LINEAR-QUADRATIC OPTIMAL CONTROL PROBLEM WITH A RANDOM HORIZON

Rocio Ilhuicatzi–Roldán*
Universidad Autónoma de Tlaxcala
Facultad de Ciencias Básicas, Ingeniería y Tecnología
Av. Ángel Solana s/n, San Luis Apizaquito, Apizaco, Tlax., 90300, México

ABSTRACT
In this paper a Linear-Quadratic control model of a discrete-time is considered (the transition law is a linear difference equation and the cost per stage has a quadratic form). Also, the expected total cost with a random horizon is considered as performance criterion, it is assumed that the horizon is independent of the control process. For the corresponding control problem the existence of the optimal solution is proved when the support of the distribution of the horizon is infinite, the proof is based on recent theoretical results concerning to the Markov decision processes with a random horizon. In addition, the rolling horizon procedure is used to obtain a control policy for the approximation of the optimal solution in the Linear-Quadratic control problem. The policy is provided through recursive equations which are programmed. In numerical cases is observed that even the policies with a small length of the rolling horizon provide good performance and a convergence of selectors in the policy of rolling horizon is observed, which allow to change the policy of rolling horizon by a stationary policy.

KEYWORDS: Optimal Stochastic Control; Dynamic Programming; Markov Decision Process; Linear-Quadratic Model.

MSC: 93E20.

RESUMEN
En este artículo se toma en cuenta un modelo de control de tipo Lineal-Cuadrático en tiempo discreto (la ley de transición es una ecuación lineal en diferencias y el costo por etapa tiene una forma cuadrática). Además, se considera como el criterio de rendimiento el costo total esperado con un horizonte aleatorio, suponiendo que el horizonte es independiente del proceso de control. Para el correspondiente problema de control, se prueba la existencia de la solución óptima cuando el soporte de la distribución del horizonte es infinito, la prueba es basada en recientes resultados teóricos concernientes a los procesos de decisión de Markov con un horizonte aleatorio. Adicionalmente, se utiliza el procedimiento de horizonte rodante para obtener una política que aproxime la solución óptima del problema de control Lineal-Cuadrático. La política propuesta es dada mediante ecuaciones recursivas las cuales han sido programadas. En casos numéricos se observa que incluso las políticas con una longitud pequeña del horizonte rodante proporcionan buen rendimiento y puede notarse una convergencia de los selectores que forman la política de horizonte rodante, lo cual permitirá cambiar la política de horizonte rodante por una política estacionaria.

1. INTRODUCTION

In this paper a control model of a discrete-time is considered where the transition law is given by a linear equation in differences and the cost per stage has a quadratic form. The corresponding optimal control problem is also known as the Linear-Quadratic (LQ) control problem or the LQ regulator problem, it is of

*rocioil@hotmail.com

great importance mainly in the area of engineering and economy. In addition, for this problem the expected total cost with a random horizon of an infinite support and independent of the control process is considered as the performance criterion.

The LQ control problem has been solved in [3] p. 34, here the expected total cost with a finite horizon is considered as the performance criterion. Also, in [3] p. 70, the LQ control problem is solved considering the expected total discounted cost with an infinite horizon as the performance criterion. In latest latest case, the discount factor is economically interpreted as a rate of return of the cost per stage. However, in [6] p. 126, it has shown that the criterion "the expected total discounted cost with an infinite horizon" and "the expected total cost with a random horizon with geometric distribution" are equivalent. Since in applications to engineering (see [5]), the cost per stage in the LQ control problem is interpreted as an abstraction of the energy and is not exactly a monetary cost, it is justifiable to use a random horizon rather than the rate of return, assuming that an unexpected event can end the process of the system at a random time.

In [4], the LQ control problem of matrix type is solved considering a random horizon with arbitrary distribution it having a finite support. Now, in this paper, a solution for the LQ control problem with a random horizon considering an arbitrary distribution of an infinite support is proposed. For this, first the existence of the optimal solution is proved using recent theoretical results concerning to the Markov decision processes with a random horizon (see [2]). Due to limitations to obtain the exact solution, a policy of rolling horizon and the corresponding performance function are obtained as an alternative to approximate the optimal solution for the Linear-Quadratic control problem. The solution is provided through programmable recursive equations. In numerical cases some policies of rolling horizon with different length are calculated and is observed that even the policies with a small length of rolling horizon provide good performance. Also, a numerical convergence of the selectors in the policy of rolling horizon to a stationary selector is observed, which allow to change the policy of rolling horizon by a stationary policy. Such stationary selector is in the optimal policy of a particular discounted LQ control problem that is associated to the LQ control problem with a random horizon.

This paper is organized as follows. Firstly, in Section 2, the theory of the Markov decision processes and general results of the Markov decision processes with a random horizon are presented. Afterwards, in Section 3, the Linear-Quadratic model with a random horizon is described and the existence of the optimal solution for the LQ control problem is proved. Finally, in Section 4, a policy of rolling horizon for the LQ control problem with a random horizon is provided, which can be used to approximate the optimal solution and the performance function is obtained, also a numerical results are presented.

2. BASIC THEORY OF THE MARKOV DECISION PROCESSES

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

A policy is a sequence $\pi = \{\pi_t : t = 0, 1, \ldots\}$ of stochastic kernels π_t on the control set A given the history $\mathbb{H}_t$ of the process up to time t where $\mathbb{H}_0 = X$ and $\mathbb{H}_t = \mathbb{K} \times \mathbb{H}_{t-1}$, $t = 1, 2, \ldots$. The set of all policies is denoted by Π.
$\mathbb{F}$ denotes the set of measurable functions $f : X \to A$ such that $f(x) \in A(x)$, for all $x \in X$. The functions in $\mathbb{F}$ are called selectors of the multifunction $x \mapsto A(x)$.
A deterministic Markov policy is a sequence $\pi = \{f_t\}$ such that $f_t \in \mathbb{F}$, for $t = 0, 1, 2, \ldots$. If $f_t = f$ for all $t = 0, 1, 2, \ldots$ the policy is a deterministic stationary policy.

In many cases, the evolution of a Markov control process is specified by a discrete time or difference equation of the form $x_{t+1} = F(x_t, a_t, \xi_t)$, $t = 0, 1, 2, \ldots$, with x_0 given, where $\{\xi_t\}$ is a sequence of independent and identically distributed random variables with values in a Borel space S and a common distribution μ, independent of the initial state x_0. In this case, the transition law Q is given by

$$Q(B|x,a) = \int_S I_B\big(F(x,a,s)\big)\mu(ds),$$

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

Let $(\Omega, \mathcal{F})$ be the measurable space consisting of the canonical sample space $\Omega = \mathbb{H}_\infty := (X \times A)^\infty$ and $\mathcal{F}$ as the corresponding product σ-algebra. Let $\pi = \{\pi_t\}$ be an arbitrary policy and μ be an arbitrary probability measure on X called the initial distribution. Then, by the theorem of C. Ionescu-Tulcea (see [3]), there exists a unique probability measure P^π_μ on $(\Omega, \mathcal{F})$ such that $P^\pi_\mu(\mathbb{H}_\infty) = 1$. The stochastic process $(\Omega, \mathcal{F}, P^\pi_\mu, \{x_t\})$ is called a discrete-time Markov control process or a Markov decision process. The expectation operator with respect to P^π_μ is denoted by E^π_μ. If μ is concentrated at the initial state $x \in X$, then P^π_μ and E^π_μ are written as P^π_x and E^π_x, respectively.

Let $(\Omega', \mathcal{F}', P)$ be a probability space and let $(X, A, \{A(x) \mid x \in X\}, Q, c)$ be a Markov decision model. Define the performance criterion as

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

$\pi \in \Pi$, $x \in X$, where τ is considered as a random variable on $(\Omega', \mathcal{F}')$ with the probability distribution $P(\tau = t) = \rho_t$, $t = 0, 1, 2, \ldots$ (i.e. with an infinite support). Moreover, it is assumed that for each $x \in X$ and $\pi \in \Pi$ the induced process $\{(x_t, a_t) \mid t = 0, 1, 2, \ldots\}$ is independent of τ.
The optimal value function is defined as

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

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

In [2] is observed that

$$\begin{aligned} E\left[\sum_{t=0}^{\tau} c(x_t, a_t)\right] &= E^\pi_x\left[\sum_{t=0}^{\infty} P_t c(x_t, a_t)\right] \\ &= E^\pi_x\left[\sum_{t=0}^{\infty} \prod_{k=n}^{t} \alpha_{k-1} c(x_t, a_t)\right], \end{aligned}$$

where

$$\alpha_t = P(\tau \geq t+1 \mid \tau \geq t) = \frac{P_{t+1}}{P_t} \qquad t = 0, 1, 2, \ldots, \tag{1}$$

with $P_t = P(\tau \geq t)$ and $\alpha_{-1} = P_0 = 1$ is considered.

The following definitions and results are retaken of [2].
For each $n = 0, 1, 2, \ldots$, defines

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

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

$$V_n^\tau(x) := \inf_{\pi \in \Pi} v_n^\tau(\pi, x),$$

$x \in X$.

Remark 2.1. *Observe that the values α_k in the definition of the functions v_n^τ are obtained means of the distribution of τ (see (1)), and with $n = 0$ it is obtained that $v_n^\tau(\pi, x) = \jmath^\tau(\pi, x)$.*

For $N > n \geq 0$, it is defined that

$$v_{n,N}^\tau(\pi, x) := E_x^\pi \left[\sum_{t=n}^{N} \prod_{k=n}^{t} \alpha_{k-1} c(x_t, a_t) \right], \tag{2}$$

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

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

$x \in X$.

Assumption 1

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

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

(c) There exists a policy $\pi \in \Pi$ such that $\jmath^\tau(\pi, x) < \infty$ for each $x \in X$.

Lemma 2.1. *Suppose that Assumption 1 holds. Then, for every $n \geq 0$ and $x \in X$,*

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

and V_n^τ is lower semicontinuous.

Theorem 2.1. *Suppose that Assumption 1 holds, then*

(a) the optimal value function V_n^τ, $n = 0, 1, 2, \ldots$, satisfies the optimality equation

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

$x \in X$, and if $\{u_n\}$ is another sequence that satisfies the optimality equations in (3), then $u_n \geq V_n^\tau$.

(b) There exists a policy $\pi^* = \{f_n \in \mathbb{F} : n \geq 0\}$ such that, for each $n = 0, 1, 2, \ldots$, the selector $f_n(x) \in A(x)$ attains the minimum in (3), i.e.

$$V_n^\tau(x) = c(x, f_n(x)) + \alpha_n \int_X V_{n+1}^\tau(y) Q(dy \mid x, f_n(x)),$$

$x \in X$, and the policy π^* is optimal.

3. LQ CONTROL PROBLEM WITH A RANDOM HORIZON AND THE EXISTENCE OF THE OPTIMAL SOLUTION

LQ control model with a random horizon is defined as follow: Let $X = A = A(x) = \mathbb{R}$. The cost function per stage is given by

$$c(x, a) = qx^2 + ra^2,$$

$(x, a) \in \mathbb{K}$, where $q \geq 0$ and $r > 0$. The transition law is induced by the following difference equation:

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

$t = 0, 1, 2, \ldots, \tau$, with x_0 known. In this case, $\gamma, \beta \in \mathbb{R}$ and $\{\xi_t\}$ is a sequence of independent and identically distributed random variables taking values in $S = \mathbb{R}$ with a continuous bounded density function Δ, such that $E[\xi_0] = 0$ and $E[\xi_0^2] = \sigma^2 < +\infty$, where ξ_0 is a generic element of the sequence $\{\xi_t\}$. Moreover, it is assumed that the distribution function of the random horizon τ has an infinite support and $E[\tau] < \infty$.

Lemma 3.2. *The LQ Model with a random horizon satisfies Assumption 1.*

Proof: Clearly, c is nonnegative and continuous. Next, let $A_\lambda(c) := \{a \in A(x) : c(x, a) \leq \lambda\}$, $\lambda \in \mathbb{R}$. Then,

$$A_\lambda(c) = \begin{cases} \emptyset & \text{if } \lambda < qx^2 \\ \{0\} & \text{if } \lambda = qx^2 \\ \left[-\sqrt{\frac{\lambda - qx^2}{r}}, \sqrt{\frac{\lambda - qx^2}{r}}\right] & \text{if } \lambda > qx^2 \end{cases}$$

Since $A_\lambda(c)$ is compact for each $x \in \mathbb{R}$, then c is a inf-compact function on $\mathbb{K}$. Now, it is verified that the transition law is strongly continuous. Let $v : X \to \mathbb{R}$ a measurable bounded function. Observe that

$$v'(x, a) = \int_X v(y) Q(dy \mid x, a) = \int_\infty^\infty v(\gamma x + \beta a + s)\Delta(s) ds.$$

Making the change of variable $u = \gamma x + \beta a + s$, it is obtained that

$$v'(x, a) = \int_\infty^\infty v(u)\Delta(u - \gamma x - \beta a) du.$$

Let $\{(x_k, a_k)\}$ a sequence such that $\lim_{k\to\infty}(x_k, a_k) = (x', a')$. By Dominated Convergence Theorem and the continuity of Δ, then $\lim_{k\to\infty} v'(x_k, a_k) = v'(x', a')$, hence v' is continuous and bounded on $\mathbb{K}$. Thus, the transition law for the LQ Model is strongly continuous. Finally, to prove Assumption 1(c), the stationary policy $h(x) = -\frac{\gamma}{\beta}x$ and $x_0 = x \in X$ are considered. In this case, applying the dynamic $x_{t+1} = \gamma x_t + \beta a_t + \xi_t$ of the system it is obtained that

$$\begin{aligned} x_1 &= \gamma x_0 + \beta a_0 + \xi_0 = \gamma x - \beta \frac{\gamma}{\beta} x + \xi_0 = \xi_0 \\ x_2 &= \gamma x_1 + \beta a_1 + \xi_1 = \gamma \xi_0 - \beta \frac{\gamma}{\beta} \xi_0 + \xi_1 = \xi_1 \\ x_3 &= \gamma x_2 + \beta a_2 + \xi_2 = \gamma \xi_1 - \beta \frac{\gamma}{\beta} \xi_1 + \xi_2 = \xi_2 \\ &\vdots \\ x_t &= \gamma x_{t-1} + \beta a_{t-1} + \xi_{t-1} = \gamma \xi_{t-2} - \beta \frac{\gamma}{\beta} \xi_{t-2} + \xi_{t-1} = \xi_{t-1}. \end{aligned}$$

Then

$$\begin{aligned} J^{\tau}(h,x) &= E_x^h\left[\sum_{t=0}^{\infty} P_t\left(q+r\frac{\gamma^2}{\beta^2}\right)\xi_{t-1}^2\right] \\ &= \left(q+r\frac{\gamma^2}{\beta^2}\right)\sigma^2(E[\tau]+1)<\infty. \end{aligned}$$

■

Once that is proved the Assumption 1, Lemma 2.1. and Theorem 2.1. provides the existence of the optimal solution for the LQ control problem with a random horizon.

4. APPROXIMATION OF THE OPTIMAL SOLUTION

In this section, an approximation of the optimal solution is obtained for the LQ control problem with a random horizon through the rolling horizon procedure (see [1]).

Next, the rolling horizon algorithm is presented.

Algorithm 1. 1. *Set $m=0$ and $n=\mathbf{N}$.*

2. *Find the policy $\pi^* = (\pi_m^*, \pi_{m+1}^*, \ldots, \pi_{n-1}^*)$, which is optimal for periods from m to n, and set $\widehat{\pi}_m = \pi_m^*$.*
3. *Let $m=m+1$ and $n=n+1$.*
4. *Go to step 2.*

The policy $\widehat{\pi} = (\widehat{\pi}_0, \widehat{\pi}_1, \widehat{\pi}_2, \ldots)$ is called a rolling horizon policy. $\mathbf{N}$ is the length of the rolling horizon and it is a positive integer.

In accordance with Algorithm 1, the decision corresponding to stage $m=t$, $t=0,1,2,\ldots$, of the rolling horizon policy is the first decision in the optimal policy of the control problem that starts at step t and ends to step $t+\mathbf{N}$. Assuming a zero terminal cost, observe that of (2), the corresponding criterion in the last problem is $v_{t,t+\mathbf{N}-1}^{\tau}$ and to solve it is used Theorem 4.1 given in [2] with the dynamic programming equation written for the LQ model as follows:

$$\begin{aligned} U_{t,\mathbf{N}}(x) &= 0 \\ U_{t,k}(x) &= \min_{a\in\mathbb{R}}\left[qx^2+ra^2+\alpha_{t+k}E[U_{t,k+1}(\gamma x+\beta a+\xi)]\right], \quad k=\mathbf{N}-1,\mathbf{N}-2,\ldots,0. \end{aligned} \tag{4}$$

Lemma 4.3. *The rolling horizon policy $\widehat{\pi}$ of N stages for the LQ control problem with a random horizon is the following: $\widehat{\pi} = (\widehat{f}_0, \widehat{f}_1, \widehat{f}_2, \ldots)$ with $\widehat{f}_t(x) = \lambda_t x$, $t=0,1,2,\ldots$, $x\in X$, where*

$$\lambda_t = \frac{-\alpha_t C_{t,1}\gamma\beta}{r+\alpha_t C_{t,1}\beta^2} \tag{5}$$

and the constant $C_{t,1}$ is obtained through the recurrence relation:

$$\begin{aligned} C_{t,N} &= 0, \\ C_{t,k} &= \frac{qr+\alpha_{t+k}C_{t,k+1}(q\beta^2+r\gamma^2)}{r+\alpha_{t+k}C_{t,k+1}\beta^2}, \quad k=N-1,N-2,\ldots,1. \end{aligned}$$

The performance function is given by $v_0^{\tau}(\widehat{\pi},x) = K_1x^2+K_2$, $x\in X$, where

$$K_1 = P_0(q+r\lambda_0^2)+P_1(q+r\lambda_1^2)(\gamma+\beta\lambda_0)^2+\sum_{t=2}^{\infty}P_t(q+r\lambda_t^2)\prod_{i=0}^{t-1}(r+\beta\lambda_i)^2 \tag{6}$$

and

$$K_2 = \sigma^2 P_1(q + r\lambda_1^2) + \sigma^2 \sum_{t=2}^{\infty} P_t(q + r\lambda_t^2) \left(1 + \sum_{i=0}^{t-2} \prod_{j=i+1}^{t-1} (\gamma + \beta\lambda_j)^2 \right). \tag{7}$$

Proof: In the rolling horizon procedure, considers the problem that starts in step $m = t$ with arbitrary t. Using the dynamic programming equation (4) with $k = \mathbf{N} - 1$, it is obtained that $f_{t,\mathbf{N}-1}(x) = 0$ and $U_{t,\mathbf{N}-1}(x) = qx^2$, $x \in X$.
For $k = \mathbf{N} - 2$, replacing $U_{t,\mathbf{N}-1}$ in (4), it is obtained that

$$\begin{aligned} U_{t,\mathbf{N}-2}(x) &= \min_{a\in\mathbb{R}} \left[qx^2 + ra^2 + \alpha_{t+\mathbf{N}-2} E[q(\gamma x + \beta a + \xi)^2] \right] \\ &= \min_{a\in\mathbb{R}} \left[qx^2 + ra^2 + \alpha_{t+\mathbf{N}-2} q(\gamma^2 x^2 + \beta^2 a^2 + 2\gamma\beta x a + \sigma^2) \right] \\ &= \min_{a\in\mathbb{R}} \left[(r + \alpha_{t+\mathbf{N}-2} q\beta^2) a^2 + 2\alpha_{t+\mathbf{N}-2} q\gamma\beta x a + q(\alpha_{t+\mathbf{N}-2}\gamma^2 + 1)x^2 + q\alpha_{t+\mathbf{N}-2}\sigma^2 \right], \end{aligned}$$

deriving the expression in brackets with respect to the variable a and equating to zero for the minimization, it is obtained that

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

where $C_{t,\mathbf{N}-1} = q$. Therefore $U_{k,\mathbf{N}-2}(x) = C_{t,\mathbf{N}-2}x^2 + D_{t,\mathbf{N}-2}$, where

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

and $D_{t,\mathbf{N}-2} = C_{t,\mathbf{N}-1}\alpha_{t+\mathbf{N}-2}\sigma^2$.
Continuing with this process, it follows that

$$f_{t,\mathbf{N}-3}(x) = \frac{-\alpha_{t+\mathbf{N}-3} C_{t,\mathbf{N}-2}\gamma\beta}{r + \alpha_{t+\mathbf{N}-3} C_{t,\mathbf{N}-2}\beta^2} x$$

and $U_{t,\mathbf{N}-3}(x) = C_{t,\mathbf{N}-3}x^2 + D_{t,\mathbf{N}-3}$, where

$$C_{t,\mathbf{N}-3} = \frac{qr + \alpha_{t+\mathbf{N}-3} C_{t,\mathbf{N}-2}(q\beta^2 + r\gamma^2)}{r + \alpha_{t+\mathbf{N}-3} C_{t,\mathbf{N}-2}\beta^2}$$

and $D_{t,\mathbf{N}-3} = \alpha_{t+\mathbf{N}-3}(C_{t,\mathbf{N}-2}\sigma^2 + D_{t,\mathbf{n}-2})$.
Making $k = 1$, it is obtained that

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

and $U_0(x) = C_{t,1}x^2 + D_{t,1}$, where

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

and $D_{t,1} = \alpha_{t+1}(C_{t,2}\sigma^2 + D_{t,2})$.
Finally, with $k = 0$, it is obtained the first optimal decision for the problem starts at stage t, that is

$$f_{t,0}(x) = \frac{-\alpha_t C_{t,1}\gamma\beta}{r + \alpha_t C_{t,1}\beta^2} x,$$

Hence

$$\widehat{f}_t(x) = \frac{-\alpha_t C_{t,1}\gamma\beta}{r + \alpha_t C_{t,1}\beta^2} x = \lambda_t x, \quad t = 0, 1, 2, \ldots, \quad x \in X.$$

To obtaining the performance function of the rolling horizon policy, applying the dynamic of the system which is $x_{t+1} = \gamma x_t + \beta a_t + \xi_t$ with $a_t = \widehat{f}_t(x) = \lambda_t x$ and $x_0 = x$, it is obtained that

$$\begin{aligned}
x_1 &= (\gamma+\beta\lambda_0)x + \xi_0 \\
x_2 &= (\gamma+\beta\lambda_0)(\gamma+\beta\lambda_1)x + (\gamma+\beta\lambda_1)\xi_0 + \xi_1 \\
x_3 &= (\gamma+\beta\lambda_0)(\gamma+\beta\lambda_1)(\gamma+\beta\lambda_2)x + (\gamma+\beta\lambda_1)(\gamma+\beta\lambda_2)\xi_0 + (\gamma+\beta\lambda_2)\xi_1 + \xi_2 \\
&\vdots \\
x_t &= \prod_{i=0}^{t-1}(\gamma+\beta\lambda_i)x + \sum_{i=0}^{t-2}\prod_{j=i+1}^{t-1}(\gamma+\beta\lambda_j)\xi_i + \xi_{t-1}.
\end{aligned}$$

Then,

$$\begin{aligned}
J^{\tau}(\widehat{\pi}, x) &= E_x^{\widehat{\pi}}\left[\sum_{t=0}^{\infty} P_t c(x_t, a_t)\right] \\
&= E_x^{\widehat{\pi}}\left[\sum_{t=0}^{\infty} P_t (q + r\lambda_t^2) x_t^2\right] \\
&= P_0(q+r\lambda_0^2)x^2 + P_1(q+r\lambda_1^2)E_x^{\widehat{\pi}}\left[((\gamma+\beta\lambda_0)x+\xi_0)^2\right] \\
&\quad + E_x^{\widehat{\pi}}\left[\sum_{t=2}^{\infty} P_t(q+r\lambda_t^2)\left(\prod_{i=0}^{t-1}(\gamma+\beta\lambda_i)x + \sum_{i=0}^{t-2}\prod_{j=i+1}^{t-1}(\gamma+\beta\lambda_j)\xi_i + \xi_{t-1}\right)^2\right] \\
&= P_0(q+r\lambda_0^2)x^2 + P_1(q+r\lambda_1^2)\left((\gamma+\beta\lambda_0)^2x^2 + \sigma^2\right) \\
&\quad + \sum_{t=2}^{\infty} P_t(q+r\lambda_t^2)\left(\prod_{i=0}^{t-1}(\gamma+\beta\lambda_i)^2x^2 + E_x^{\widehat{\pi}}\left[\left(\sum_{i=0}^{t-2}\prod_{j=i+1}^{t-1}(\gamma+\beta\lambda_j)\xi_i\right)^2\right] + \sigma^2\right) \\
&= P_0(q+r\lambda_0^2)x^2 + P_1(q+r\lambda_1^2)\left((\gamma+\beta\lambda_0)^2x^2 + \sigma^2\right) \\
&\quad + \sum_{t=2}^{\infty} P_t(q+r\lambda_t^2)\left(\prod_{i=0}^{t-1}(\gamma+\beta\lambda_i)^2x^2 + \sum_{i=0}^{t-2}\prod_{j=i+1}^{t-1}(\gamma+\beta\lambda_j)\sigma^2 + \sigma^2\right) \\
&= \left(P_0(q+r\lambda_0^2) + P_1(q+r\lambda_1^2)(\gamma+\beta\lambda_0)^2 + \sum_{t=2}^{\infty} P_t(q+r\lambda_t^2)\prod_{i=0}^{t-1}(r+\beta\lambda_i)^2\right)x^2 \\
&\quad + \sigma^2 P_1(q+r\lambda_1^2) + \sigma^2\sum_{t=2}^{\infty} P_t(q+r\lambda_t^2)\left(1 + \sum_{i=0}^{t-2}\prod_{j=i+1}^{t-1}(\gamma+\beta\lambda_j)^2\right) \\
&= K_1x^2 + K_2,
\end{aligned}$$

concluding of this form the proof. ■

The formulas given in Lemma 4.3. have been programmed in order to obtain numerical results for the LQ control problem with a random horizon. Without loss of generality, the following values for the parameters of the model have been considered: $\gamma = \beta = q = r = \sigma^2 = 1$. Furthermore, it is assumed that the horizon τ has a Logarithmic distribution, that is $P(\tau = t) = -\frac{(1-p)^{t+1}}{(t+1)\ln p}$, $t = 0, 1, 2, \ldots$, with $p = 0.8$.

In Table 1, the approximate values of the constants given in (6) and (7) are showed evaluating 900 stages of the rolling horizon policies with various values of $\mathbf{N}$. Note that in [1], it is proved that when $\mathbf{N} \to \infty$ the error of the approximation tends to zero. As is observed in the numerical results, there is no big difference in the performance of policies and even rolling horizon with small lengths a good performance is obtained.

N	K_1	K_2
100	1.07820310519224237224	0.09445482558131939907
50	1.07820310519224237224	0.09445482558131939907
20	1.07820310519224237224	0.09445482558131939907
15	1.07820310519224237224	0.09445482558131945282
10	1.07820310519224237267	0.09445482558535835057
5	1.07820310944105568810	0.09445512530972944205
3	1.07825877384652960600	0.09448857885511955891

Table 1: Approximate coefficients of the performance function of rolling horizon policies with different lengths.

In Table 2, with $\mathbf{N} = 20$, some values of the constants λ_t given in (5) are presented. Note that there seems to be a convergence of these values.

t	970	971	972	973	974	975
λ_t	-.14662424	-.14662439	-.14662453	-.14662468	-.14662483	-.14662498
t	976	977	978	979	980	981
λ_t	-.14662512	-.14662527	-.14662542	-.14662556	-.14662571	-.14662586

Table 2: Coefficients of the selectors in the rolling horizon policy.

In [2], for the logarithmic distribution with parameter p it is proved that $\lim_{t\to\infty} \alpha_t = 1-p$ and $\alpha_t \geq 1-p$, where α_t is defined in (1). Also, in [2], under this condition it is proved that there exist a relation between "the optimal control problem with a random horizon" and "a discounted problem with $\alpha = \lim_{t\to\infty} \alpha_t$ as the discount factor".

For the LQ control problem with a random horizon the associated problem is the discounted problem with $\alpha = 1 - p$. The optimal policy for this problem is the following (see [3], p. 70):

$$\pi = (f, f, f, \ldots)$$

with the selector f given by

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

$x \in X$, where

$$C = \frac{-\left(r - \alpha(r\gamma^2 + q\beta^2)\right) + \sqrt{\left(r - \alpha(r\gamma^2 + q\beta^2)\right)^2 + 4\alpha q r \beta}}{2\alpha\beta^2}$$

In the numerical case, the evaluation of the coefficient in the selector f is obtained, obtaining that $\lambda = -1.46768836$, so it can be stated that

$$\lim_{t\to\infty} \widehat{f}_t = f.$$

The proof is not trivial, but in sufficient numerical cases can be seen something similar. Such convergence would be very useful to the impossibility of compute $\widehat{f}_t$ for t large.

5. CONCLUSION

In this paper the existence of the optimal solution is proved for a Linear-Quadratic control problem in which the performance function is the expected total cost with a random horizon. The case of random horizon with infinite support is considered. To guarantee the existence of the optimal solution were used the theoretical results provided in [2]. Given the difficulty to obtain the exact optimal solution, in this paper a policy of rolling horizon is obtained to approximate the solution. A deterministic Markov policy through recursive equations is obtained which are programmed to obtain numerical results. It is shown that use small values of the length in the rolling horizon is sufficient to calculate the selectors. Moreover, a convergence of these selectors is observed in various numerical cases which allow to change the policy of rolling horizon by a stationary policy.

6. ACKNOWLEDGMENTS

El autor agradece a la Sociedad Matemática Mexicana y Fundación Sofía Kovalévskaia el apoyo otorgado para la presentación de este artículo.

RECEIVED: JULY, 2014.
REVISED: DECEMBER, 2014.

REFERENCES

[1] ALDEN, J. M.and SMITH, R. L.(1992): Rolling Horizon Procedures in Nonhomogeneous Markov Decision Process, **Operations Research**, 40, 183-194.

[2] CRUZ-SUÁREZ, H., ILHUICATZI-ROLDÁN, R., MONTES-DE-OCA, R.(2012): Markov Decision Processes on Borel Spaces with Total Cost and Random Horizon, **Journal of Optimization Theory and Applications**, DOI 10.1007/s10957-012-0262-8

[3] HERNÁNDEZ-LERMA, O., LASSERRE, J. B. (1996): **Discrete-Time Markov Control Processes: Basic Optimality Criteria**. Springer-Verlag, New York.

[4] KOZLOWSKI, E.(2010): The Linear-Quadratic Stochastic Optimal Control Problem with Random Horizon at the Finite Number of Infinitesimal Events, **Annales UMCS Informatica**, 1, 103-115.

[5] MORI, K. (2012): Approximate Linear Quadratic Regulator Problem and its Aplications to Optimal Control in Discrete-Time LTI Systems, **Intelligente Control and Innovative, Lecture Notes in Electrical Engineering**, 110, 53-64.

[6] PUTERMAN, M. L. (1994): **Markov Decision Process: Discrete Stochastic Dynamic Programming**, John Wiley & Sons.