Una Estimación Puntual para la Política Óptima en Modelos Lineales con Costo Cuadrático.

Daniel CRUZ-SUÁREZ
División Académica de Ciencias Básicas
Universidad Juárez Autónoma de Tabasco
Apdo. Postal 5, Cunduacán, Tabasco 86690, México.
e-mail: daniel.cruz@basicas.ujat.mx

RESUMEN

El problema de resolver la ecuación de Ricatti algebraica a tiempo discreto, asociada a los sistemas lineales con costo cuadrático, que surge en la determinación de la política óptima en Procesos de Control de Markov (PCMs) Descontados, se complica para dimensiones mayores que 1. Aquí se propone, por medio de un algoritmo, estimar la política óptima en PCMs Descontados, sin tener que recurrir a la solución de la ecuación de Ricatti. Se presentan los resultados que garantizan la convergencia de tal algoritmo.

Palabras Claves: Procesos de Control de Markov Descontados, Horizonte de Pronóstico, Ecuación de Ricatti, Ecuación de Programación Dinámica, Iteración de Valores, Modelo Lineal con Costo Cuadrático.

1. INTRODUCCIÓN.

Los sistemas lineales con costo cuadrático son comunes en la teoría de control automático de un movimiento o de un proceso (ver [1, 9]).

El modelo que consideramos en el presente trabajo es el modelo estudiado en [2], con la diferencia que aquí consideramos el caso estacionario. En [2] se resuelve el problema de control de Markov de horizonte infinito, mostrando la existencia de una solución. El problema de dicha solución radica en que se presenta por medio de una relación matricial, que involucra la solución de la llamada ecuación de Ricatti, que en dimensiones superiores no es simple su solución.

El algoritmo que proponemos (ver [3]), nos aproxima tal solución por medio de la política óptima del problema truncado, es decir, de horizonte finito. El algoritmo nos permite estimar la política óptima de un Proceso de Control de Markov (PCM) Descontado, a tiempo discreto, estacionario y de horizonte infinito. Esta política óptima se aproxima por las políticas óptimas de los problemas correspondientes de horizonte finito.

El problema de horizonte finito permite hacer cálculos computacionales simples y obtener una solución.

Previamente hemos abordado el concepto de Horizonte de Pronóstico (HP), (ver [5, 6]). Se ha presentado un resultado que garantiza la existencia del HP, para ciertos modelos (ver [5]) y se ha dado una revisión de todos los resultados de HP en sistemas de inventarios controlados (ver [6]). La estimación de la política óptima de un PCM descontado permite dar una estimación del HP en PCMs descontados.

El trabajo se organiza de la siguiente manera. En la Sección 2 se proporciona la teoría de los sistemas lineales con costo cuadrático. En la Sección 3 se propone un problema equivalente compactificando el conjunto de controles, en la Sección 4 se proporciona la estimación de la política óptima y finalmente se presenta un ejemplo numérico de un modelo lineal con costo cuadrático.

2. SISTEMAS LINEALES CON COSTO CUADRÁTICO.

En esta sección presentamos el problema en el que estamos interesados, es una versión estacionaria del modelo estudiado en [2]. Damos la teoría básica, de la misma manera que en [3, 4, 5, 6, 7, 8], consideremos un modelo de Control estándar, (*X, A,* {*A(x):x*Î*X* }, *F, c*), (ver [5, 6]), el cual consiste de un espacio de estados *X,* de un espacio de acciones o controles *A,* de un conjunto de controles admisibles *A(x),* de una dinámica *F* y una función de costo *c:X´A®IR.* En este modelo se supone que *X*= *A*= *A(x),* para cada *x*Î*X.*

Para el presente trabajo, es suficiente suponer que *X* y *A* son subconjuntos, no vacíos, de espacios euclidianos y que *c* es una función medible definida en *X´A. F:X´A´S®X,* es una función medible, donde *S* es también un subconjunto, no vacío, de espacios euclidianos. Aquí *F* se interpreta como la dinámica siguiente a tiempo discreto:

$$x_{t+1} = Bx_t + Da_t + w_t, \quad t = 0,1,... \tag{1}$$

donde, para cada *t,* $x_t \in X$ denota el estado del sistema, $a_t \in A$ es el control correspondiente, w_t representa el ruido o perturbación aleatoria con valores en S y B, D son matrices dadas de dimensión adecuada. La función de costo cuadrático de un paso se define como:

$$c(x,a) = x'Qx + a'Ra, \tag{2}$$

donde Q y R son matrices simétricas dadas de dimensiones adecuadas, Q semidefinida positiva, R definida positiva, $x \in X$, $a \in A$ y donde " ´ " representa la transpuesta de la matriz. A este modelo le llamaremos el modelo general.

Denotemos por *IF* el conjunto de selectores medibles:

$$IF = \{ f : X \to A \mid f \text{ es medible y } f(x) \in A(x) \forall x \}.$$

Una *Política Markoviana* se define como una sucesión $p = \{f_t\}_{t=0}^{\infty}$ tal que $f_t \in IF$, para todo *t.* Si $f_t \equiv f \in IF$, para todo *t,* diremos que p es una *Política Estacionaria*. Por convención, identificaremos a IF con el conjunto de políticas estacionarias. Sea Π el conjunto de todas las políticas.

El objetivo es minimizar el costo total descontado esperado, el cual se define a continuación.

Si la política usada es $\pi \in \Pi$, dado el estado inicial x, el costo total descontado esperado se define por

$$V(\pi, x) = E_x^{\pi} \sum_{t=0}^{N} \alpha^t c(x_t, a_t), \quad N \leq \infty. \qquad (3)$$

Donde, N es el horizonte del problema, $\{x_t\}$ y $\{a_t\}$ denotan las sucesiones de estados y controles respectivamente, $\alpha \in (0,1)$ es el factor de descuento y E_x^{π} es el operador esperanza con respecto a la medida de probabilidad P_x^{π} inducida por el par (π, x).

El problema de control óptimo consiste en encontrar una política π^* tal que $V(\pi^*, x) = V^*(x)$, para todo $x \in X$, donde

$$V^*(x) = \inf_{\pi \in \Pi} V(\pi, x), \quad x \in X. \qquad (4)$$

En este caso, se dice que π^* es una política óptima y que V^* es la función de valores óptimos. Existen diversas metodologías para resolver este problema, por ejemplo, Programación Dinámica e Iteración de Valores, (ver [7, 8]).
Para el problema lineal con costo cuadrático, se tienen los siguientes resultados, adaptados al caso estacionario considerado aquí (ver [2]). La ley de control óptimo para cualquier k tiene la forma:

$$f_k(x) = L_k x, \qquad (5)$$

donde las matrices L_k están dadas por la ecuación:

$$L_k = -\alpha\left(\alpha B' K_{k+1} B + R\right)^{-1} B' K_{k+1} A, \qquad (6)$$

y donde las matrices simétricas, semidefinidas positivas K_k están dadas por

$$K_0 = Q, \qquad (7)$$

$$K_{k+1} = A'\left(\alpha K_k - \alpha^2 K_k B\left(\alpha B' K_k B + R\right)^{-1} B' K_k\right) A + Q. \qquad (8)$$

La ecuación anterior se llama la ecuación de Ricatti a tiempo discreto.
En el presente contexto estacionario, es decir, en donde A, B, Q, R son constantes en cada etapa, la solución K_k converge cuando $k \to +\infty$ a una solución K que satisface la ecuación de Ricatti algebraica:

$$K = A'\left(\alpha K - \alpha^2 K B\left(\alpha B' K B + R\right)^{-1} B' K\right) A + Q. \qquad (9)$$

Es posible demostrar que cada f_k de (5) son distintas entre sí y que el límite de ellas, denotado por f^*, también es distinto a cada una de ellas.
La ley de control óptima para el problema infinito está dada por la relación siguiente:

$$f^*(x) = Lx \qquad (10)$$

donde la matriz L está dada por:

$$L = -\alpha\left(\alpha B' K B + R\right)^{-1} B' K A \qquad (11)$$

Obsérvese que la ecuación (9) en una dimensión, es fácil resolverla, pero en dimensiones superiores se complica la solución. Además, las funciones de iteración de valores están dadas por:

$$V_{k+1}(x) = x' K_k x + \sum_{m=0}^{k-1} \alpha^m E[w' K_m w], \qquad (12)$$

y la función de valores óptimos es:

$$V^*(x) = x' K x + \frac{\alpha}{1-\alpha} E[w' K w]. \qquad (13)$$

3. UN PROBLEMA EQUIVALENTE.

El algoritmo que se propone requiere de una tasa de convergencia de las funciones de iteraciones de valor a la función de valor, para proponerla seguimos [8], en donde las hipótesis que se consideran son las siguientes:

HIPOTESIS (H1).
Para cualquier $x \in X$:

a. El conjunto de controles permitidos *A(x)* es compacto.
b. $c(\cdot,\cdot)$ es no-negativo, semicontinuo inferiormente.
c. La función

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

es continua en $a \in A(x)$ para cualquier función $u \in IB(X)$, donde $IB(X)$ denota el espacio de Banach de funciones u sobre X, real valuadas medibles y acotadas, con la norma del supremo $\|u\| = \sup_X |u(x)|$.

d. Existe $\pi \in \Pi$, tal que $V(\pi, x) < +\infty, \quad \forall x \in X$.

HIPOTESIS (H2)
Existen constantes no negativas $\bar{c}$ y b, con

$$1 \leq b \leq 1/\alpha$$

y una función de peso $W \geq 1$ sobre X tal que para cualquier estado $x \in X$:

a. $\sup_{A(x)} |c(x,a)| \leq \bar{c} W(x)$;
b. $\sup_{A(x)} \int W(y) Q(dy \mid x, a) \leq b\, W(x)$.

Para que nuestro modelo satisfaga las hipótesis anteriores necesitamos modificarlo de manera que los conjuntos de controles admisibles *A(x)* sean compactos.
Consideremos, de nuevo, un modelo lineal con costo cuadrático como el definido en la Sección anterior, es decir, un modelo de Control estándar, $(X, A, \{A(x): x \in X\}, F, c)$, (ver [7, 8]), tal que X es el espacio de estados, A es el espacio de acciones. La dinámica se denota por F y la función de costo por $c: X \times A \to IR$.
Como antes, suponemos que X y A son subconjuntos euclidianos, no vacíos, que la dinámica es:

$$x_{t+1} = Bx_t + Da_t + w_t\,, \quad t = 0,1,... \qquad (14)$$

donde, para cada t, $x_t \in X$, $a_t \in A$, w_t Î S y B, D son matrices dadas de dimensión adecuada.
La función de costo cuadrático de un paso se define como:

$$c(x,a) = x'Qx + a'Ra \qquad (15)$$

donde Q y R son matrices dadas de dimensión adecuada y $x \in X$, $a \in A$.
Ahora, consideramos al conjunto de restricciones compacto, como se indica en las siguientes hipótesis.

HIPOTESIS (H3).
a. Suponemos que los vectores aleatorios w_t, *t=0, 1, …* son independientes e idénticamente distribuidos con valores en *S*, con función de densidad conjunta continua dada por Δ, con media cero y matriz de varianzas y covarianzas con entradas cero fuera de la diagonal principal, con varianza común para cada variable aleatoria igual a s^2.
b. El conjunto de restricciones se define, para cada $x \in X$, como:

$$A(x) = \{a \in A \,|\, |a| \le |Ux|\}, \qquad (16)$$

donde $|\cdot|$, representa la norma del espacio correspondiente y U es la cota dada en el Lema 2 (ver abajo).

A este modelo lo llamaremos el modelo con restricciones. Demostraremos que este nuevo modelo lineal con costo cuadrático satisface las hipótesis H1 y H2.

LEMA 1.
Supóngase que el modelo con restricciones satisface H3. Entonces, también satisface H1 y H2.

Demostración.
Obsérvese que a y b de H1 son inmediatas probaremos H1c. Demostraremos que $Q(\cdot|x,a), x \in X, a \in A(x)$, inducido por (1) es continuo. Sea B un elemento de la σ-álgebra de Borel de *X*, $x \in X, a \in A(x)$. Entonces

$$Q(B|x,a) = \int \cdots \int I_B(Bx + Da + s)\Delta(s)ds$$

$$= \int \cdots \int_B \Delta(l - (Bx + Da))dl,$$

donde la última igualdad se debe al Teorema de Cambio de Variables y $ds = (ds_1, \cdots, ds_n)$, si n es la dimensión, lo mismo vale para dl. Por la hipótesis H3 se sigue que $\Delta(\cdot - (Bx + Da))$ es continua en (x,a). Entonces, por el Ejemplo C.6 del Apéndice C de [7], obtenemos que Q es fuertemente continuo.
Para probar que cumple H2, supongamos por simplicidad que *B*, *D* y Q son matrices de dimensión $n \times n$ y que la matriz *R* tiene dimensión $m \times m$. Ahora, definamos las siguientes expresiones:

$$|Q| = \max\{|q_{ij}|\,|\, 1 \le i \le j \le n\},$$

$$|R| = \max\{|r_{ij}|\,|\, 1 \le i \le j \le m\},$$

$$|B| = \max\{|b_{ij}|\,|\, 1 \le i \le j \le n\},$$

$$|D| = \max\{|d_{ij}|\,|\, 1 \le i \le j \le n\},$$

$$c = \sum_{k=1}^{n} x_k^2\,, \quad x = (x_1, \cdots, x_n)\,.$$

Entonces, definiendo

$$\overline{c} = 1,$$

$$W(x) = \left[|Q|n^2 + |R|m^2\right]c\,,$$

$$b = n^3(|B| + |D|)^2\,,$$

$$s^2 = \frac{\left[|Q|n^2 + |R|m^2\right][b-1]}{n}\,,$$

se demuestra que se satisface H2. ÿ

Ahora, demostramos que el modelo con restricciones tiene la misma solución que el modelo general.

LEMA 2.
La solución óptima del problema con restricciones es la misma que la del problema general.
Para la demostración, sólo obsérvese que como L_k ®L, la sucesión $\{L_k\}$ está acotada, digamos por *U*. De donde *A*(*x*), definido en (16), contiene a la solución del problema general.

4. ESTIMACIÓN.

Ahora, presentamos los resultados para estimar la política óptima en el modelo lineal con costo cuadrático.

LEMA 3.
Para el problema con restricciones se tiene que:

a. La política óptima $f^*(x) = Lx$, es única.
b. $f_n(\cdot) \xrightarrow{U|C} f^*(\cdot)$ en donde C es un subconjunto compacto de *X*.

Demostración.
a. Esta afirmación se encuentra en [4].
b. Sean $C \subset X$, un subconjunto compacto de X y $x \in C$ fijo. Sin pérdida de generalidad, se supone que C se puede encerrar en una bola de radio $M > 0$, con centro en el origen.
Se tiene que $f_k(x) = L_k x$ y $L_k \to L$, cuando k tiende a infinito. Entonces dado *e* >0, existe *N* tal que:

$$|L_n - L| < \frac{e}{M}\,,$$

Entonces,

$$\left|f_n(x) - f^*(x)\right| = \left|L_n - L\right|\left|x\right| \leq \left|L_n - L\right| M < e \ ,$$

para todo $x \in C$, lo que demuestra :

$$f_n(\cdot) \xrightarrow{U|C} f^*(\cdot). \qquad \text{ÿ}$$

Antes de dar el siguiente resultado consideremos la notación siguiente.

Notación.
Para abreviar usaremos la notación:

$$G_n(x,a) := c(x,a) + a \int V_n(y) Q(dy|x,a),$$

$$G(x,a) := c(x,a) + a \int V^*(y) Q(dy|x,a),$$

$$D_n(x,a) := \left|G_n(x,a) - V_n(x)\right| .$$

TEOREMA 4.
Sea $B = B_e(f_n(x))$. Entonces, existe un entero N^* tal que $\inf_{a \notin B} D_{N^*}(x) > O(N^*)$, donde

$$O(n) = \frac{2\overline{c}(ab)^{n+1} W(x)}{1 - ab} > 0 \ .$$

Demostración.
Supóngase lo contrario, es decir, para toda n existe a_n tal que:

$$D_n(x, a_n) \leq O(n),$$

Como $a_n \in A(x) - B$ (compacto) existe una sub-sucesión $\{a_{nk}\}$ y $\overline{a} \in A(x) - B$ tal que $a_{nk} \to \overline{a}$, entonces

$$D_{nk}(x, a_{nk}) \to 0, \quad k \to +\infty ,$$

de donde

$$G(x, \overline{a}) = G(x, f^*(x)),$$

lo cual implica que

$$\overline{a} = f^*(x),$$

por unicidad, pero

$$d(a_{nk}, f_{nk}(x)) \geq e \ ,$$

para toda k, si hacemos k ® +¥, se tiene que

$$d(\overline{a}, f^*(x)) \geq e$$

lo cual es una contradicción, por lo tanto, existe N^* con la propiedad requerida. ÿ

El resultado anterior es posible extenderlo, para encontrar un N^* uniforme en x, para $x \in C$, donde C es un subconjunto compacto de X. En cuyo caso el algoritmo es de la forma siguiente.

ALGORITMO
Para un $x \in X$ fijo encontrar el mínimo n tal que:

$$\inf_{a \notin B} D_n(x,a) > O(n),$$

donde $B = B_e(f_n(x))$.

5. EJEMPLO.

Consideremos el caso en el cual $X=A=IR^2$, el conjunto de restricciones es:

$$A(x,y) = \left\{(a,b) \in A \,\middle|\, a^2 + b^2 \leq x^2 + y^2\right\},$$

la dinámica o transición está dada por:

$$x_{k+1} = Bx_k + Da_k + x_k \ ,$$

donde, para cada k, $x_k \in X$ denota el estado del sistema, $a_k \in A$ es el control correspondiente. B y D son matrices identidad 2×2, $x_k = (u_k, g_k)$ donde u_k, g_k son variables aleatorias independientes e idénticamente distribuidas con media cero y varianza común igual a 19/4. El costo de una etapa está definido como:

$$c(\overline{x}, \overline{a}) = \overline{x}' Q \overline{x} + \overline{a}' R \overline{a} \ ,$$

donde $\overline{x} \in X, \overline{a} \in A$ y Q, R son matrices 2×2, R es la identidad y

$$Q = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},$$

En este ejemplo, resolver la ecuación de Ricatti, implica resolver un sistema de cuatro ecuaciones con cuatro incógnitas con términos cuadráticos y mixtos. Aunque, en esta dimensión 2×2, es manejable la solución, para dimensiones mayores se complica.

LEMA 5.
Si

$\overline{c} = 1,$

$W(x,y) = 2(x^2 + xy + y^2) + 1,$

$b = 20,$

$a = \dfrac{1}{30},$

entonces se satisfacen a y b de H3. Además, el orden de convergencia de V_n a V^* (ver [8]), está dado por:

$$2\overline{c}\frac{(ab)^{n-1}}{1 - ab} W(x,y) = 18\left(\frac{2}{3}\right)^{n-1} .$$

Con los valores dados en el Lema 6, podemos calcular para cada *N*, el orden de convergencia. Para los primeros 20 valores se tiene:

N	$18(2/3)^{n-1}$
1	18.
2	12.
3	8.
4	5.33333
5	3.55556
6	2.37037
7	1.58025
8	1.0535
9	0.702332
10	0.468221
11	0.312148
12	0.208098
13	0.138732
14	0.0924882
15	0.0616588
16	0.0411058
17	0.0274039
18	0.0182693
19	0.0121795
20	0.00811967

LEMA 6.
Para el ejemplo se tiene que:

$$V_n(\overline{x}) = 1.03219(x+y)^2 + 0.373716,$$

$$f_n(\overline{x}) = \begin{bmatrix} -0.0321912 & -0.0321912 \\ -0.0321912 & -0.0321912 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix},$$

para toda *n³ 5*, $\overline{x} = (x, y)$.

Para el ejemplo que estamos considerando tomemos $\overline{x} = (1,0)$. Dado *e* = 0.1, definimos $B = B_{0.1}(f_n(1,0))$. Entonces, con los anteriores valores se puede encontrar que:

$$\inf_{a \notin B} D_n(1,0) = 0.0298415$$

para todo *n³ 5*. Por lo tanto, comparando con la tabla anterior obtenemos que N^* = 17. Lo que nos dice que para valores mayores a 17, la política óptima del problema finito f_n, en el punto $\overline{x} = (1,0)$, está dentro de la *e*-vecindad, para *e* =0.1, de la política óptima del problema de horizonte infinito. Entonces, podemos aproximar la política óptima de horizonte infinito f^* por f_n.
Obsérvese que para el caso $(x, y) = (1,0)$ el coeficiente de $f_n(1,0)$ es la solución de la ecuación de Ricatti para $n \geq 3$.

CONCLUSIONES.
En este trabajo hemos resuelto el problema lineal con costo cuadrático, también conocido como del regulador lineal, por medio de un algoritmo. Este algoritmo, estima la política óptima sin tener que recurrir a la solución explicita de la ecuación de Ricatti. Presentamos un ejemplo numérico, para ilustrar la teoría.

6. REFERENCIAS.

1. Athans M., "The Role and Use of the Stochastic Linear-Quadratic-Gaussian Problem in Control System Design", **IEEE Transactions on Automatic Control**, Vol. AC-16, No. 6, December 1971, pp. 529-552.

2. Bertsekas D.P., **Dynamic Programming: Deterministic and Stochastic Models**. Prentice-Hall, Englewood Cliffs, NJ, 1987.

3. Cruz-Suárez D. and Montes-De-Oca R., "Minimizers of the Value Iteration Algorithm of Discounted Markov Decision Processes: Convergence to Optimal Policies", sometido 2004.

4. Cruz-Suárez D., Montes-De-Oca R. and Salem-Silva F., "Conditions for the Uniqueness of Optimal Policies of Discounted Markov Decisión Processes", **Mathematical Methods of Operations Research,** Vol. 60, No. 3, 2004, pp. 415-436.

5. Cruz-Suárez H. D., "Horizonte de pronóstico en procesos de control de Markov", **Memorias CISCI 2002**, Vol. 2, Pag. 269.

6. Cruz-Suárez H. D., "Horizonte de pronóstico en sistemas de inventarios controlados: una revisión panorámica", **Memorias CISCI 2003**, Vol. 2, Pag. 185.

7. Hernández-Lerma O. and Lasserre, J.B., **Discrete Time Markov Control Processes**, Springer-Verlag, New York, 1996.

8. Hernández-Lerma O. y Lasserre, J.B., **Further Topics on Discrete-Time Markov Control Processes**, Springer-Verlag, New York, 1999.

9. Martínez-Morales M., "Adaptive Premium Control in an Insurance Risk Processes", Ph. D. Thesis, Texas Tech University, 1991.