1) below).

Lemma 3.7. *Under Assumption 3.4, Example 3.3 satisfies Assumptions 2.5 and 2.6.*

Proof. For each $n = 1, 2, \cdots$, the uniqueness of f_n follows directly from (28), and using a similar argument it is possible to prove that f^* is unique. Then, Example 3.3 satisfies Assumption 2.5.

Obviously Assumptions 2.6b and 2.6c hold for Example 3.3 (see (15), (16), and (19)).

Now, from (14), (16), and (19) it results that for each $(x, a) \in \mathbb{K}$ and $n = 1, 2, \cdots$

$$\int V_n(y)\, Q(dy \mid x, a) = U_n(\gamma x + \beta a)^2 + \sigma^2 U_n + H_n \tag{36}$$

and

$$\int V^*(y)\, Q(dy \mid x, a) = U(\gamma x + \beta a)^2 + \sigma^2 U + H. \tag{37}$$

Hence, Assumption 2.6d clearly holds.

Finally, let $\{x_n\}$ be a sequence in $\mathbb{R}$, such that $x_n \to x$, if $n \to \infty$, where $x \in \mathbb{R}$, and let

$$y_n \in \left[-\left|\gamma x_n/\beta\right|, \left|\gamma x_n/\beta\right|\right] \quad = \quad \left[-\left|\gamma/\beta\right|\left|x_n\right|, \left|\gamma/\beta\right|\left|x_n\right|\right]$$

$$\subset \quad \left[-\left|\gamma/\beta\right| M, \left|\gamma/\beta\right| M\right],$$

$n = 1, 2, \cdots$, where $M > 0$ is a bound for $\{x_n\}$. Then, evidently, there exists a subsequence $\{y_{n_k}\}$ of $\{y_n\}$ and y, such that

$$y_{n_k}, y \in \left[-\left|\gamma/\beta\right| M, \left|\gamma/\beta\right| M\right], \tag{38}$$

for all $k = 1, 2, \cdots$, and $y_{n_k} \to y$, but

$$-\left|\gamma/\beta\right| \left|x_{n_k}\right| \leq y_{n_k} \leq \left|\gamma/\beta\right| \left|x_{n_k}\right| \tag{39}$$

for all $k = 1, 2, \cdots$. Hence letting $k \to \infty$ in (39) yields that

$$y \in \left[-\left|\gamma/\beta\right| \left|x\right|, \left|\gamma/\beta\right| \left|x\right|\right] = A(x).$$

Therefore, Lemma 2.20 in [2] implies that Assumption 2.6a holds. □

4. Suboptimality of the VI-Policies

Let $(X, A, \{A(x) \mid x \in X\}, Q, c)$ be a fixed Markov control model.

Notation 4.1. a. For $\varsigma \subset X$, where ς is a nonempty compact set, denote $\mathbb{K}_\varsigma := \{(x, a) \in \mathbb{K} \mid x \in \varsigma, a \in A(x)\}$.

b. 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.

$$B_\varepsilon(f_n(x)) = \{a \in A(x) \mid d(a, f_n(x)) < \varepsilon\}. \tag{40}$$

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

$$B_\varepsilon^c(f_n(x)) = \{a \in A(x) \mid d(a, f_n(x)) \geq \varepsilon\}. \tag{41}$$

c. For each $\varepsilon > 0$, and $n = 1, 2, \cdots$, denote

$$\mathbb{K}_n := \{(x, a) \in \mathbb{K} \mid x \in \varsigma, a \in B_\varepsilon^c(f_n(x))\}. \tag{42}$$

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

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

where $V_{n-1}(\cdot)$ is the function given in (3),

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

f. Denote

$$D_n(x,a) = G_n(x,a) - V_n(x), \tag{43}$$

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

$$D(x,a) = G(x,a) - V^*(x), \tag{44}$$

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

For MDPs that satisfy Assumption 2.2, denote

$$\widehat{O}(n,\varsigma) = \overline{c}\,(\alpha\eta)^n\, \widehat{l}/(1-\alpha\eta), \tag{45}$$

for each $n = 1,2,\cdots$, where $\varsigma \subset X$ is a nonempty compact set, and take

$$\widehat{l} = \sup_{x\in\varsigma} w(x) \tag{46}$$

as finite.

Lemma 4.2. *Suppose that Assumptions 2.1, 2.2, 2.5, and 2.6 hold. Let $\varsigma \subset X$ be a nonempty compact set and let $\varepsilon > 0$. Then $\mathbb{K}_n$ is a compact set, for each $n = 1,2,\cdots$. Moreover, $\mathbb{K}_\varsigma$ is a compact set (observe that $\mathbb{K}_n \subset \mathbb{K}_\varsigma$, for all $n = 1,2,\cdots$).*

Proof. See Lemma 3.2 in [5]. □

The proof of Lemma 4.3 below is similar to the proof of Lemma 3.4 in [5]. However, for the completeness of the article that proof will be presented here.

Lemma 4.3. *Suppose that Assumptions 2.1, 2.2, 2.5, and 2.6 hold. Let ς be a compact fixed subset of X, and let $\varepsilon > 0$. Let n be a positive integer such that $\mathbb{K}_n \neq \emptyset$, and $\inf_{(x,a)\in\mathbb{K}_n} D_n(x,a) > 2\widehat{O}(n,\varsigma)$. Then,*

$$d(f_n(x), f^*(x)) < \varepsilon, \tag{47}$$

for all $x \in \varsigma$, i.e. f_n is an ε-approximation to f^* on ς (note that Lemma 4.2 implies that $\mathbb{K}_n$ is a compact set, and that (43) and Assumption 2.6 permit to obtain that $D_n(\cdot,\cdot)$ is a continuous function).

Proof. It will be shown that $D(x',a') > 0$ for all $(x',a') \in \mathbb{K}_n$. Observe that $D(x, f^*(x)) = 0$ for all $x \in X$ as a consequence of (6) and (44). Hence, for $x' \in \varsigma$ it follows that $(x', f^*(x')) \notin \mathbb{K}_n$ and, consequently, $f^*(x') \notin B^c_\varepsilon(f_n(x'))$, i.e. $d(f_n(x'), f^*(x')) < \varepsilon$.

Now, let $(x', a') \in \mathbb{K}_n$ be fixed. Using (7), (43), (44), (45), (46), and Assumption 2.2b, it follows that

$$\begin{aligned} D\left(x', a'\right) &= c\left(x', a'\right) + \alpha \int V^*(y)\, Q\left(dy \mid x', a'\right) - V^*\left(x'\right) \\ &= D_n\left(x', a'\right) + \alpha \int \left[V^*(y) - V_{n-1}(y)\right] Q\left(dy \mid x', a'\right) \\ &\quad + \left(V_n\left(x'\right) - V^*\left(x'\right)\right) \\ &\geq D_n\left(x', a'\right) - \alpha \frac{\overline{c}(\alpha\eta)^{n-1}}{1-\alpha\eta} \int w(y) Q\left(dy \mid x', a'\right) - O(n, x') \\ &\geq \inf_{(x,a)\in\mathbb{K}_n} D_n(x,a) - \frac{\overline{c}(\alpha\eta)^n}{1-\alpha\eta} w(x') - O(n, x') \\ &= \inf_{(x,a)\in\mathbb{K}_n} D_n(x,a) - 2O\left(n, x'\right) \\ &\geq \inf_{(x,a)\in\mathbb{K}_n} D_n(x,a) - 2\widehat{O}(n, \varsigma) > 0. \end{aligned}$$

Therefore, since (x', a') is arbitrary, Lemma 4.3 follows. □

Now, the main result of the paper will be presented.
Define, for $n = 1, 2, \cdots,$

$$\Lambda(n) = \sqrt{\frac{2\overline{c}(\alpha\eta)^n \widehat{l}}{(r + \alpha\beta^2 U_{n-1})(1-\alpha\eta)}} \tag{48}$$

(note that from Assumption 3.4, $1 - \alpha\eta > 0$).

Theorem 4.4. *Consider the one-dimensional LQ-model described in Example 3.3. Suppose that Assumption 3.4 holds. Let $\varsigma \subset \mathbb{R}$ be a nonempty fixed compact set. Let n be a positive integer.*

a. Assume that there is $z \in \varsigma$ such that $\Lambda(n) < |\frac{\gamma}{\beta} z| - f_n(z)$. Then for each $\varepsilon \in (\Lambda(n), |\frac{\gamma}{\beta} z| - f_n(z)]$ it follows that

$$| f_n(x) - f^*(x) | < \varepsilon, \tag{49}$$

for all $x \in \varsigma$.

b. Assume that there is $z \in \varsigma$ such that $\Lambda(n) < |\frac{\gamma}{\beta} z| + f_n(z)$. Then for each $\varepsilon \in (\Lambda(n), |\frac{\gamma}{\beta} z| + f_n(z)]$ it follows that

$$| f_n(x) - f^*(x) | < \varepsilon, \tag{50}$$

for all $x \in \varsigma$. (observe that since $f_n(z) \in A(z)$, it results that both $|\frac{\gamma}{\beta} z| + f_n(z)$ and $|\frac{\gamma}{\beta} z| - f_n(z)$ are positive).

Proof. Now the case a will be proved (the proof for the case b is similar).

Suppose that there is $z \in \varsigma$ such that $\Lambda(n) < |\frac{\gamma}{\beta}z| - f_n(z)$. Following the notation given in Lemma 3.5, it results that

$$\begin{aligned}
&\inf_{(x,a)\in\mathbb{K}_n} D_n(x,a) = \inf_{(x,a)\in\mathbb{K}_n} \left[c(x,a) + \alpha \int V_{n-1}(y) Q(dy \mid x,a) - V_n(x) \right] \\
&= \inf_{(x,a)\in\mathbb{K}_n} [qx^2 + ra^2 + \alpha U_{n-1}(\gamma x + \beta a)^2 + \alpha U_{n-1}\sigma^2 \\
&+ \alpha H_{n-1} - U_n x^2 - H_n] \qquad (51) \\
&= \inf_{(x,a)\in\mathbb{K}_n} \{(r + \alpha\beta^2 U_{n-1}) \left(a + \frac{\alpha U_{n-1}\beta^2\gamma^2}{r + \alpha\beta^2 U_{n-1}} x \right)^2 \\
&+ (R + SU_{n-1})^{-1}[(R + SU_{n-1})(q + \alpha U_{n-1}\gamma^2) \\
&- P - \overline{Q}U_{n-1} - \beta^2\gamma^2\alpha^2 U_{n-1}^2]x^2\} \\
&= \inf_{(x,a)\in\mathbb{K}_n} \left\{ (r + \alpha\beta^2 U_{n-1}) \left(a + \frac{\alpha\beta\gamma U_{n-1}}{R + SU_{n-1}} x \right)^2 \right\} \\
&= (r + \alpha\beta^2 U_{n-1}) \inf_{(x,a)\in\mathbb{K}_n} \left(a + \frac{\alpha\beta\gamma U_{n-1}}{R + SU_{n-1}} x \right)^2. \qquad (52)
\end{aligned}$$

Take $\varepsilon \in (\Lambda(n), |\frac{\gamma}{\beta}z| - f_n(z)]$.

Notice that by definition of $\mathbb{K}_n, (a - f_n(x))^2 \geq \varepsilon^2$, for all $(x,a) \in \mathbb{K}_n$, hence

$$\inf_{(x,a)\in\mathbb{K}_n} (a - f_n(x))^2 \geq \varepsilon^2. \qquad (53)$$

Since $(z, f_n(z) + \varepsilon) \in \mathbb{K}_n$ (observe that ε is positive, and that $0 < \varepsilon \leq |\frac{\gamma}{\beta}z| - f_n(z)$ implies that $-|\frac{\gamma}{\beta}z| < f_n(z) + \varepsilon \leq |\frac{\gamma}{\beta}z|$), and

$$(f_n(z) + \varepsilon - f_n(z))^2 = \varepsilon^2, \qquad (54)$$

from (53) it results that

$$\inf_{(x,a)\in\mathbb{K}_n} \left(a + \frac{\alpha\beta\gamma U_{n-1}}{R + SU_{n-1}} x \right)^2 = \inf_{(x,a)\in\mathbb{K}_n} (a - f_n(x))^2 = \varepsilon^2. \qquad (55)$$

Therefore from (52), it results that

$$\inf_{(x,a)\in\mathbb{K}_n} D_n(x,a) = (r + \alpha\beta^2 U_{n-1})\varepsilon^2. \qquad (56)$$

Finally, observe that

$$(r + \alpha\beta^2 U_{n-1})\varepsilon^2 > 2\overline{c}\,(\alpha\eta)^n\,\widehat{l}/\,(1 - \alpha\eta) \qquad (57)$$

if and only if

$$\varepsilon > \sqrt{\frac{2\overline{c}(\alpha\eta)^n\widehat{l}}{(r+\alpha\beta^2U_{n-1})(1-\alpha\eta)}} = \Lambda(n). \tag{58}$$

Hence, if

$$\varepsilon \in \left(\Lambda(n), |\frac{\gamma}{\beta}z| - f_n(z)\right], \tag{59}$$

using Lemma 4.3, it is obtained that

$$\mid f_n(x) - f^*(x) \mid < \varepsilon, \tag{60}$$

for all $x \in \varsigma$. □

Remark 4.5. a. Due to the fact that Examples 3.1 and 3.3 are equivalent, in the sense explained in the paragraph just previous to Example 3.3, it follows that f^* is also optimal for the classical LQ-model presented in Example 3.1. Therefore, under Assumption 3.4, Theorem 4.4 also provides a method to approximate uniformly on compact sets the optimal policy for Example 3.1.

b. Note that as a consequence of Assumption 2.2 and Lemma 3.5, it is obtained that $\Lambda(n) \to 0$, when $n \to \infty$.

c. Theorem 4.4 allows to obtain, for each n when $\mathbb{K}_n \neq \emptyset$, the *suboptimality* of f_n in the sense that determines the values of ε such that f_n is an ε-approximation of f^* on ς.

5. A Specific Example

Example 5.1. Consider Example 3.3 with $\gamma = 10/19, \beta = 1, q = 2,$ and $r = 5$, i.e. $x_{t+1} = \frac{10}{19}x_t + a_t + \xi_t, t = 0, 1, \cdots$, and $c(x,a) = 2x^2 + 5a^2$, $x \in \mathbb{R}, a \in [-\left|10x/19\right|, \left|10x/19\right|]$.

For Example 5.1 the following assumption will be taken into account.

Assumption 5.2. The disturbances ξ_t, $t = 0, 1, \cdots$ are i.i.d. random variables with values in $\widehat{S} = \mathbb{R}$. Moreover, suppose that ξ_0 has a continuous density Δ, zero mean value, and a finite variance $\sigma^2 = 2$.

Consider $\alpha = 9/10$.

Remark 5.3. Observe that the linear quadratic model proposed in Example 5.1 satisfies Assumption 3.4. For this example it is obtained that $\overline{c} = 1, \eta = 400/361$, w is given by $w(x) = \frac{1222}{361}x^2 + \frac{2444}{39}$, $x \in \mathbb{R}$, and for the compact set $[-100, 100]$, it is easy to verify that the value of $\widehat{l}$ is 33913.1000.

Moreover, straightforward computations allow to obtain that the main coefficient of the optimal policy f^* for Example 5.1 is an irrational number given by $2(-1763+\sqrt{32430394})/3249$ (see Remark 3.6).

The next array (constructed with the help of the formulas provided in Lemma 3.5 and Theorem 4.4), for the compact set $[-100,100]$, shows the sets given in Theorem 4.4a in which, for $n = 2000, 3000, 4000, 5000, 6000, 7000$ f_n is an ε-approximation to f^* on ς. Take $z = 100$.

(Note that in the computations there are considered only four decimals.)

n	U_{n-1}	$f_n(x), x \in \mathbb{R}$	$f_n(100)$	$(\Lambda(n), \vert\frac{10}{19}(100)\vert - f_n(100)]$
2000	2.4202	$-0.1597x$	-15.9700	$(115.2770, 294.6610]$
3000	2.4202	$-0.1597x$	-15.9700	$(28.7999, 294.6610]$
4000	2.4202	$-0.1597x$	-15.9700	$(7.1951, 294.6610]$
5000	2.4202	$-0.1597x$	-15.9700	$(1.7975, 294.6610]$
6000	2.4202	$-0.1597x$	-15.9700	$(0.4490, 294.6610]$
7000	2.4202	$-0.1597x$	-15.9700	$(0.1120, 294.6610]$

For instance, f_{7000} (given by $f_{7000}(x) = -0.1597x, x \in \mathbb{R}$) is suboptimal on $\varsigma = [-100, 100]$, for $\varepsilon \in (0.1120, 294.6610]$.

References

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

[2] 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*, **12**, No. 3 (2006), 133-148.

[3] 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.

[4] D. Cruz-Suárez, R. Montes-de-Oca, F. Salem-Silva, Pointwise approximations of discounted Markov decision processes to optimal policies, *International Journal of Pure and Applied Mathematics,* **28**, No. 2 (2006), 265-281.

[5] D. Cruz-Suárez, R. Montes-de-Oca, F. Salem-Silva, Uniform approximations of discounted Markov decision processes to optimal policies, in: *Proceedings of Prague Stochastics 2006* (Ed-s: M. Huskova, M. Janzura), Matfyzpress (2006), 278-287.

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

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