Una cota de convergencia para los minimizadores en un modelo lineal cuadrático

R. Israel Ortega-Gutiérrez
Facultad de Ciencias Físico Matemáticas,
Benemérita Universidad Autónoma de Puebla,
Ave. San Claudio y Río Verde, Col. San Manuel, CU,
Puebla, Pue. 72570, México

Hugo Cruz-Suárez
Facultad de Ciencias Físico Matemáticas,
Benemérita Universidad Autónoma de Puebla,
Ave. San Claudio y Río Verde, Col. San Manuel, CU,
Puebla, Pue. 72570, México

J. Daniel Velázquez-Martínez
Facultad de Ciencias Químicas-Ingeniería Industrial y Administración,
Universidad Autónoma de Nuevo León,
Av. Universidad, sin número, Ciudad Universitaria,
San Nicolás de Los Garza, Nuevo León 66455, México.

RESUMEN

En este artículo se presenta un modelo lineal cuadrático estocástico en tiempo discreto. Se considera como criterio de rendimiento al costo total descontado esperado con horizonte infinito. Para este problema se da la primera y segunda derivada del lado derecho de la ecuación de programación dinámica. Con estas derivadas se desarrolla la serie de Taylor para el lado derecho de la ecuación de programación dinámica alrededor de la política óptima. Con dicha serie de Taylor se relacionan las funciones de valor con los minimizadores provenientes del algoritmo de iteración de valores y con ello se determina una cota de convergencia.

Palabras Claves: Procesos de Control de Markov Descontados, Programación Dinámica, Modelos de Control de Markov, Modelos Lineales Cuadráticos.

1. INTRODUCCIÓN

En este artículo se presenta el modelo Lineal Cuadrático (LQ, por sus siglas en inglés), el cual se modela usando la teoría de Procesos de Control de Markov (PCM), en particular se observa dicho proceso en tiempo discreto y con horizonte infinito (véase [4]).

El modelo LQ ha mostrado ser una herramienta eficaz para la solución de problemas de control en diversas áreas aplicadas, por ejemplo, puede consultar referencias en Ingeniería ([6] y [9]), Inteligencia Artificial ([7] y [12]), Economía ([10] y [13]), entre otras. El contexto general de aplicación del modelo LQ es implementado en problemas dinámicos que siguen un comportamiento lineal en sus transiciones y el controlador tiene la necesidad de que dicho movimiento se mantenga cercano a un objetivo específico, de esta manera se propone una función de penalidad a movimientos alejados del objetivo fijo, mediante una función de costo. Por tal motivo se busca determinar una estrategia que minimice los costos de penalización, en consecuencia, se encontrará una estrategia que mantenga cercano al sistema estocástico del objetivo dado. En este trabajo, sin pérdida de generalidad, el objetivo fijo se centra en el origen de un eje cartesiano de los números reales.

La metodología para analizar el problema descrito es mediante la técnica de programación dinámica estocástica ([5]). De esta manera en el artículo, en una primera etapa, validamos Programación Dinámica verificando las suposiciones clásicas de la literatura referentes a continuidad y compacidad en las componentes del modelo de control. Posteriormente aplicando la técnica de iteración de valores se determinan aproximaciones a la función de valor óptimo, la cual corresponde al costo mínimo incurrido en las transiciones del sistema aplicando la estrategia óptima. La metodología descrita hasta este momento es canónica y ha sido aplicada para el modelo LQ en diferentes trabajos, por ejemplo: [3], [4] y [14]. Por otro lado, la teoría de programación dinámica garantiza la existencia de una estrategia de operación estacionaria en el tiempo, considerando el problema a horizonte infinito, sin embargo, dicha política no es determinada de manera directa, en muchos casos se procede a solucionar la ecuación de programación dinámica, o bien, en casos muy particulares, determinar aproximaciones mediante los minimizadores de iteración de valores ([2]), la cual ha sido aplicada al modelo LQ en [1]. La aportación de este trabajo va a encaminada a proponer una tasa de convergencia para los minimizadores de iteración de valores con respecto a la política óptima. Dicha propuesta permite determinar aproximaciones numéricas de la política óptima vía los minimizadores de iteración de valores y determinar el error de aproximación vía la cota de convergencia, la cual tiene un orden de convergencia geométrico (ver Teorema 13 en este artículo). El procedimiento que seguimos para determinar dicha cota de convergencia se encuentra sustentada en la forma funcional cuadrática del lado derecho de la ecuación de programación dinámica y su representación en series de potencias.

El manuscrito se encuentra organizado de la siguiente manera, en la Sección 2 se presenta la teoría básica de Procesos de Control de Markov, posteriormente, en la Sección 3 se introduce el modelo lineal cuadrático. En la Sección 4 se presenta el teorema principal del trabajo de investigación junto con algunos resultados clásicos referentes al modelo LQ. Finalmente, en la Sección 6 se proporcionan algunos resultados numéricos.

2. PROCESOS DE CONTROL DE MARKOV DESCONTADOS

Primero se dará brevemente la definición de un proceso de control de Markov (para más detalles y terminología véase [5]).

Un Modelo de Control de Markov (MCM), estacionario, a tiempo discreto (véase [5]), consiste de una quíntupla:

$$\left(X, A, \left\{A(x) \middle| x \in X\right\}, Q, c\right), \tag{1.1}$$

donde

a) X es un espacio de Borel no vacío (i.e., un subconjunto de Borel de un espacio métrico completo y separable), llamado espacio de estados. A los elementos de X se les llama estados;

b) A, es el conjunto de acciones (o controles), el cual también es un espacio de Borel no vacío;

c) $A(x) \subset A$ es una familia de subconjuntos medibles no vacíos $\mathrm{A(x)}$ de A, donde $\mathrm{A(x)}$ denota el conjunto de acciones admisibles para el estado $x \in X$ y con la propiedad de que el conjunto
$$\mathbb{K} := \left\{(x,a) \middle| x \in X, a \in A(x)\right\}, \tag{1.2}$$
de parejas de estado-acción admisible, es un subconjunto medible de $\mathrm{X} \times \mathrm{A}$;

d) Q es un kérnel estocástico (véase [5]) definido en X dado $\mathbb{K}$;

e) $\mathrm{c} : \mathbb{K} \to \mathbb{R}$ es una función medible llamada función de costo en un paso.

El modelo de control (1.1) representa un sistema de control estocástico el cual es observado de manera periódica en los tiempos $\mathrm{t} = 0,1,2,\ldots$ Sean $\mathrm{x_t = x \in X}$ y $\mathrm{a_t = a \in A(x)}$, el estado del sistema y la acción (o control) aplicada en el tiempo t, entonces ocurre lo siguiente:

a) se paga un costo $\mathrm{c(x,a)}$ y,

b) el sistema se moverá a un nuevo estado $\mathrm{x_{t+1}}$ con la distribución de probabilidad $\mathrm{Q}\left(\cdot \middle| \mathrm{x,a}\right)$, i.e.,
$$\mathrm{Q}\left(\mathrm{B}\middle|\mathrm{x,a}\right) = \Pr\left(\mathrm{x_{t+1}} \in \mathrm{B}\middle|\mathrm{x_t = x, a_t = a}\right), \tag{1.3}$$
donde B es un elemento de la $\sigma-\text{álgebra}$ de Borel de X. Una vez que el sistema pasa a un nuevo estado, se elige nuevamente una acción y el proceso anterior se repite.

Sea Π el conjunto de todas las políticas admisibles (posiblemente aleatorizada o dependiente de la historia). En particular, una **política estacionaria** es definida como una función medible $f : X \to A$ tal que $f(x) \in A(x)$, para todo $x \in X$. El conjunto de todas las políticas estacionarias será denotado por $\mathbb{F}$. Ahora, para cada $\pi \in \Pi$ y un estado inicial $x \in X$, sea

$$V(\pi, x) = E_x^{\pi}\left[\sum_{\mathrm{t}=0}^{\infty} \alpha^{\mathrm{t}} \mathrm{c}(\mathrm{x_t}, \mathrm{a_t})\right], \tag{1.4}$$

el **costo total esperado descontado** cuando se aplica la política π, dado un estado inicial x. La constante $\alpha \in (0,1)$ es el **factor de descuento,** el cual es fijo. Las sucesiones de estados y controles se denotan por $\{x_t\}$ y $\{a_t\}$, respectivamente. También E_x^{π} denota la esperanza con respecto a la medida de probabilidad P_x^{π} la cual está definida en el espacio canónico $(\Omega, \mathfrak{F})$ y es inducida por el Teorema de Ionesco Tulcea [5].

Observación 1:
En algunos casos la ley de transición Q es inducida por un sistema de ecuaciones de la forma $x_{t+1} = F(x_t, a_t, \xi_t)$, con $t = 0,1,\cdots$, donde $F : X \times A \times S \to X$ es una función medible y $\{\xi_t\}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas (i.i.d) con valores en $S \subseteq \mathbb{R}$, y densidad común Δ.

Definición 2:
Una política π^* es **óptima** si $V(\pi^*, x) = \upsilon^*(x)$ para todo $x \in X$, donde

$$\upsilon^*(x) := \inf_{\pi \in \Pi} V(\pi, x), \quad x \in X. \tag{1.5}$$

υ^* es llamada la función de valor óptimo.

Suposición 3:

a) c es una función continua en $\mathbb{K}$.

b) La ley de transición Q es fuertemente continua, es decir,
$$\theta(x,a) = \int_X u(y) Q(dy \mid x,a), \ (x,a) \in \mathbb{K}, \tag{1.6}$$
es una función continua y acotada en $\mathbb{K}$, para cada función medible y acotada u en X.

c) Supóngase que para cada $x \in X$, los conjuntos $A(x)$ son compactos y la multifunción $x \mapsto A(x)$ es continua.

d) Supóngase que existe una constante $\hat{k}$, δ con $1 < \delta < \dfrac{1}{\alpha}$ y $\varpi : X \to [1,\infty)$ una función medible y continua tal que
$$\sup_{a \in A} \left|c(x,a)\right| \leq \hat{k}\varpi(x), \tag{1.7}$$
y
$$\sup_{a \in A} \int \varpi(y) Q(dy \mid x,a) \leq \delta \varpi(x), \tag{1.8}$$
para todo $x \in X$.

Definición 4:
Las funciones de iteración de valores son definidas como

$$V_n(x) = \min_{a\in A(x)} \left\{ c(x,a) + \alpha \int_X V_{n-1}(y) Q(dy \mid x,a) \right\}, \qquad (1.9)$$

para todo $x \in X$ y $n = 1,2,\ldots,$ con $V_0(\cdot) = 0.$

La demostración del siguiente lema puede ser consultada en [5].

Lema 5:
Asuma que la Suposición 3 se cumple.
Entonces

a) La función de valor óptimo υ^* es una solución de

$$\upsilon^*(x) = \min_{a\in A(x)} \left[c(x,a) + \alpha \int_X \upsilon^*(y) Q(dy \middle| x,a) \right] \qquad (1.10)$$

para cada $x \in X$ y si u es otra solución de la ecuación, entonces $u(\cdot) \ge \upsilon^*(\cdot).$

b) Existe un selector $f^* \in \mathbb{F}$ tal que en (1.10) se alcanza el mínimo, es decir,

$$\upsilon^*(x) = c(x, f^*(x)) + \alpha \int_X \upsilon^*(y) Q(dy \middle| x, f^*(x)) \qquad (1.11)$$

y f^* es óptima.

c) Para toda $x \in X$, $V_n(x) \uparrow \upsilon^*(x)$ con V_n definida como en (1.9).
Además, para cada $n \in \mathbb{N}$ existe $f_n \in \mathbb{F}$ tal que

$$\begin{aligned} &\min_{a\in A(x)} \left\{ c(x,a) + \alpha \int_X V_{n-1}(y) Q(dy \mid x,a) \right\} \\ &= c(x, f_n(x)) + \alpha \int_X V_{n-1}(y) Q(dy \mid x, f_n(x)), \end{aligned} \qquad (1.12)$$

para todo $x \in X.$ f_n son los minimizadores provenientes del método de iteración de valores.

3. MODELO LINEAL CUADRÁTICO (LQ)

Sea $X = A = A(x) = \mathbb{R}$ y $x \in X.$ Suponga que la transición del sistema es dominada por la siguiente ecuación e diferencias:

$$x_{t+1} = x_t + a_t + \xi_t, \ t = 0,1,\ldots, \text{ y } x_0 = x \in X \ , \qquad (1.13)$$

donde las variables aleatorias ξ_t, para cada $t = 0,1,\ldots$, son independientes e idénticamente distribuidas con valores en $S = \mathbb{R}$. Además, supóngase que ξ_0 un elemento genérico tiene densidad continua Δ, con media 0 y varianza σ^2, es decir,

$$E(\xi_0) = 0, \qquad (1.14)$$

y

$$E(\xi_0^2) = \sigma^2 < \infty. \qquad (1.15)$$

En este caso, la ley de transición es determinada para $x \in X$, $a \in A(x)$ y $y \in \mathbb{R}$ como:

$$\begin{aligned} Q\left(x_1 \in (-\infty, y] \middle| x_0 = x, a_0 = a\right) &= \Pr\left(x + a + \xi_0 \le y\right) \\ &= \int I_{(-\infty, y-x-a)}(s)\Delta(s)ds \end{aligned}$$

(1.16)

Además, supóngase que lo costos de penalización son modelados por la siguiente función:

$$c(x,a) = x^2 + a^2, \ (x,a) \in \mathbb{K}. \qquad (1.17)$$

Lema 6:
Las componentes del modelo LQ satisfacen la Suposición 3.

Demostración
La prueba de (a) y (b) es dada en [5].
(c) En este caso, primero se probará que la multifunción es inferiormente semicontinua (i.s.c.). Para ello, sea $x_n \to |x| \in X$ y $a \in A(x) = \left[-|x|, |x|\right].$ Definiendo $a_n = -x_n$ o $a_n = x_n$, se obtiene que $a_n \to -x$ o $a_n \to x$, cuando $n \to \infty.$ Por otro lado, si $a \in \left(-|x|, |x|\right)$, entonces $a = \beta(-|x|) + (1-\beta)|x|$, para algún $\beta \in (0,1).$ Tomando $a_n = \beta(-x_n) + (1-\beta)x_n$, se obtiene que $a_n \to a$ cuando $n \to \infty.$ Por lo tanto, la multifunción $x \mapsto A(x)$ es i.s.c.

Ahora, se probará que la multifunción es superiormente semicontinua (s.s.c.), dado $x_n \to x \in X$ y $a_n \in A(x_n) = \left[-x_n, x_n\right].$ Entonces $-x_n \le a_n \le x_n$, tomando el límite cuando $n \to \infty$ se obtiene que $-x \le \liminf_{n\to\infty} a_n \le \limsup_{n\to\infty} a_n \le x.$ Lo anterior implica que la multifunción $x \mapsto A(x)$ es s.s.c. (véase Teorema 3.17 en [11]).
Por lo tanto, la multifunción $x \mapsto A(x)$ es continua.
Para probar (d), basta con tomar $\hat{k} = 2,$ $\delta = 1$ y $\varpi(x) = x^2 + \sigma^2.$

Lema 7:
La política óptima $f^*(x) \in \left(-|x|, |x|\right)$, para todo $x \in X.$ En efecto, note que si $\hat{f}(x) = 0$ entonces

$$V(\hat{f}, x) = \frac{x}{1-\alpha} + \frac{\sigma^2\alpha}{(1-\alpha)^2}. \qquad (1.18)$$

Por otro lado, si $\overline{f}(x) = -x$ entonces

$$V(\overline{f}, x) = \frac{\sigma^2}{1-\alpha}, \qquad (1.19)$$

y si $\tilde{f}(x) = x$ entonces

$$V(\tilde{f}, x) = x^2 \sum_{t=0}^{\infty} (t+1)\alpha^t + \frac{\sigma^2\alpha}{1-\alpha}. \qquad (1.20)$$

Observe que, en cualquier caso

$$V(\hat{f},x) \le V(\overline{f},x) \qquad (1.21)$$

y

$$V(\hat{f},x) \le V(\tilde{\mathrm{f}},x). \qquad (1.22)$$

Entonces, $\overline{f}(x) = -x$ y $\tilde{f}(x) = x$ no son óptimas.

Por lo tanto, $f^*(x) \in \left(-|x|,|x|\right)$ para todo $x \in X$.

4. ITERACIÓN DE VALORES PARA EL MODELO LQ

En lo subsecuente la siguiente notación será considerada.

Notación 8:

Defina las siguientes funciones

$$G(x,a) := x^2 + a^2 + \alpha E\left[\upsilon^*\left(x+a+\xi\right)\right], \qquad (1.23)$$

y

$$G_n(x,a) := x^2 + a^2 + \alpha E\left[V_{n-1}\left(x+a+\xi\right)\right], \qquad (1.24)$$

para cada $(x,a) \in \mathbb{K}$. A G se le llama el lado derecho de la ecuación de programación dinámica.

Teorema 9:

Para el modelo LQ se cumple que

$$V_n(x) = K_n x^2 + C_n \sigma^2 \qquad (1.25)$$

y

$$f_n(\mathrm{x}) = -\frac{\alpha K_{n-1} x}{1+\alpha K_{n-1}}. \qquad (1.26)$$

Demostración

Usando la Ec. (1.9), el Lema 5 e iniciando con $V_0(x) = 0$, se obtiene que

$$\begin{aligned} V_1(x) &= \min_{a\in A(x)} \left\{x^2 + a^2 + \alpha E\left[V_0(x+a+\xi\right]\right\} \\ &= \min_{a\in A(x)} \left\{x^2 + a^2\right\} . \end{aligned} \qquad (1.27)$$

Derivando con respecto a a la parte que esta entre llaves de la Ec. (1.27), se obtiene que $f_1(x) = 0$. Sustituyendo $f_1(x) = 0$ en la Ec. (1.27), se llega a que

$$V_1(x) = x^2 . \qquad (1.28)$$

De manera similar,

$$\begin{aligned} V_2(x) &= \min_{a\in A(x)} \left\{x^2 + a^2 + \alpha E\left[V_1(x+a+\xi)\right]\right\} \\ &= \min_{a\in A(x)} \left\{x^2 + a^2 + \alpha E\left[(x+a+\xi)^2\right]\right\} \\ &= \min_{a\in A(x)} \left\{x^2 + a^2 + \alpha E\left[(x+a)^2 + 2(x+a)\xi + \xi^2\right]\right\} \\ &= \min_{a\in A(x)} \left\{x^2 + a^2 + \alpha(x+a)^2 + \alpha\sigma^2\right\} . \end{aligned}$$

$$(1.29)$$

Entonces

$$f_2(x) = -\frac{\alpha x}{1+\alpha}. \qquad (1.30)$$

De esta forma sustituyendo en la Ec. (1.29), se llega a que

$$V_2(x) = \frac{2\alpha+1}{1+\alpha} x^2 + \alpha\sigma^2. \qquad (1.31)$$

En general, supóngase que

$$V_{n-1}(x) = K_{n-1} x^2 + C_{n-1}\sigma^2 \qquad (1.32)$$

De donde se sigue que

$$\begin{aligned} V_n(x) &= \min_{a\in A(x)} \left\{x^2 + a^2 + \alpha E\left[V_{n-1}(x+a+\xi)\right]\right\} \\ &= \min_{a\in A(x)} \left\{x^2 + a^2 + \alpha E\left[\mathrm{K}_{n-1}(x+a+\xi)^2 + C_{n-1}\sigma^2\right]\right\} \\ &= \min_{a\in A(x)} \left\{x^2 + a^2 + \alpha \mathrm{K}_{n-1}(x+a)^2 + \alpha \mathrm{K}_{n-1}\sigma^2 + \alpha C_{n-1}\sigma^2\right\} . \end{aligned}$$

$$(1.33)$$

Así,

$$f_n(\mathrm{x}) = -\frac{\alpha K_{n-1} x}{1+\alpha K_{n-1}}. \qquad (1.34)$$

Observe que para todo $\mathrm{n} \ge 1$, se cumple que $\mathrm{f_n(x)} \in \mathrm{int(A(x))}$ (donde, int(D) denota el interior del conjunto D).

Sustituyendo en la Ec. (1.33), se llega a que

$$\begin{aligned} V_n(x) &= x^2 + \frac{\alpha^2 \mathrm{K}_{n-1}^2 x^2}{\left(1+\alpha K_{n-1}\right)^2} \\ &\quad + \alpha \mathrm{K}_{n-1}\left(x - \frac{\alpha \mathrm{K}_{n-1} x}{1+\alpha K_{n-1}}\right)^2 + \alpha\sigma^2\left(\mathrm{K}_{n-1} + C_{n-1}\right) \\ &= \frac{1+2\alpha \mathrm{K}_{n-1}}{1+\alpha K_{n-1}} x^2 + \alpha\sigma^2\left(\mathrm{K}_{n-1} + C_{n-1}\right). \end{aligned}$$

$$(1.35)$$

Por tanto,

$$K_n = \frac{1+2\alpha K_{n-1}}{1+\alpha K_{n-1}} \quad \text{y} \quad \mathrm{K_0}=0. \qquad (1.36)$$

Además,

$$C_n = \sum_{i=1}^{n} K_i \quad \text{y} \quad \mathrm{C_0}=0. \qquad (1.37)$$

Observación 10:

a) Por el Lema 5 y el Lema 6 existe f^* y se tiene que $\mathrm{V_n(x)} \to \upsilon^*(\mathrm{x})$ para cada $\mathrm{x} \in \mathrm{X}$. Siguiendo una prueba similar a la prueba del Lemma 4 en [4], p. 312, se concluye que $K_n \to K$ y $C_n \to C$. De estas relaciones se deduce que

$$K = \frac{1+2\alpha K}{1+\alpha K} \qquad (1.38)$$

Resolviendo la única solución debe ser

$$K = \frac{2\alpha - 1 + \sqrt{4\alpha^2 + 1}}{2\alpha} \qquad (1.39)$$

ya que la otra solución no está en el conjunto de acciones admisibles.

b) Note que $G(x,\cdot)$ es estrictamente convexa, para cada $x \in X$. En consecuencia se tiene la unicidad de la política óptima f^*, véase [3].

5. TASA DE CONVERGENCIA EN EL MODELO LQ

En esta sección, se obtiene una tasa de convergencia explícita para los minimizadores del enfoque de iteración de valores, lo cual es una consecuencia de la diferenciabilidad del lado derecho de la ecuación de programación dinámica. Recuérdese que las políticas provenientes del método de iteración de valores se denotan por $f_n(x) \in arg\min\{G_n(x,a)\}$.

Lema 11:
Bajo la Suposición 3. Se cumplen las siguientes desigualdades para todo $n \geq 1$

a)

$$\left|V_n(x) - \upsilon^*(x)\right| \leq \frac{2\left(x^2 + \sigma^2\right)\alpha^n}{1-\alpha}, \qquad (1.40)$$

b)

$$|G(x, f_n(x)) - G(x, f^*(x))| \leq \frac{4\left(x^2 + \sigma^2\right)\alpha^n}{1-\alpha}, \qquad (1.41)$$

para cada $a \in A(x)$ y $x \in X$, donde $f_n(x)$ son las políticas provenientes del método de iteración de valores.

Demostración:

a) Es un resultado inmediato del Teorema 8.3.6 en [4].

b) Sea $x \in X$ y $n \geq 1$ fijo. Entonces,

$$\begin{aligned}
\left|G(x, f_n(x)) - G(x, f^*(x))\right| &= \left|\begin{matrix} c(x, f_n(x)) + \alpha\int_X \upsilon^*(y)Q(dy \mid x, f_n(x)) \\ +V_n(x) - V_n(x) - \upsilon^*(x)\end{matrix}\right| \\
&\leq \left|\alpha\int_X \left(\upsilon^*(y) - V_{n-1}(y)\right)Q(dy \mid x, f_n(x)) + V_n(x) - \upsilon^*(x)\right| \\
&\leq \alpha\int_X \left|\upsilon^*(y) - V_{n-1}(y)\right|Q(dy \mid x, f_n(x)) + \left|V_n(x) - \upsilon^*(x)\right| \\
&\leq \frac{2\alpha^n}{1-\alpha}\int_X \varpi(y)Q(dy \mid x, f_n(x)) + \frac{2\varpi(x)\alpha^n}{1-\alpha} \\
&\leq \frac{2\alpha^n}{1-\alpha}\varpi(x) + \frac{2\varpi(x)\alpha^n}{1-\alpha} \\
&= \frac{4\varpi(x)\alpha^n}{1-\alpha} = \frac{4\left(x^2 + \sigma^2\right)\alpha^n}{1-\alpha},
\end{aligned} \qquad (1.42)$$

Como x es arbitrario el resultado se sigue para todo $x \in X$.

Lema 12:
Asuma que la Suposición 3 se cumple. Entonces, la primera y segunda derivada de G, denotada por G_a y G_{aa}, respectivamente están dadas por

$$G_a(x,a) = 2\left(a + \alpha K\left(x + a\right)\right), \qquad (1.43)$$

$$G_{aa}(x,a) = 2\left(1 + \alpha K\right), \qquad (1.44)$$

Para cada $a \in int(A(x))$ y $x \in X$. En particular, como $f^*(x) \in int(A(x))$ (véase Lema 7)

$$G_{aa}(x, f^*(x)) = 2\left(1 + \alpha K\right). \qquad (1.45)$$

Demostración:
Sea $x \in X$ fijo. De la Ec. (1.23) y usando la siguiente forma funcional para $\upsilon^*(x) = Kx^2 + C$, se tiene que

$$\begin{aligned}
G(x,a) &= x^2 + a^2 + \alpha E\left[\upsilon^*\left(x + a + \xi\right)\right] \\
&= x^2 + a^2 + \alpha E\left[K\left(x + a + \xi\right)^2 + C\sigma^2\right] \\
&= x^2 + a^2 + \alpha K\left(x + a\right)^2 + \alpha\sigma^2 + \alpha C\sigma^2.
\end{aligned} \qquad (1.46)$$

Entonces

$$G_a(x,\mathrm{a}) = 2a + 2\alpha K\left(x + a\right). \qquad (1.47)$$

Derivando nuevamente la Ec (1.47), se llega a que

$$G_{aa}(x,a) = 2\left(1 + \alpha K\right), \qquad (1.48)$$

Como x es arbitrario el resultado se sigue para todo $x \in X$.

Teorema 13:
En el modelo LQ se satisface para cada $x \in X$ y $n \geq 1$ la siguiente relación:

$$\left|f_n(x) - f^*(x)\right| \leq 2\sqrt{\frac{\left(x^2 + \sigma^2\right)}{\left(1 + \alpha K\right)\left(1 - \alpha\right)}}\alpha^{n/2}. \qquad (1.49)$$

Demostración:

Sea $x \in X$ fijo. Ahora, la serie de Taylor para $\mathrm{G}(x, f_n(x))$ en $f^*(x)$, donde f_n son los minimizadores provenientes del método de iteración de valores. Usando que $G_a(x, f^*(x)) = 0$, debido a que f^* es óptima, por el Teorema 3 p. 196, véase [8] se obtiene que

$$\begin{aligned}
\mathrm{G}(x, f_n(x)) &= G(x, f^*(x)) + \frac{G_a(x, f^*(x))}{1!}(f_n(x) - f^*(x)) \\
&+ \frac{G_{aa}(x, f^*(x))}{2!}(f_n(x) - f^*(x))^2,
\end{aligned} \qquad (1.50)$$

debido a que las derivadas de órdenes superiores son cero. Aplicando el Lema 12 a la Ec. (1.50) se llega a

$$\begin{aligned}\mathrm{G}(x, f_n(x)) &= G(x, f^*(x)) + \frac{1}{2}\left(2\left(1+\alpha K\right)\right)(f_n(x) - f^*(x))^2 \\ &= G(x, f^*(x)) + \left(1+\alpha K\right)(f_n(x) - f^*(x))^2.\end{aligned} \tag{1.51}$$

Entonces, de la Ec. (1.41) se obtiene que

$$\begin{aligned}\left(1+\alpha K\right)(f_n(x) - f^*(x))^2 &= G(x, f_n(x)) - G(x, f^*(x)) \\ &\leq \frac{4\left(x^2+\sigma^2\right)\alpha^n}{1-\alpha}.\end{aligned} \tag{1.52}$$

Así, para todo $n \geq 1$

$$\begin{aligned}(f_n(x) - f^*(x))^2 &\leq G(x, f_n(x)) - G(x, f^*(x)) \\ &\leq \frac{4\alpha^n\left(x^2+\sigma^2\right)}{\left(1+\alpha K\right)\left(1-\alpha\right)}.\end{aligned} \tag{1.53}$$

Claramente de la Ec. (1.53) se obtiene que

$$\left| f_n(x) - f^*(x) \right| \leq 2\sqrt{\frac{\alpha^n\left(x^2+\sigma^2\right)}{\left(1+\alpha K\right)\left(1-\alpha\right)}}. \tag{1.54}$$

Observación 14:

Nótese que en la Ec. (1.54), si se toma el límite cuando $n \to \infty$ los minimizadores provenientes del método de iteración de valores convergen puntualmente a la política óptima, es decir, $f_n(x) \longrightarrow f^*(x)$ para todo $x \in X$.

6. RESULTADOS NUMÉRICOS

Como consecuencia de la cota dada en Teorema 19, Ec. (1.49) se obtiene el siguiente resultado numérico. Para $\varepsilon > 0$ dado, se tiene lo siguiente

$$2\sqrt{\frac{\left(x^2+\sigma^2\right)}{\left(1+\alpha K\right)\left(1-\alpha\right)}}\alpha^{n/2} = \varepsilon \tag{1.55}$$

Despejando n de la Ec. (1.55) se concluye que

$$\mathrm{n} = \frac{1}{\ln(\alpha)}\left[\ln(\varepsilon^2) + \ln(1+\alpha\mathrm{K}) + \ln(1-\alpha) - \ln(4) - \ln(\mathrm{x}^2+\sigma^2)\right] \tag{1.56}$$

En particular, si $\alpha = 1/4$, $\sigma = 1$ y $\mathrm{x} = 1$. Entonces se obtienen los siguientes resultados

ε	n	$f^*(1)$
0.1	5	-0.2359551
0.01	9	-0.236068
0.001	12	-0.236068
0.0001	15	-0.236068

7. CONCLUSIONES

En este artículo se usa la diferenciabilidad del modelo LQ para poder desarrollar la serie de Taylor centrada en la política óptima con el objetivo de obtener una relación entre las políticas provenientes del método de iteración de valores y las funciones de valor óptimo y con ello se determinó una cota para estos minimizadores.

8. REFERENCIAS

[1] Cruz-Suárez D. Una Estimación Puntual para la Política Óptima en Modelos Lineales con Costo Cuadrático. Memorias de la Octava Conferencia Iberoamericana en Sistemas, Cibernética e Informática, Vol. 2, No. 1, 88-92, 2005.

[2] Cruz-Suárez D. and Montes-de-Oca, R. Uniform Convergence of the Value Iteration Policies for Discounted Markov Decision Processes. Boletín de la Sociedad Matemática Mexicana, Vol. 12, 133-152, 2006.

[3] Cruz-Suárez D., Montes-de-Oca R. and Salem-Silva F.: Conditions for the uniqueness of optimal policies of discounted Markov decision processes. Mathematical Methods of Operations Research, Vol. 60, 415-436, 2004.

[4] Cruz-Suárez H. and Montes-de-Oca R.: An envelope theorem and some applications to discounted Markov decision processes. Mathematical Methods of Operations Research, Vol. 67, No. 2, 299-321, 2008.

[5] Hernández-Lerma O. and Lasserre J. B.: Further Topics on Discrete-Time Markov Control Processes. Springer Verlag, 1999.

[6] Jaen, C., Pou, J., Pindado, R., Sala, V., & Zaragoza, J. A linear-quadratic regulator with integral action applied to pwm dc-dc converters. In IECON 2006-32nd Annual Conference on IEEE Industrial Electronics, 2280-2285, 2006.

[7] Kang, J., Coates, J. T., Strawderman, R. L., Rosenstein, B. S., & Kerns, S. L. Radiogenomics models in precision radiotherapy: from mechanistic to machine learning. arXiv preprint arXiv:1904.09662, 2019.

[8] Marsden J. E. and Tromba A. J.: Vector Calculus, Freeman and Company, New York, Fifth Edition, 2003.

[9] Prakash, S., Sahoo, S., & Mishra, S. A linear quadratic regulator for small signal stability improvement of grid connected PMSG. In 2018 IEEMA Engineer Infinite Conference, 1-6, 2018.

[10] Realdon, M., & Boonyanet, W. Linear–quadratic term structure models for negative euro area yields. Economics Letters, Vol. 155, 149-153, 2017.

[11] Rudin W.: Principles of Mathematical Analysis. McGraw-Hill, Third Edition, 1976.

[12] Umenberger, J., & Schön, T. B. Learning convex bounds for linear quadratic control policy synthesis. In Advances in Neural Information Processing Systems, 9561-9572, 2018.

[13] West, K. D. Hansen and Sargent's Recursive Models of Dynamic Linear Economies: A Review Essay. Journal of Economic Literature, Vol. 55, No. 1, 173-81, 2017.

[14] Zacarías G. y Cruz-Suárez H.: Un Método Iterativo para resolver un Modelo Lineal Cuadrático, Memorias de la Octava Conferencia Iberoamericana en Sistemas, Cibernética e Informática, Vol.1, 286-291, 2009.