Procesos de Decisión de Markov Descontados: Soluciones Óptimas mediante Problemas de Control Determinista Diferenciables

Hugo Cruz-Suárez
Departamento de Matemáticas
Universidad Autónoma Metropolitana-Iztapalapa,
Av. San Rafael Atlixco 186, Col. Vicentina,
México D.F. 09340, México

RESUMEN

Este artículo está relacionado con la teoría de Procesos de Control de Markov. En él se tratarán problemas de control descontados a tiempo discreto, con horizonte infinito en su versión determinista y estocástica (ver [2, 4, 5, 8]). Para los PCMs que consideramos, suponemos que el espacio de estados y controles son espacios euclidianos n-dimensionales ($n\geq 1$). Entonces dado un Proceso de Control de Markov descontado, cuya dinámica está perturbada por un ruido aleatorio (por simplicidad lo llamaremos: Problema Estocástico), le asociaremos cierto Proceso de Control de Markov descontado determinista (llamado: Problema Determinista) del cual suponemos conocida su solución, y que dicha solución es diferenciable. Por tanto, la idea general del trabajo consiste en obtener la solución del Problema Estocástico a partir de la solución del Problema Determinista. La teoría desarrollada será ejemplificada con algunos modelos clásicos de la teoría de control.

Palabras Claves: Procesos de Control de Markov Descontados, Diferenciabilidad de soluciones óptimas, Ecuación de Euler, Regulador Lineal, Crecimiento Económico.

1. INTRODUCCIÓN.

Este trabajo tratará con Procesos de Control de Markov (PCMs) con horizonte infinito y con costo total descontado. Para los PCMs considerados, suponemos que los espacios de estados y controles son espacios euclidianos n-dimensionales ($n\geq 1$). También suponemos que la ley de transición es inducida por una ecuación en diferencias estocásticas. La meta principal en problemas de control óptimo es determinar la función de valores óptimos y la política de control óptimo. Desafortunadamente, este objetivo es frecuentemente muy difícil o quizás imposible de obtenerse explícitamente. En este artículo daremos una posible forma de abordar este problema.

Supondremos que conocemos la política óptima y el valor óptimo de cierto problema determinista asociado al caso estocástico y que dichas funciones son diferenciables en su dominio respectivo. Bajo ciertas condiciones sobre el modelo de control probaremos que la solución estocástica del problema puede ser inducida por la solución determinista.

El artículo se organiza de la forma siguiente, comenzamos haciendo una revisión de los Modelos de Control Estocásticos y Deterministas, también presentamos algunos ejemplos deterministas resueltos por métodos variacionales, alternativos y/o complementarios a Programación Dinámica. En la sección siguiente, se dan algunos elementos de la teoría de control, los cuales serán necesarios en el desarrollo subsecuente. Después, presentamos los resultados principales del artículo y retomamos los ejemplos deterministas para analizar sus versiones estocásticas con los resultados obtenidos. Finalmente se da una sección de conclusiones y otra de referencias.

2. MODELOS DE CONTROL.

Problemas de control estocástico
Consideremos un *Modelo de Control de Markov* (MCM) a tiempo discreto y estacionario, $(X, A, \{A(x): x\in X\}, Q, c)$ (ver [2, 4, 5, 8]), donde:

a) X y A son espacios de Borel no vacíos, llamados el espacio de estados y controles respectivamente;

b) $\{A(x)\,|\,x\in X\}$ es una familia de subconjuntos medibles no vacíos $A(x)$ de A, donde a cada estado $x\in X$ le asociamos $A(x)$, cuyos elementos son las acciones admisibles o controles, cuando el sistema está en el estado x. El conjunto IK de pares de estados-acciones admisibles, está definido por:

$$\text{IK} := \{(x, a)\,|\,x\in X, a\in A(x)\},$$

el cual se supone que es un conjunto medible del espacio producto X×A.

c) Q es la ley de transición o kérnel estocástico sobre X dado IK (i.e para $B\in B(X)$, $(x, a)\in$ IK y $t=0, 1,\ldots$, $Q(B\,|\,x, a) := \text{Prob}(x_{t+1}\in B\,|\,x_t = x, a_t = a)$, donde $\{x_t\}$ representa la sucesión de estados y $\{a_t\}$ la sucesión de controles, Q satisface: $Q(B\,|\,\bullet)$ es medible y $Q(\bullet\,|\,x, a)$ es una medida de probabilidad sobre X). Supondremos que la ley de transición es inducida por una ecuación en diferencias estocásticas de la forma siguiente

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

para $t=0,1,2\ldots$ con un estado inicial x_0 dado. Donde $G: X\times A\times S\to X$ ($S\subseteq\Re^k$, $k\geq 1$) es una función medible conocida y $\{\xi_t\}$ es una sucesión de variables aleatorias, independientes e idénticamente distribuidas (i.i.d.), independientes de x_0;

d) r: IK $\to\Re$ es una función medible y la llamaremos la función de recompensa en un paso.

Definición 1. IF denota la colección de funciones medibles $f : X \to A$ tal que $f(x) \in A(x)$ para todo $x \in X$.

Definición 2. En general, una política de control es una sucesión $\pi = \{\pi_t,\ t = 0, 1, \ldots\}$ donde para cada $t=0,1,\ldots$, $\pi_t(\bullet \mid h_t)$ es un kernel estocástico definido en la σ-álgebra de Borel $\mathcal{B}(A)$ dada la historia

$$h_t=(x_0,a_0,\ldots,x_{t-1},a_{t-1},x_t)$$

y la cual satisface la condición $\pi_t(A(x_t) \mid h_t)=1$. En este caso $\{x_t\}$ y $\{a_t\}$ denotan la sucesión de estados y controles respectivamente. Se denotará por Π al conjunto de todas las *políticas de control*. En especial estamos interesados en la clase de *políticas estacionarias*. Una política es estacionaria, si existe $f \in$ IF tal que $\pi_t = f$ para cada t.

Definición 3. Una vez que se tiene un MCM y un conjunto de políticas Π, definimos el siguiente criterio de rendimiento. Para cualquier política $\pi \in \Pi$ y estado inicial $x \in X$,

$$V(\pi,x) := E_x^{\pi}\left[\sum_{t=0}^{\infty} \alpha^t c(x_t,a_t)\right], \qquad (2)$$

$V(\pi, x)$ es llamado *el Costo Total Esperado α-descontado*. Donde $\alpha \in (0, 1)$ es conocido como el factor de descuento y E_x^{π} denota la esperanza respecto a la medida de probabilidad P_x^{π}, inducida por $x \in X$ y $\pi \in \Pi$ (ver [5]).

Problema de Control Óptimo
Sean $(X, A, \{A(x) \mid x \in X\}, Q, c)$ un MCM, Π el conjunto de políticas y como criterio de rendimiento considérese la recompensa total descontada $V(\pi, x)$, donde $\pi \in \Pi$ y $x \in X$ (ver (2)). El *problema de control óptimo* consiste en determinar una política $\hat{\pi} \in \Pi$, tal que:

$$V(\hat{\pi},x) = \inf_{\pi} V(\pi,x),\ x \in X. \qquad (3)$$

Toda política que satisface la relación anterior será llamada *óptima*. La función V definida por:

$$V(x) := \inf_{\pi} V(\pi,x),\ x \in X, \qquad (4)$$

se llamará la *función de valores óptimos*.

Problemas de control deterministas
Los problemas de control óptimo determinista que consideraremos son aquellos que presentan el mismo modelo de control descrito al inicio de esta sección, a excepción de la dinámica, donde únicamente dependerá del estado y de la acción, es decir

$$x_{t+1} = F(x_t, a_t),\ t = 0,1,\ldots$$

Donde F: $X \times A \to X$ es una función medible conocida. El conjunto de políticas admisibles al problema serán la clase de políticas deterministas, las cuales son definidas sobre el espacio de historias admisibles y con valores en el espacio de controles (ver [4, 8]).

Para dar solución al problema de control determinista se cuenta con métodos iterativos como iteración de valores o de políticas (ver [2, 4, 5, 8]) ó variacionales como el principio del máximo y la ecuación de Euler (ver [1,6, 7]).
Para un desarrollo detallado del problema de control determinista puede consultarse [2].

Nota: El desarrollo de esta sección es posible hacerlo en términos de recompensa solo es necesario cambiar ínfimo por supremo en (3) y (4).

Ejemplo
1. Regulador Lineal (ver [2])
Sea $X=A=A(x)=\Re$, $x \in X$ y consideremos constantes reales γ, β, q y r con $\gamma \cdot \beta \neq 0$. La dinámica y la función de costo son,

$$\begin{aligned} x_{t+1} &= \gamma x_t + \beta a_t, \\ c(x_t,a_t) &= q x_t^2 + r a_t^2. \end{aligned} \qquad t=0,1,2\ldots$$

Solución. La ecuación de programación dinámica del problema para $x \in X$ es,

$$W(x) = \inf_{a \in A(x)} [q x^2 + r a^2 + \alpha W(\gamma x + \beta a)].$$

Suponiendo que la función de valor y la política óptima son diferenciables (ver [3]), procedemos de la forma siguiente.

Derivando el lado de derecho de la ecuación de programación dinámica e igualando a cero, obtenemos la ecuación conocida como *condición de primer orden*.

La condición de primer orden se satisface para una política $g \in$ IF, entonces para $x \in X$,

$$\begin{aligned} 2rg(x) + \alpha\beta W'(\gamma x + \beta g(x)) &= 0 \\ \alpha W'(\gamma x + \beta g(x)) &= -\frac{2rg(x)}{\beta} \end{aligned} \qquad (5)$$

La política $g \in$ IF minimiza el lado derecho de la ecuación de programación dinámica, es decir,

$$W(x) = q x^2 + r(g(x))^2 + \alpha W(\gamma x + \beta g(x)), \qquad (6)$$

$x \in X$.

Derivando (6),

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

$x \in X$. Utilizando (5) en la derivada de W, se obtiene

$$W'(x) = 2qx - \frac{2\gamma r g(x)}{\beta}, \qquad (7)$$

$x \in X$.
Despejando g de (7), se encuentra

$$g(x) = \frac{\beta}{2\gamma r}(2qx - W'(x)), \qquad (8)$$

$x \in X$.
Sustituyendo (8) en (5), obtenemos la ecuación funcional conocida como *Ecuación de Euler (EE)*,

$$\alpha W'\left[\frac{2\gamma^2 rx+\beta^2(2qx-W'(x))}{2\gamma r}\right]=-\frac{1}{\gamma}(2qx-W'(x)),$$

$x\in X$.

Evaluando la ecuación anterior (EE) en el punto de equilibrio $\bar{x}$, es decir, en el punto que satisface

$$\bar{x}=\gamma\bar{x}+\beta g(\bar{x}), \tag{9}$$

se tiene

$$\alpha W'\left[\frac{2\gamma^2 r\bar{x}+\beta^2(2q\bar{x}-W'(\bar{x}))}{2\gamma r}\right]=-\frac{1}{\gamma}(2q\bar{x}-W'(\bar{x}))$$

Nótese que $\bar{x}$ satisface,

$$\begin{aligned}\bar{x}&=\gamma\bar{x}+\beta g(\bar{x}),\\&=\frac{2\gamma^2 r\bar{x}+\beta^2(2q\bar{x}-W'(\bar{x}))}{2\gamma r}.\end{aligned}$$

La última igualdad es debido a (8). Utilizando ésta información en la EE (evaluada en $\bar{x}$), se obtiene

$$W'(\bar{x})=\frac{-2q}{\gamma\alpha-1}\bar{x}$$

Sustituyendo en (9), encontramos que $\bar{x}=0, W'(\bar{x})=0$, es decir, $W'(0)=0$. Entonces por (8) se encuentra que

$$g(0)=0.$$

Por lo tanto, evaluando en cero la ecuación (6), se determina que $W(0)=0$.

Derivando implícitamente la Ecuación de Euler de manera sucesiva y evaluando en el punto de equilibrio, obtenemos

$\alpha\beta^2(W''(0))^2-2(\gamma^2\alpha r-r+q\alpha\beta^2)W''(0)-4qr=0$ y
$W^{(n)}(0)=0,$ para $n\geq 3$.

Entonces haciendo una expansión en series alrededor de $\bar{x}=0$,

$$W(x)=W(0)+W'(0)+\frac{1}{2}W''(0)x^2,\ x\in X.$$

Lo cual implica,

$$W(x)=Qx^2,\ x\in X.$$

Donde Q satisface la ecuación de *Ricatti* (ver [2]),

$$\alpha\beta^2Q^2-(\gamma^2\alpha r+q\alpha\beta^2-r)Q-qr=0. \tag{10}$$

La política óptima se obtiene sustituyendo el valor óptimo W, ya conocido, en (8). Resolviendo obtenemos para x en X,

$$g(x)=-\frac{Q\alpha\beta\gamma}{r+Q\alpha\beta^2}x.$$

3. RESULTADOS AUXILIARES

Sea MCM: (X, A, $\{A(x):x\in X\}$, Q, c) y V la función de valores óptimos definida por (4).

Definición 4. Sea $M(X)^+$ el conjunto de las funciones medibles no negativas definidas en X. Para cada $u\in M(X)^+$ definimos,

$$Tu(x):=\min_{a\in A(x)}\left[c(x,a)+\alpha E[V(G(x,a,\xi))]\right],\ x\in X.$$

El operador T es conocido como el *operador de programación dinámica*.

Definición 5. Una función ν:IK$\to\Re$ se dice que es *inf-compacta* sobre IK si, para cada $x\in X$ y $r\in\Re$ el conjunto $\{a\in A(x): \nu(x,a)\leq r\}$ es compacto.

Definición 6. Diremos que la ley de transición Q es *fuertemente continua* si la función

$$v'(x,a)=\int v(y)Q(dy|x,a)$$

es continua y acotada en IK para cada función medible y acotada v definida en X.

Definición 7. Sea Z un espacio topológico y υ una función definida en Z. La función υ se dice que es inferiormente semicontinua si el conjunto $\{z\in Z:\upsilon(x)\leq r\}$ es cerrado para cada $r\in\Re$.

Hipótesis 1.

a) El costo es inferiormente semicontinuo, no negativo e inf-compacto en IK.

b) La ley de transición es fuertemente continua.

c) Existe una política π tal que $V(\pi, x)<\infty$, para $x\in X$.

La prueba del Lema siguiente puede consultarse en [5].

Lema 1. Bajo las Hipótesis 1 (a) y 1 (b). Para cada $u\in M(X)^+$, Tu esta en $M(X)^+$. Además, existe un selector $f\in$IF tal que, para cada x en X,

$$Tu(x)=c(x,f(x))+\alpha\int_X u(y)Q(dy|x,f(x)).$$

La prueba del Lema 2 se incluye por completitud del artículo y debido a que es uno de los resultados claves en el desarrollo de los resultados principales. El Lema 2, es un resultado establecido en [5].

Lema 2. Supóngase que las Hipótesis 1 se cumplen, entonces,

1. Si $u\in M(X)^+$ y $u\geq Tu$ entonces $u\geq V$.
2. Si $u\in M(X)^+$, $u\leq Tu$ y para cada $\pi\in\prod$ y $x\in X$,

$$\lim_{n\to\infty}\alpha^n E_x^{\pi}[u(x_n)]=0 \tag{11}$$

entonces $u\leq V$.

Demostración. (a) Sea $u \in M(X)^+$ tal que $u \geq Tu$ y sea $f \in$ IF entonces,

$$u(x) \geq c(x, f(x)) + \alpha \int u(y) Q(dy|x, f(x)), \qquad (12)$$

$x \in X$.

Iterando (12), se obtiene

$$u(x) \geq E_x^f \left[\sum_{t=0}^{n-1} \alpha^t c(x_t, f(x_t)) \right] + \alpha^n \int u(y) Q^n(dy|x, f),$$

para $n \geq 1$ y $x \in X$.

Como u es no negativa,

$$u(x) \geq E_x^f \left[\sum_{t=0}^{n-1} \alpha^t c(x_t, f(x_t)) \right],$$

$n \geq 1$ y $x \in X$. Si $n \to \infty$

$$\begin{aligned} u(x) &\geq \lim_{n\to\infty} E_x^f \left[\sum_{t=0}^{n-1} \alpha^t c(x_t, f(x_t)) \right], \\ &= V(f, x), \\ &\geq V(x), \end{aligned}$$

$x \in X$. Por lo tanto, $u \geq V$.

(b) Ahora, supongamos que $u \leq Tu$ y seleccionemos $\pi \in \prod$ y $x \in X$. Utilizando la propiedad de Markov,

$$\begin{aligned} &E_x^\pi \left[\alpha^{t+1} u(x_{t+1}) | h_t, a_t \right] = \alpha^{t+1} \int_X u(y) Q(dy|x_t, a_t), \\ &= \alpha^t \left[c(x_t, a_t) - c(x_t, a_t) + \alpha \int_X u(y) Q(dy|x_t, a_t) \right], \\ &= \alpha^t \left[c(x_t, a_t) + \alpha \int_X u(y) Q(dy|x_t, a_t) \right] - \alpha^t c(x_t, a_t), \\ &\geq \alpha^t [u(x_t) - c(x_t, a_t)]. \end{aligned}$$

Por otro lado, tenemos

$$-\alpha^t c(x_t, a_t) \leq E_x^\pi [\alpha^{t+1} u(x_{t+1}) - \alpha^t u(x_t) | h_t, a_t].$$

Al tomar esperanza E_x^π y sumar desde 0 hasta *n-1*, se tiene

$$E_x^\pi \left[\sum_{t=0}^{n-1} \alpha^t c(x_t, a_t) \right] \geq u(x) - \alpha^n E_x^\pi [u(x_n)],$$

$n \geq 1$.

Utilizando (11), se tiene al tender n a infinito,

$$\lim_{n\to\infty} \left[\sum_{t=0}^{n-1} \alpha^t c(x_t, a_t) \right] \geq u(x)$$

Entonces $V(\pi,x) \geq u(x)$. Como π y x eran arbitrarios, se concluye que $V \geq u$. □

4. RESULTADOS PRINCIPALES

En esta sección se presentan dos Teoremas los cuales dan condiciones para resolver el problema de control estocástico vía un problema determinista asociado.

Consideremos un MCM: $(X, A, \{A(x): x \in X\}, Q, c)$ y el conjunto de políticas estacionarias IF con el criterio de rendimiento costo total descontado. Supondremos que la dinámica estocástica (1) es de la forma siguiente,

$$G(x, a, \xi) = L(F(x, a), \xi) .$$

Donde L: $X \times S \to X$ es una función medible conocida y F: $X \times A \to X$ es precisamente la función que describe a la dinámica determinista.

Para este problema estocástico consideramos el problema de control determinista definido en el mismo espacio de estados, controles, restricciones y con el mismo costo en común. La única diferencia es en la dinámica la cual presenta la forma siguiente,

$$x_{t+1} = F(x_t, a_t) , \; t=0,1,2,\ldots$$

Hipótesis 2.

(a) Supongamos que se conocen el valor óptimo W y la política óptima g del problema de control determinista.

(b) Supongamos que existe una función medible no negativa h tal que se satisface la relación siguiente,

$$E[W(L(F(x,a),\xi))] = W(F(x,a)) + E[h(\xi)] \text{ ó}$$
$$W(F(x,a)) \cdot E[h(\xi)],$$

con $E[h(\xi)] < \infty$ y $(x, a) \in$ IK.

Supongamos que se satisface la Hipótesis 2 (b), es decir,

$$E[W(L(F(x,a),\xi)] = W(F(x,a)) + E[h(\xi)],$$

con $E[h(\xi)] < \infty$ y $(x, a) \in$ IK.

Consideremos la siguiente familia funcional definida en términos del valor óptimo determinista W,

$$\mathcal{L}_1 = \{W(x) + \lambda: \lambda \in \Re \text{ y } x \in X\}.$$

Definición 5. Sea $\Re \subseteq M(X)^+$. Diremos que $\Re$ es invariante con respecto al operador de programación dinámica, si para cada $U \in \Re$ se tiene que $TU \in \Re$.

Lema 3. La familia $\mathcal{L}_1$ es invariante con respecto al operador de programación dinámica.

Demostración. Sea $U \in \mathcal{L}_1$ entonces $U(x) = W(x) + \lambda$, con $\lambda \in \Re$ y $x \in X$. Aplicando el operador de programación dinámica,

$$\begin{aligned} TU(x) &= \min_{A(x)} \{c(x,a) + \alpha E[W(L(F(x,a),\xi)) + \lambda]\}, \\ &= \min_{A(x)} \{c(x,a) + \alpha(W(F(x,a)) + E[h(\xi)] + \lambda)\}, \end{aligned}$$

$$= \min_{A(x)}\{c(x,a) + \alpha W(F(x,a))\} + \alpha(E[h(\xi)] + \lambda),$$
$$= W(x) + \alpha(E[h(\xi)] + \lambda).$$

Si hacemos

$$\lambda = \frac{\alpha}{1-\alpha} E[h(\xi)],$$

Se tiene

$$TU(x) = W(x) + \alpha\left[\frac{\alpha E[h(\xi)] + (1-\alpha)E[h(\xi)]}{1-\alpha}\right].$$

Por lo tanto $TU(x)=W(x)+\lambda \in \mathcal{L}_1$ para cada $x \in X$. □

Nota: En este caso no solo se probó que $\mathcal{L}_1$ es invariante con respecto al operador T, sino que T tiene un punto fijo en $\mathcal{L}_1$.

Teorema 3. Bajo las Hipótesis 1 y 2 la función de valores óptimos V se encuentra en $\mathcal{L}_1$.

Demostración. Por el Lema 3 se tiene que $TU=U$ donde U esta dado por

$$U(x) = W(x) + \frac{\alpha}{1-\alpha} E[h(\xi)],$$

$x \in X$.

Por el Lema 2 solo será necesario probar que se satisface la condición de transversalidad (11),

$$\lim_{n\to\infty} \alpha^n E_x^{\pi}(U(x_n)) = \lim_{n\to\infty} \alpha^n E[U(L(F(x_{n-1}, a_{n-1}), \xi_{n-1}))],$$
$$= \lim_{n\to\infty} \alpha^n W(F(x_{n-1}, a_{n-1})) + \alpha^n E[h(\xi)],$$
$$= \lim_{n\to\infty} \alpha^n W(F(x_{n-1}, a_{n-1})) + \lim_{n\to\infty} \alpha^n E[h(\xi)] = 0.$$

La última igualdad se sigue por la condición de transversalidad del problema determinista. Por lo tanto, utilizando el Lema 2 concluimos que $V=U$. □

Ahora analizaremos el caso que satisface la Hipótesis 2 (b) en su versión multiplicativa,

$$E[W(L(F(x,a),\xi)] = W(F(x,a)) \cdot E[h(\xi)],$$

con $E[h(\xi)]<\infty$ y $(x, a) \in \text{IK}$.

Para este caso consideramos la siguiente familia funcional, inducida por la función de valor determinista,

$$\mathcal{L}_2 = \{ W(x)\lambda : \lambda \in \Re \text{ y } x \in X \}.$$

Teorema 4. Bajo las Hipótesis 1 y 2 la función de valores óptimos es un elemento de $\mathcal{L}_2$.

Demostración. La demostración la haremos por casos.
CASO 1. Supongamos que $E[h(\xi)] \geq 1$.Escojamos un valor λ tal que $1-\lambda \geq 0$ y $U \in \mathcal{L}_2$ (i.e. $U(x)=\lambda W(x)$). Entonces,

$$U(x) = \lambda W(x) = \lambda \inf_{A(x)}\left[c(x,a) + \alpha W(F(x,a))\right],$$
$$\leq \lambda c(x,a) + \alpha\lambda E[h(\xi)]W(F(x,a)),$$
$$= \left[c(x,a) + \alpha\lambda E[h(\xi)]W(F(x,a))\right] + (\lambda - 1)c(x,a),$$
$$\leq c(x,a) + \alpha\lambda E[h(\xi)]W(F(x,a)),$$

$(x, a) \in \text{IK}$. Entonces,

$$U(x) \leq \inf_{a \in A(x)}\{c(x,a) + \alpha\lambda E[h(\xi)]W(F(x,a))\}.$$

Lo cual implica que, $U \leq TU$. Por lo tanto, del Lema 2 concluimos que $U \leq V$.

Para probar la desigualdad inversa procedemos por contradicción. Supongamos que para todo $\lambda \in \Re$, existe x_λ tal que,

$$\lambda W(x_\lambda) < T(\lambda W)(x_\lambda).$$

Entonces,

$$\lambda W(x_\lambda) < T(\lambda W)(x_\lambda) \leq$$
$$\leq c(x_\lambda, g) + \alpha\lambda E[h(\xi)]W(F(x_\lambda, g)).$$

Es decir,

$$\lambda W(x_\lambda) < c(x_\lambda, g) + \alpha\lambda E[h(\xi)]W(F(x_\lambda, g)), \qquad (13)$$

Donde g es la política óptima del problema determinista.
Por otro lado

$$\lambda c(x_\lambda, g) + \alpha\lambda W(F(x_\lambda, g)) = \lambda W(x_\lambda). \qquad (14)$$

Utilizando (13) y (14), obtenemos

$$(\lambda - 1)c(x_\lambda, g) < \alpha\lambda(1 - E[h(\xi)])W(F(x_\lambda, g)) \leq 0.$$

Entonces,

$$(\lambda - 1)c(x_\lambda, g) < 0.$$

Como λ es arbitrario escojamos $\lambda - 1 > 0$. Entonces $c(x_\lambda, g) < 0$, lo cual es una contradicción ya que el costo es no negativo.

Ahora, probaremos la condición de transversalidad (11),

$$\lim_{n\to\infty} \alpha^n E[U(x_n)] = \lim_{n\to\infty} \alpha^n E[h(\xi)]W(F(x_{n-1}, a_{n-1})),$$
$$= E[h(\xi)] \lim_{n\to\infty} \alpha^n W(x_n) = 0.$$

Por el Lema 2, tenemos que $V=U$.

Nota: Observemos que para probar la condición de transversalidad no se utilizo la condición $E[h(\xi)] \geq 1$.

CASO 2. Sea $E[h(\xi)] \leq 1$, probaremos que $V \in \mathcal{L}_2$, es decir, existe $\lambda \in \Re$ tal que $V=\lambda W$.

Para la primera parte escogemos un valor λ tal que $1-\lambda E[h(\xi)] \leq 0$ y tomemos $U \in \mathcal{L}_2$. Entonces

$$TU(x) = \inf_{A(x)}\{c(x,a) + \alpha E[U(L(F(x,a),\xi))]\},$$
$$= \inf_{A(x)}\{c(x,a) + \alpha\lambda E[h(\xi)]W(F(x,a))\},$$
$$\leq c(x,g) + \alpha\lambda E[h(\xi)]W(F(x,g)),$$

$$= \lambda E[h(\xi)][c(x,g) + \alpha W(F(x,g))]$$
$$+ (1 - \lambda E[h(\xi)])c(x,g),$$
$$\leq \lambda E[h(\xi)]W(x) \leq \lambda W(x) = U(x).$$

Por el Lema 2, $V \leq U$. Para probar la desigualdad inversa, procedemos por contradicción. Supongamos que para todo $\lambda \in \Re$, $\exists\, x_\lambda \in X$ tal que $T(\lambda W)(x_\lambda) < \lambda W(x_\lambda)$. Entonces

$$TU(x_\lambda) = \inf_{A(x_\lambda)} \left[c(x_\lambda, a) + \alpha \lambda E[h(\xi)]W(F(x_\lambda, a))\right],$$
$$\geq \inf_{a \in A(x)} \left[c(x_\lambda, a) + \alpha W(F(x_\lambda, a))\right] +$$
$$+ \inf_{a \in A(x)} \alpha(\lambda E[h(\xi)] - 1)W(F(x_\lambda, a)),$$
$$= W(x_\lambda) + \inf_{a \in A(x)} \alpha(\lambda E[h(\xi)] - 1)W(F(x_\lambda, a)).$$

Si suponemos que $\lambda E[h(\xi)] - 1 \geq 0$ tenemos,

$$W(x_\lambda) \leq TU(x_\lambda).$$

Como λ es arbitrario tomemos 1>λ. Entonces por la hipótesis de contradicción

$$0 > W(x_\lambda).$$

Por lo tanto existe $\lambda \in \Re$ tal que $T(\lambda W)(x) > \lambda W(x)$ para cada $x \in X$. Así, $TU \geq U$ y por la condición de transversalidad probada en el caso anterior, concluimos que $V = U$. □

Ejemplo
En esta parte retomamos el ejemplo trabajado en la segunda sección. Para ellos verificaremos únicamente que se satisfacen las Hipótesis 2.

Regulador Lineal versión estocástica (ver [2])
Consideraremos el mismo modelo que el dado en la segunda sección, solo que ahora perturbamos la dinámica de la forma siguiente,

$$x_{t+1} = \gamma x_t + \beta a_t + \xi_t,\ t = 0,1,2,\ldots$$

Las variables $\xi_t \in \Re$ son variables aleatorias i.i.d. con media cero y varianza finita.
Comprobaremos el inciso b) de la Hipótesis 2, para $(x, a) \in$ IK,

$$E[W(\gamma x + \beta a + \xi)] = E[Q(\gamma x + \beta a + \xi)^2],$$
$$= Q(\gamma x + \beta a)^2 + QE[\xi(\gamma x + \beta a) + \xi^2].$$

Como la medio es cero, se concluye

$$E[W(\gamma x + \beta a + \xi)] = Q(\gamma x + \beta a)^2 + QE[\xi^2],$$

$(x, a) \in$ IK.

Se obtiene, para $(x, a) \in$ IK

$$E[W(\gamma x + \beta a + \xi)] = Q(\gamma x + \beta a)^2 + E(h(\xi)),$$

donde $h(u) = Qu^2$. Por lo tanto, la función de valor V de acuerdo al Teorema 3 es,

$$V(x) = Qx^2 + \frac{\alpha}{1-\alpha} Q\sigma^2,$$

$x \in X$.

La política óptima se obtiene sustituyendo el valor óptimo V en la Ecuación de Programación Dinámica,

$$f(x) = -\frac{\alpha\beta\gamma Q}{\gamma + \alpha\beta^2 Q} x,$$

$x \in X$.

5. CONCLUSIONES Y PROBLEMAS ABIERTOS

En este artículo, se presenta una forma de dar solución a problemas de control estocástico. La metodología fue considerar cierto problema de control determinista asociado al problema original. Debido a que el problema determinista, en principio, es más sencillo de resolver o por lo menos contamos con más herramienta para encontrar su solución. Es de importancia señalar que la función de valor del problema estocástico hereda las propiedades cualitativas (monotonicidad, convexidad, continuidad, diferenciabilidad,…) de la función de valor determinista.

En el presente artículo, se supone que la política y la función de valor tienen derivadas hasta de orden p ($p \geq 1$) alrededor del punto de equilibrio, dicha condición es necesaria justificarla. También la Hipótesis 2 (b) sería interesante dar condiciones sobre el modelo que permitan concluir dicha condición.

6. REFERENCIAS

[1] Arkin, V. I. &Evstigneev I. V., **Stochastic Models of Control and Economic Dynamics**, Academic Press Inc., 1987.

[2] Bertsekas, D. P., **Dynamic Programming and Optimal Control**, Athena Scientific, Belmont Massachusetts, 1995.

[3] Blume, L., Easley D. & O'Hara M., "Characterization of optimal plans for stochastic dynamic programs", **Journal of Economic Theory**, Vol. 28, 1982, pp. 221-234,

[4] Dynkin, E. B., Yushkevich, A. A., **Controlled Markov Processes**, Springer Verlag, New York, 1979.

[5] Hernández-Lerma, O. & Lasserre J. B., **Discrete Time Markov Control Processes**, Springer Verlag, 1996.

[6] Levhari, D. & Srinivasan, T. N., "Optimal savings under uncertainty", **Review of Economic Studies**, 1969, pp. 153-164.

[7] Pontryagin, L. S., Boltyanskii V. G., Gamkrelidze R. V. & Mishchenko E. F., **The Mathematical Theory of Optimal Processes**, Interscience Publishers, 1962.

[8] Puterman, M. L., **Markov Decision Processes: Discrete Stochastic Dynamic Programming**, Wiley, New York, 1994.

[9] Stokey, N. L. & Lucas, R. E., **Recursive Methods in Economic Dynamics**, Harvard University Press, 1989.