International Journal of Pure and Applied Mathematics

Volume 58 No. 1 2010, 77-85

OPTIMAL POLICIES IN THE CLASS OF INFINITELY DIFFERENTIABLE FUNCTIONS FOR DISCOUNTED LINEAR-QUADRATIC MODELS

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

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

[2]Facultad de Ciencias Físico-Matemáticas
Benemérita Universidad Autónoma de Puebla
San Claudio Av., San Manuel, Puebla, 1152, MEXICO
e-mail: hcs@fcfm.buap.mx

Abstract: This paper deals with one-dimensional linear-quadratic (LQ) models. These models are presented as Markov decision processes with the total discounted cost as the objective function. The discussion about the determination of the optimal policy of LQ models is divided into two cases: the analysis of the deterministic LQ models, and the analysis of the stochastic case. Specifically, (a) assuming that the optimal value function of the deterministic LQ models is the class C^{∞}, the optimal policy is obtained by means of the Euler equation, and (b) the optimal policy for stochastic LQ models is obtained from the dynamic programming equation (DPE), using as a fixed point of the DPE, the optimal value function for the deterministic case adjusted with a suitable additive constant.

AMS Subject Classification: 90C40, 93E20
Key Words: linear-quadratic model, discounted Markov decision process, optimal policy, Euler equation, Taylor's series around an equilibrium point

Received: December 15, 2009

§Correspondence author

1. Introduction

This article deals with one-dimensional linear-quadratic (LQ) models which are presented as discounted Markov decision processes (MDPs) with discrete-time and infinite horizon (see [1]).

The standard technique to find the solution of the LQ models has been mainly based on the value iteration algorithm to obtain the well-known Ricatti's equation while the corresponding optimal solution is obtained with a limit procedure (see [1]).

Here, a *different* approach is used in order to solve LQ models (in the best of our knowledge this approach has not been previously used to solve them). In particular, the LQ models are solved without the use of the value iteration algorithm. The solutions obtained are constructed directly on the dynamic programming equation (DPE) (see [5]). Obviously, the same solution as in the Ricatti's technique is obtained.

The solution presented is divided into two parts.

Firstly, deterministic LQ models, that is LQ models for which the dynamics of the systems do not include an additive random perturbation, are considered. Assuming that the optimal value function is of class C^{∞}, using the Euler equation, it is possible to obtain an optimal policy which is the expansion in Taylor's series around a convenient equilibrium point of the dynamics of the system. Then with the knowledge of the optimal policy, the optimal value function is computed.

Secondly, the stochastic case, that is LQ models with an additive random noise, is considered. In this part, the optimal value function of the deterministic LQ model plus a suitable constant which depends on the parameters of the random noise of the dynamics of the system is used as a fixed point in the DPE, obtaining that this function is the optimal value function for the stochastic case, and the optimal policy is the same as in the deterministic LQ model.

The approach presented in this article, was partially inspired in [2]. But, it is important to observe that Assumptions 3b and 3c in [2] are not necessary for the present work.

The paper is organized as follows. In Section 2, the basics on the Markov decision theory are presented, and in Section 3, the LQ models and their solutions are provided.

2. Discounted Markov Control Processes

Let $(X, A, \{A(x) : x \in X\}, G, c)$ be a Markov control model (see [1], [4] and [5]), which consists of the *state space* X, the *control set* A (where X and A are nonempty Borel subsets of Euclidean spaces). The Borel σ-algebras of X and A will be denoted by $\mathcal{B}(X)$ and $\mathcal{B}(A)$, respectively. $\{A(x) : x \in X\}$ is a family of nonempty measurable subsets $A(x) \in \mathcal{B}(A)$, whose elements are *feasible controls* when the system is in a state $x \in X$. The set $\mathbb{K} = \{(x, a) : x \in X, a \in A(x)\}$ of admissible state-actions pairs is assumed to be a measurable subset of the Cartesian product $X \times A$. The following component is the *transition law* G, and c is a real-valued measurable function on $\mathbb{K}$ called the *cost function*.

Let x_t and a_t be the state and the control at a time t, respectively. If the system is in the state $x_t = x$ at a time t and the control action $a_t = a \in A(x)$ is applied, then a cost $c(x, a)$ will be paid and the system moves to a new state x_{t+1} by means of the transition law, given by the following difference equation:

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

$t = 0, 1, ...$, where $\{\xi_t\}$ are independent and identically distributed (i.i.d.) random variables taking values on a Borel space S with density Δ. Let ξ be a generic element of the sequence $\{\xi_t\}$. $G : \mathbb{K} \times S \to X$ is a given measurable function.

The transition law (1) induces a stochastic kernel given by

$$Q(B\,|x, a) = \int I_B(G(x, a, s))\Delta(s)ds,$$

$B \in \mathcal{B}(X)$ and $(x, a) \in \mathbb{K}$, where $I_B(\cdot)$ denotes the indicator function of a measurable set B.

Remark 2.1. A stochastic kernel Q satisfies the following properties: a) $Q(B\,|\cdot)$ is a measurable function on $\mathbb{K}$, for each $B \in \mathcal{B}(X)$, and b) $Q(\cdot\,|k)$ is a probability measure on $\mathcal{B}(X)$, for each $k \in \mathbb{K}$.

In some cases the transition law is given by a deterministic difference equation

$$x_{t+1} = F(x_t, a_t), \tag{2}$$

$t = 0, 1, ...$ and $x_0 = x \in X$ where $F : \mathbb{K} \to X$ is a given measurable function. Observe that in this case the stochastic kernel induced by (2) is

$$Q(B\,|x, a) = I_B(F(x, a)),$$

$B \in \mathcal{B}(X)$ and $(x, a) \in \mathbb{K}$.

In general, a policy is a sequence $\pi = \{\pi_t : t = 0, 1, ...\}$ of stochastic kernels, defined on A given the history of the controlled process. The set of policies will be denoted by Π. Let $\mathbb{F} := \{f : X \to A : f$ be measurable and $f(x) \in A(x), x \in X\}$. A sequence $\pi = \{f_t : t = 0, 1, ...\}$ of functions $f_t \in \mathbb{F}$ is called a *Markov policy*. The set of the Markov policies will be denoted by $\mathbb{M}$. A Markov policy $\pi = \{f_t : t = 0, 1, ...\}$ is said to be a *stationary policy* if $f_t = f \in \mathbb{F}$, for all t.

Let $(X, A, \{A(x) : x \in X\}, G, c)$ be a fixed control model. For each policy $\pi \in \mathbb{F}$ and state $x \in X$, it is defined that

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

$v(\pi, x)$ is called the *total discounted cost*, where $\alpha \in (0, 1)$ is the discount factor, and E_x^π denotes the expectation with respect to the canonical probability P_x^π (see [5] for the construction of P_x^π).

The *optimal control problem* consists of determining a policy $\pi^* \in \mathbb{F}$, such that

$$v(\pi^*, x) = \inf_{\pi \in \mathbb{F}} v(\pi, x),$$

$x \in X$, and π^* will be called the *optimal policy*. The function V defined by

$$V(x) = \inf_{\pi \in \mathbb{F}} v(\pi, x),$$

$x \in X$, will be called the *optimal value function*.

Remark 2.2. In fact the optimal control problem could be established on the class Π of all policies, but in this work the existence of stationary optimal policies will be assumed for the models taken into account. Hence

$$\inf_{\pi \in \Pi} v(\pi, x) = \inf_{\pi \in \mathbb{F}} v(\pi, x),$$

for all $x \in X$.

Lemma 2.3. *Under certain assumptions (see Remark 2.4 below), it results that:*

a) The optimal value function V is a solution for the following equation (known as the Dynamic Programming Equation):

$$V(x) = \min_{a \in A(x)} \{c(x, a) + \alpha E [V(G(x, a, \xi))]\}, \tag{3}$$

$x \in X$.

b) There exists $f \in \mathbb{F}$ such that $f(x) \in A(x), x \in X$, attains the minimum

on the right-hand side of (3), i.e.,

$$V(x) = c(x, f(x)) + \alpha E\left[V(G(x, f(x), \xi))\right], \tag{4}$$

$x \in X$ and f is optimal.

Remark 2.4. The assumptions and the proof of Lemma 2.3 can be consulted in [5]. In particular, if the transition law is given by (2), a similar lemma holds and the assumptions can be verified in the following way. If the cost function is bounded, see conditions (a), (b) and (c) in Theorem 2.8 in [4], p. 23. And for unbounded costs, see conditions in Theorems 3, 4, 5, and 6 in [6].

Remark 2.5. Throughout the rest of the paper, MCPs which have the structure of linear-quadratic models are taken into account. It is well-known that for each of these MCPs, (a) the corresponding optimal value function satisfies (3), and (b) there exists an optimal policy characterized by (4) (see, [1], [5], and [6]).

Definition 2.6. An *equilibrium point* $\overline{x}$ of the system (2) is defined by the equation

$$\overline{x} = F(\overline{x}, f(\overline{x})),$$

where f is the optimal policy of the optimal control problem.

3. Linear-Quadratic Models

3.1. Deterministic Version

In this subsection there is considered an LQ problem with a stationary, scalar, deterministic, linear system

$$x_{t+1} = \gamma x_t + \beta a_t,$$

$t = 0, 1, 2, ...$ and $x_0 = x$, where γ and β are real constants with $\gamma \cdot \beta \neq 0$. The cost function is given by

$$c(x_t, a_t) = q x_t^2 + r a_t^2,$$

$t = 0, 1, 2, ...$, where $q > 0$ and $r > 0$ are given constants. It will be assumed that the space of states is $X = \mathbb{R}$, the space of controls is $A = \mathbb{R}$ and the problem is unconstrained, i.e. $A(x) = A, x \in X$.

Notation. For deterministic LQ models, the optimal value function will be denoted by W and the optimal policy by g (see Remark 2.5).

For a positive integer p, $C^p(X, \mathbb{R}) = \{\theta : X \to \mathbb{R} :$ the first derivatives of θ

exist and are continuous}, and $C^\infty(X,\mathbb{R}) = \{\theta : X \to \mathbb{R} : \theta \in C^p(X,\mathbb{R})$, for all positive integer $p\}$.

Assumption I. $W \in C^\infty(X,\mathbb{R})$.

Lemma 3.1. *Under Assumption I,* $g \in C^\infty(X,A)$.

Proof. Let x be a fixed state. Suppose that for each positive integer $p \geq 1$, $W \in C^p(X,\mathbb{R})$. Let

$$\widehat{G}(x,a) := qx^2 + ra^2 + \alpha W(\gamma x + \beta a),$$

$a \in A(x)$. Since the optimal policy g takes values in $A(x) = \mathbb{R}$, and $\mathbb{R}$ is open, it follows that

$$\widehat{G}_a(x, g(x)) = 0,$$

and $\widehat{G}_{aa}(x, g(x)) > 0$. Then, using the Implicit Function Theorem (see [3], p. 205) it is obtained that $g(x) \in C^{p-1}(X,A)$. Since x is an arbitrary state, Lemma 3.1 follows. □

Lemma 3.2. *The optimal policy g satisfies the following equation*

$$rg(x) = \alpha\gamma r g(\gamma x + \beta g(x)) - q\alpha\beta\gamma x - q\alpha\beta^2 g(x), \tag{5}$$

$x \in \mathbb{R}$. *This equation is known as Euler equation (EE). In a similar way the EE for the optimal value function W is given by*

$$\alpha\gamma W'\left(\gamma x + \beta^2 \frac{(2qx - W'(x))}{2\gamma r}\right) = W'(x) - 2qx, \tag{6}$$

$x \in \mathbb{R}$.

Proof. Let $x \in \mathbb{R}$ be fixed. The DPE for this problem is given by

$$W(x) = \min_{a\in\mathbb{R}} \left[qx^2 + ra^2 + \alpha W(\gamma x + \beta a)\right].$$

Then first-order condition is given by

$$2rg(x) + \alpha W'(\gamma x + \beta g(x))\beta = 0,$$

where g is the optimal policy.

On the other hand, g satisfies

$$W(x) = qx^2 + rg(x)^2 + \alpha W(\gamma x + \beta g(x)).$$

Derivating the last equation with respect to x and using the first-order condition, it is obtained that

$$W'(x) = 2qx + \alpha W'(\gamma x + \beta g(x))\gamma. \tag{7}$$

Again using the first-order condition in (7), it results that

$$W'(x) = 2qx - 2\gamma r\beta^{-1} g(x). \tag{8}$$

Substituting (8) in the first-order condition, (5) yields.

To obtain the EE in terms of the value function, the following approach is used. From (8) it is obtained that

$$g(x) = \left(2qx - W'(x)\right)\beta(2\gamma r)^{-1}.$$

Substituting the last equation in the first-order condition, (6) holds. □

Theorem 3.3. *The optimal policy and the optimal value function are given, respectively, by*

$$g(x) = \lambda x \quad \text{and} \quad W(x) = \frac{q\beta - \gamma r\lambda}{\beta}x^2,$$

$x \in \mathbb{R}$, *where* λ *satisfies the following:*

$$\alpha\beta\gamma r\lambda^2 + \left[\alpha\left(\gamma^2 r - q\beta^2\right) - r\right]\lambda - q\alpha\beta\gamma = 0 \text{ and} \tag{9}$$

$$|\gamma + \beta\lambda| \leq 1.$$

Proof. Consider the EE (5). This equation is not possible to solve, in general, for each $x \in X$. That is why (5) will be solved, initially, at the equilibrium point $\overline{x}$. The equilibrium point satisfies the following equation

$$\overline{x} = \gamma\overline{x} + \beta g(\overline{x}). \tag{10}$$

The EE at the equilibrium point can be reduced, using (10), in the equation

$$rg(\overline{x}) = \alpha\gamma r g(\overline{x}) - q\alpha\beta\gamma\overline{x} - q\alpha\beta^2 g(\overline{x}). \tag{11}$$

Solving (10) and (11) simultaneously it is obtained that it is possible to consider $\overline{x} = 0$, while $g(\overline{x}) = 0$ necessarily.

Now derivating (5) implicitly with respect to x, it results that

$$rg'(x) = -q\alpha\beta\gamma - \alpha\beta^2 q g'(x) + \alpha\gamma r g'(\gamma x + \beta g(x))(\gamma + \beta g'(x)),$$

$x \in \mathbb{R}$. Evaluating the previous equation at the equilibrium point $\overline{x} = 0$, it is obtained that

$$\alpha\beta\gamma r g'(0)^2 + g'(0)\left[\alpha(\gamma^2 r - q\beta^2) - r\right] - q\alpha\beta\gamma = 0. \tag{12}$$

It is direct to verify that the quadratic equation (12) has at least one real solution for $g'(0)$.

Derivating again equation (5) with respect to x, it results that

$$\begin{aligned} rg''(x) &= -\alpha\beta^2 q g''(x) + \alpha\gamma r[g''(\gamma x + \beta g(x))(\gamma + \beta g'(x))^2 \\ &\quad + g'(\gamma x + \beta g(x))\beta g''(x)]. \end{aligned}$$

Considering in the last equation that $x = \overline{x} = 0$, it is obtained that $g''(0) = 0$.

In general, it is possible to obtain that $g^{(k)}(0) = 0$, for $k > 1$.

Now, making an expansion in Taylor's series (see [3]) around $x = 0$ and

using the information obtained, it results that

$$g(x) = \lambda x,$$

$x \in \mathbb{R}$, where $\lambda = g'(0)$ is a real solution of (12) and $|\gamma + \beta\lambda| \leq 1$, and

$$W(x) = v(g, x) = (q + r\lambda)x^2 \sum_{t=0}^{\infty} \left[\alpha(\gamma + \beta\lambda)^2\right]^t, \tag{13}$$

$x \in \mathbb{R}$. (Observe that it is possible to choose λ such that $|\gamma + \beta\lambda| \leq 1$, because it is assumed that $v(g, x)$, $x \in \mathbb{R}$, is finite.) □

Remark 3.4. It is important to observe that the EE (see (6)) could be obtained in terms of the optimal value function W, and in this case there will be obtained a quadratic equation similar to equation (9) which corresponds to the Riccati's equation (see [1]).

3.2. Stochastic Version

Now, consider the following variant of the LQ model:

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

$t = 0, 1, 2, ...$ and $x_0 = x$, where $\{\xi_t\}$ are i.i.d. random variables with zero mean, finite variance σ^2 and density Δ. This model is the deterministic LQ given in the previous subsection, perturbed with an additive random noise ξ. The solution in this case can be induced using the solution given in the previous subsection, as the following theorem suggests.

Notation. For stochastic LQ models, the optimal value function will be denoted by V and the optimal policy by f (see Remark 2.5).

Theorem 3.5. *The LQ model with dynamics (14) has the following solution:*

$$\begin{aligned} V(x) &= \frac{(q + r\lambda)x^2}{1 - \alpha(\gamma + \beta\lambda)^2} + \frac{\alpha}{1 - \alpha} M\sigma^2, \\ f(x) &= \lambda x, \end{aligned}$$

$x \in R$, where λ satisfies the conditions of Theorem 3.3, and $M = (q\beta - \gamma r\lambda)/\beta$.

Proof. The idea of the proof is to induce the solution using the deterministic case. For this consider

$$V(x) = W(x) + N, \tag{15}$$

$x \in R$, where W is the optimal value function of the previous problem (see (13)) and N is a constant to determine. The constant N will be adjusted using

the DPE, substituting (15) in the DPE:

$$\begin{aligned} W(x)+N &= \min_{a\in A(x)}\left[qx^2+ra^2+\alpha\int W(\gamma x+\beta a+s)\Delta(s)ds+\alpha N\right] \\ &= \min_{a\in A(x)}\left[qx^2+ra^2+\alpha\int W(\gamma x+\beta a)\Delta(s)ds\right]+\alpha N+\alpha M\sigma^2 \\ &= W(x)+\alpha N+\alpha M\sigma^2, \end{aligned}$$

$x\in\mathbb{R}$. Then

$$N=\frac{\alpha}{1-\alpha}M\sigma^2,$$

and the optimal policy is g, i.e. $f=g$. □

Acknowledgments

This work was partially supported by Consejo Nacional de Ciencia y Tecnología (CONACyT) under grant PCI6480408.

References

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

[2] H. Cruz-Suárez, R. Montes-de-Oca, Discounted Markov control processes induced by deterministic systems, *Kybernetika*, Prague, **42**, No. 6 (2006), 647-664.

[3] A. de la Fuente, *Mathematical Methods and Models for Economists*, Cambridge University Press, Cambridge (2000).

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

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

[6] J.L. Rincón-Zapatero, C. Rodríguez-Palmero, Existence and uniqueness of solutions to the Bellman equation in the unbounded case, *Econometrica*, **71** (2003), 1519-1555.