El Principio del Máximo en Procesos de Control de Markov con Horizonte Finito

Lilia González de la Palma

Julio 2008

Índice general

Agradecimientos

Porque tú eres grande, y hacedor de maravillas;
sólo tú eres DIOS.
Salmos 86:10

Principalmente a DIOS, quien es mi fortaleza y mi escudo, en Él confió mi corazón y fui ayudada. Mis ojos han visto su respaldo y fidelidad, muestras de su perfecto y eterno amor hacía mí. Gracias a Él puedo caminar segura, pues está conmigo, su providencia, su bondad y su gracia son sobre mi vida.

A mis PADRES Y HERMANOS por los ánimos que me dan continuamente, siempre me han mostrado su confianza, cada esfuerzo realizado a favor mío es apreciado en gran manera, ustedes son valiosos, los respeto y los amo.

Al Dr. Hugo Adán Cruz Suárez por su paciencia y ayuda durante éstos dos años de aprendizaje, gracias por su respaldo y amistad.

A mis maestros quienes colaboraron a mi formación académica durante mi estancia en la Facultad de Ciencias Físico Matemáticas de la Benemérita Universidad Autónoma de Puebla.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico otorgado para la realización de esta maestría.

A mis amigas y amigos por su exhortación, comprensión y paciencia. No hay ayuda insignificante, recuerdo y valoro a cada uno de ustedes.

Capítulo 1

Introducción

El presente trabajo se encuentra relacionado con la teoría de Procesos de Control de Markov (PCMs) a tiempo discreto con horizonte finito. Los PCMs son una clase de problemas de optimización estocástica, los cuales son utilizados en diversas áreas como: ingeniería, economía, investigación de operaciones, entre otras. El problema de control óptimo con el que se trabajará consiste de un sistema dinámico el cual es observado en instantes de tiempo discreto $t = 0, 1, 2, ..., N$. En dicho sistema se busca influenciar o regular su comportamiento de forma óptima, mediante la elección de ciertas variables del sistema llamadas controles (o acciones). Es decir, supóngase que al momento t el sistema se encuentra en un estado $x_t = x$, entonces, al aplicar un control $a_t = a$ suceden dos cosas, mediante una ley de transición el sistema se traslada a un nuevo estado x_{t+1} en el instante $t + 1$ y, se obtiene una recompensa por haber aplicado el control a_t. Al encontrarse en el estado x_{t+1} se repite el procedimiento descrito. De esta manera se obtiene una sucesión de controles $\{a_t\}$ con $t = 0, 1, 2, ..., N - 1$, llamada política. Para medir la calidad de los controles aplicados al sistema se tiene un criterio de rendimiento o función objetivo. En este trabajo se considera las funciones objetivo conocidas como recompensa total y recompensa total descontada.

Así, el problema de control de Markov consiste en encontrar una política que optimice el criterio de rendimiento. La política que optimiza la función objetivo se le llama política óptima, y al criterio de rendimiento evaluado en la política óptima se le conoce como la función de valor óptimo. Aquí, N es referido como el horizonte del proceso, cuando N es finito se dice que el proceso es a tiempo discreto con horizonte finito, o bien que el proceso es a tiempo discreto con horizonte infinito, si N es infinito.

Esta tesis está enfocada a analizar el problema de control óptimo vía el Principio del Máximo (PM). El PM ha sido trabajado en distintos contextos tanto a tiempo discreto como a tiempo continuo (véase Araujo & Scheinkman [2], Arkin & Evstigneev [4], Arkin & Krechetov [5], Arutyunov [6], Basile & Mininni [8], Burmeister & Dobell [13], Cadenillas & Karatzas [14], Canon et. al. [15], [16], Clarke & Vinter [19], Fan & Wang [23], Fattorini [24], Fleming & Rishel [25], Halkin [26], Hartl et. al. [29], Haussmann [30], Holtzman [32], [33], Intriligator [34], Medhin [35], Peng [36], Pham [37], Shell [43], Tang & Hou [44], Urzula & Heinz [45], Xu [47] y Yong & Xu [48]). El PM es un método de optimización debido en gran parte al trabajo del matemático ruso Lev Semenovich Pontryagin y sus colaboradores (véase Pontryagin et. al. [38]), el cual generaliza la ecuación de Euler-Lagrange del cálculo de variaciones. El problema que trabajó Pontryagin et. al. fue para un proceso a tiempo continuo y se centró en maximizar una función objetivo cuya dinámica está regida por una ecuación diferencial. Actualmente existen diversas versiones del PM para problemas de control determinista a tiempo continuo dependiendo de las características del problema (véase Arutyunov [6]). En contraste, en la literatura revisada para problemas de control estocástico a tiempo discreto no se encontró una versión del PM que determinará de forma explícita la solución del problema de control óptimo. Por esta razón uno de los propósitos de este trabajo es presentar una versión del PM, el cual se pueda utilizar para resolver de forma explícita algunos problemas de la teoría de PCMs.

Inicialmente se estudia una versión del PM para problemas de control determinista a tiempo discreto basado en las ideas del trabajo hecho por Canon et. al. véase [15], [16]. La idea de la prueba de este principio consiste en trasladar el problema de control óptimo a un problema de optimización general. Posteriormente, se presenta una versión del PM en el contexto de PCMs. Dicho principio es obtenido, en esta tesis, a partir de la Ecuación de Programación Dinámica. Esta idea ha sido trabajada para procesos de control a tiempo continuo por Arkin [4], [5], Clarke [19], Shimomura [46], entre otros. En la prueba que se da en esta tesis se hace uso de la diferenciabilidad de la función de valor, esta propiedad ha sido estudiada por: Araujo [2], Cruz-Suárez & Montes-de-Oca [20], [21], Blume [12]. La diferenciabilidad de la función de valor óptimo no sólo es importante como una propiedad cualitativa del modelo sino también porque hace posible el estudio de la estabilidad del modelo (véase Araujo [3]), permite una caracterización de la función de valor óptimo y también ha sido utilizada en procedimientos de linealización (véase Santos [42]). El PM que se presenta en este trabajo es utilizado para resolver algunos problemas de la teoría de PCMs, los cuales en la bibliografía existente no se encuentra su solución por dicho principio.

La tesis está organizada de la siguiente manera, el capítulo 2 presenta la teoría referente a procesos de control de Markov, el capitulo 3 está dedicado al Principio del Máximo determinista y estocástico, en el capítulo 4 se presentan ejemplos resueltos mediante el PM. Los ejemplos están clasificados de manera que los dos primeros están relacionados con problemas lineales cuadráticos (LQ) y los siguientes están relacionados con problemas económicos. Finalmente, en el capítulo 5 se encuentran las conclusiones obtenidas del trabajo realizado.

Capítulo 2

Procesos de Control de Markov

Este capítulo provee la teoría necesaria para la comprensión de los capítulos siguientes. Una vez establecidas las componentes del Modelo de Control y especificado el problema de control, se describe el método de solución de Programación Dinámica. La idea de este procedimiento es llevar el problema de control óptimo a un problema equivalente, el cual consiste en resolver una ecuación funcional para la función de valores óptimos, conocida como Ecuación de Programación Dinámica (EPD) (véase Hernández-Lerma [31]).

Cabe mencionar que los resultados aquí presentados son únicamente los relevantes para el desarrollo de los capítulos subsecuentes. Para un estudio detallado acerca de los Procesos de Control de Markov véase: Bertsekas & Shreve [11], Hernández-Lerma [31] y Puterman [39].

2.1. Procesos de Control de Markov

Definición 2.1.1 *Un Modelo de Control de Markov (MCM) consiste de:*

a) *X un espacio de Borel, llamado espacio de estados.*

b) *A un espacio de Borel, llamado espacio de acciones o controles.*

c) $\{A(x) \mid x \in X\}$ *una familia de subconjuntos medibles,* $A(x) \subseteq A$, *donde* $x \to A(x)$ *para cada* $x \in X$, *los elementos de estos subconjuntos son los controles o acciones admisibles cuando el sistema se encuentra en el estado* $x \in X$.

d) Q *un kérnel estocástico(véase Apéndice, Definición A.1.1) definido en* X *dado*

$$\mathbb{K} := \{(x, a) \mid x \in X, \ a \in A(x)\},$$

con $\mathbb{K} \subseteq X \times A$, *el conjunto medible de parejas estado-acción admisibles.* Q *es llamado ley de transición.*

e) $r : \mathbb{K} \to \mathbb{R}$, *una función medible, llamada función de recompensa en un paso.*

En varias aplicaciones la ley de transición Q es inducida por una ecuación de la forma siguiente:

$$x_{t+1} = F(x_t, a_t, \xi_t), \qquad t = 0, 1, 2, ... \tag{2.1}$$

donde $F : X \times A \times S \to X$ es una función medible conocida, con x_0 dado, $\{\xi_t\}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas (v.a.i.i.d) tomando valores en un espacio de Borel $S \subseteq \mathbb{R}^k$ $(k \geq 1)$ las cuales son independientes de x_0, y si ξ denota un elemento genérico de las $\{\xi_t\}$ entonces se denotará a la distribución de ξ por μ y la esperanza con respecto a μ por E. En este caso la ley de transición Q está dado por:

$$\begin{aligned} Q(B \quad | \quad & x, a) = \mu(\{s \in S \mid F(x, a, s) \in B\}) \\ = \quad & \int_S I_B \left[F\left(x, a, s\right)\right] \mu\left(ds\right) \\ = \quad & EI_B \left[F\left(x, a, \xi\right)\right], \end{aligned}$$

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

Observación 2.1.2 *Supóngase que existe una función medible* $f : X \to A$ *tal que* $f(x) \in A(x)$ *para todo* $x \in X$*, a esta familia de funciones se le denotará por* $\mathbb{F}$ *y sus elementos son llamados selectores de la multifunción* $x \to A(x)$*. Este supuesto indica que* $\mathbb{K}$ *contiene la gráfica de una función medible de* X *en* A*.*

Condiciones suficientes para el supuesto anterior están dados por los teoremas de selección medibles (véase Hernández-Lerma [31]), algunas de las cuales se presentarán en la Sección 2.2.2.

Una de las razones por la que recibe el nombre de modelo Markov es que la probabilidad de transición y la función de recompensa dependen únicamente del estado presente del sistema y de la acción que seleccione el controlador en ese estado.

2.1.1. Políticas de Control

Considere un Modelo de Control de Markov como en la Definición 2.1.1, y defínase para cada $t = 0, 1, 2, ...$ el espacio $\mathbb{H}_t$ de *historias admisibles* del proceso de control hasta el tiempo t, como

$$\begin{aligned} \mathbb{H}_0 &= X, \\ \mathbb{H}_t &= \mathbb{K} \times \mathbb{H}_{t-1}. \end{aligned}$$

Un elemento h_t de $\mathbb{H}_t$ llamada $t-historia$ es un vector de la forma

$$(x_0, a_0, x_1, a_1, ...x_{t-1}, a_{t-1}, x_t),$$

donde $(x_i, a_i) \in \mathbb{K}$ para $i = 0, 1, 2, ..., t-1$ y $x_t \in X$. Nótese que, para cada t, $\mathbb{H}_t$ es un subespacio de

$$\begin{aligned} \overline{\mathbb{H}}_t &= (X \times A)^t \times X = (X \times A) \times \overline{\mathbb{H}}_{t-1}; \quad \text{para } t = 1, 2, ... \\ \overline{\mathbb{H}}_0 &= \mathbb{H}_0 = X. \end{aligned}$$

Definición 2.1.3 *Una política aleatorizada o simplemente política, es una sucesión* $\pi = \{\pi_t, t = 0, 1, 2, ..., \}$ *de kérneles estocásticos definidos sobre* A *dada la historia del proceso* $\mathbb{H}_t$ *tal que*

$$\pi_t(A(x_t) \mid h_t) = 1 \quad \textit{para todo } h_t \in \mathbb{H}_t, \qquad t = 0, 1, 2, ...$$

Al conjunto de todas las políticas se denotará por Π.

De acuerdo con esta definición, una política $\pi = \{\pi_t\}$ puede interpretarse como una sucesión $\{a_t\}$ de variables aleatorias sobre A tales que para cada $t-historia$ y $t = 0, 1, 2, ...$, la distribución de a_t es $\pi_t(\cdot \mid h_t)$, la cual está concentrada en el conjunto de acciones admisibles $A(x_t)$, en otras palabras, cuando usamos una política arbitraria, la acción en cualquier tiempo t es una variable aleatoria y depende de todas las $t - historias$.

Denótese a la familia de kérneles estocásticos sobre A dado X, como $\mathcal{P}(A \mid X)$ (véase Apéndice A.1.).

Sea Φ el conjunto de todos los kérneles estocásticos φ en $\mathcal{P}(A \mid X)$ tales que para toda $x \in X$ se tiene $\varphi(A(x) \mid x) = 1$.

Observación 2.1.4 *Por la Observación 2.1.2 se tiene que* $\mathbb{F} \subset \Phi$.

Definición 2.1.5 *Una política* $\pi \in \Pi$ *es:*

Determinista Markoviana (Π_{DM})*. Si existe una sucesión* $\{f_t\}$ *de funciones medibles* $f_t : X \to A$ *(o* $f_t \in \mathbb{F}$*), tales que* $f_t(x_t) \in A(x_t)$ *y* $\pi_t(\cdot \mid h_t)$ *está concentrada en* $f_t(x_t)$ *para cada* $h_t \in \mathbb{H}_t$ *y* $t = 0, 1, 2, ...$

Determinista Markoviana Estacionaria (Π_{DS})*. Si existe una función medible* $f : X \to A$ *(o* $f \in \mathbb{F}$*), tal que* $f(x_t) \in A(x_t)$ *y* $\pi_t(\cdot \mid h_t)$ *está concentrada en* $f(x_t)$ *para cada* $h_t \in \mathbb{H}_t$ *y* $t = 0, 1, 2, ...$

Se dice que $\pi(\cdot \mid h)$ está concentrada en $g(h)$, si, $\pi(C \mid h) = I_C(g(h))$ para cada $C \in \mathcal{B}(A)$. Donde I_C es la función indicadora del conjunto C.

Observación 2.1.6 *Nótese que,* $\Pi_{DS} \subset \Pi_{DM} \subset \Pi$.

Construcción Canónica

Sea $(\Omega, \mathcal{F})$ un espacio medible que consiste del espacio muestral canónico $\Omega := \overline{H}_\infty = (X \times A)^\infty$ y $\mathcal{F}$ su correspondiente σ-álgebra producto. Los elementos de Ω son de la forma $w = (x_0, a_0, x_1, a_1, ...)$ con $x_t \in X$ y $a_t \in A$ para toda $t = 0, 1, 2, ...,$

las proyecciones x_t y a_t de Ω sobre X y A son llamados estado y acción, respectivamente.

Obsérvese que $H_\infty = \mathbb{K}^\infty \subset \Omega$ es el conjunto de parejas estado acción admisible. Sean $\pi \in \Pi$ una política arbitraria y $x_0 = x \in X$. Entonces por el Teorema de Ionescu-Tulcea (véase Apéndice, Proposición A.1.7), existe una única medida de probabilidad P_x^π sobre $(\Omega, \mathcal{F})$. Además, para cada $C \in \mathcal{B}(A)$, $B \in \mathcal{B}(X)$, $h_t \in \mathbb{H}_t$ y $t = 0, 1, 2, ...$, se tiene que

$$P_x^\pi(a_t \in C \mid h_t) = \pi_t(C \mid h_t), \tag{2.2}$$

$$P_x^\pi(x_{t+1} \in B \mid h_t, a_t) = Q(B \mid x_t, a_t). \tag{2.3}$$

El proceso estocástico $((\Omega, \mathcal{F}, P_x^\pi), \{x_t\})$ es llamado un *Proceso de Control de Markov a tiempo discreto* o *Proceso de Decisión de Markov (PDM).*

La esperanza con respecto a P_x^π será denotada por E_x^π.

Observación 2.1.7 *En general, en lugar de dar $x_0 = x \in X$, se puede dar una medida de probabilidad ν sobre X, referida como distribución inicial y se cumple que*

$$P_v^\pi(x_0 \in B) = \nu(B),$$

para cada $B \in \mathcal{B}(X)$.

El MCM descrito aquí, es llamado estacionario porque sus componentes no dependen del parámetro tiempo t, en caso de que esto suceda se conoce como modelo no estacionario, es decir, un modelo de la forma $(X_t, A_t, \{A_t(x) \mid x \in X_t\}, Q_t, C_t)$ para $t = 0, 1, 2, ...$

La Observación 2.1.4 asegura que $\mathbb{F}$ es no vacío por lo que Π tampoco lo es. Esto se debe a que cada $f(x) \in \mathbb{F}$ puede ser identificada por un kérnel estocástico φ, de la siguiente forma: $\varphi(C \mid x) = I_C(f(x))$ para toda $C \in \mathcal{B}(A)$ y $x \in X$.

2.1.2. La Propiedad de Markov

En este apartado se mostrará que al usar una política de Markov en un proceso de Markov a tiempo discreto el resultado es un Proceso de Control de Markov. Defínase para $x \in X$ y $\varphi \in \Phi$

$$r(x, \varphi) := \int_A r(x, a)\varphi(da \mid x),$$

$$Q(\cdot \mid x, \varphi) := \int_A Q(\cdot \mid x, a)\varphi(da \mid x).$$

En particular, si $f \in \mathbb{F}$,

$$r(x, f) := r(x, f(x)),$$

$$Q(\cdot \mid x, f) := Q(\cdot \mid x, f(x)).$$

Definición 2.1.8 *Sea $\{Z_t\}$ una sucesión de kérneles estocásticos en $P(X \mid X)$ (véase Apéndice, Sección A.1.), y sea $\{x_t\}$ un proceso estocástico con valores en un espacio X. Se dice que $\{x_t\}$ es un proceso de Markov no homogéneo con kérneles de transición $\{Z_t\}$, si para cada $B \in \mathcal{B}(X)$ y $t = 0, 1, 2, ...$, se tiene que*

$$P(x_{t+1} \in B \mid x_0, x_1, ...x_t) = P(x_{t+1} \in B \mid x_t) = Z_t(B \mid x_t),$$

con P la medida de probabilidad sobre el espacio en el cual $\{x_t\}$ está definida. A la igualdad anterior se le conoce como la propiedad de Markov.

Si $\{Z_t\}$ es invariante en el tiempo, es decir, Z_t es igual a Z para toda $t = 0, 1, 2, ...$, con $Z \in P(X \mid X)$. Entonces $\{x_t\}$ se llama proceso de Markov homogéneo con kérnel de transición Z.

La prueba de la siguiente proposición se puede consultar en Hernández-Lerma [31], pp. 19-20.

Proposición 2.1.9 *Sea v una distribución inicial arbitraria y $\pi = \{\varphi_t\} \in \Pi_{RM}$, entonces $\{x_t\}$ es un proceso de Markov no* *homogéneo con kérneles de transición* $\{Q(\cdot \mid \cdot, \varphi_t)\}$*, esto es, para cada* $B \in \mathcal{B}(X)$ *y* $t = 0, 1, 2, ...,$

$$P_v^\pi(x_{t+1} \in B \mid x_0, x_1, ...x_t) = P_v^\pi(x_{t+1} \in B \mid x_t) = Q(B \mid x_t, \varphi_t).$$

En particular, si $\pi = \{f_t\} \in \Pi_{DM}$*, los kérneles de transición son* $Q(\cdot \mid \cdot, f_t)$*. Además, para políticas estacionarias* $f \in \Pi_{DS}$*, el proceso es de Markov homogéneo con kérnel de transición y* $Q(\cdot \mid \cdot, f)$*.*

Observación 2.1.10 *Sea* $Q(\cdot \mid \cdot, \varphi)$ *un kérnel de transición como en la proposición anterior, con* $\varphi \in \Phi$*. Se denotará a las probabilidades de transición en* n*-etapas como* $Q^n(\cdot \mid \cdot, \varphi)$*, donde*

$$\begin{aligned} Q^n(B \mid x, \varphi) &= \int Q(B \mid x, \varphi) Q^{n-1}(dy \mid x, \varphi), \\ &= \int Q^{n-1}(B \mid x, \varphi) Q(dy \mid x, \varphi), \end{aligned}$$

para $n \geq 1$*.*

2.1.3. Criterio de Rendimiento

Cada PCM estará dotado de una función llamada función objetivo o criterio de rendimiento, la cual medirá en algún sentido la calidad de cada política.

Considérese un modelo de control de Markov con horizonte de planeación $N < +\infty$ y un conjunto de políticas Π. Defínase para cada $x \in X$ y $\pi \in \Pi$ a

$$v(\pi, x) := E_x^\pi \left[\sum_{t=0}^{N-1} r(x_t, a_t) + r(x_N) \right], \tag{2.4}$$

$v(\pi, x)$ es llamada **Recompensa Total Acumulada Esperada.** Otro criterio de rendimiento, es el de recompensa total esperada $\alpha - \textit{descontada}$ por unidad de tiempo, ésta se describe a continuación.

Sea $\pi \in \Pi$ y $x \in X$ se define la **Recompensa Total** α-**Descontada Esperada** como

$$v(\pi, x) := E_x^{\pi}\left[\sum_{t=0}^{N-1} \alpha^t r(x_t, a_t) + \alpha^N r\left(x_N\right)\right],$$

con $\alpha \in (0, 1)$. A α se le conoce como **factor de descuento**.

En ambos criterios de rendimiento $r\left(x_N\right)$ es una función medible definida en X tomando valores en $\mathbb{R}$, llamada recompensa terminal.

Definición 2.1.11 *Para cada $x \in X$ sea $V\left(x\right)$ la* ***función de valor óptimo,*** *donde*

$$V(x) = \sup_{\pi \in \Pi} v(\pi, x).$$

Definición 2.1.12 *Una política $\pi^* \in \Pi$, es* ***óptima****, si*

$$V(\pi^*, x) = \sup_{\pi \in \Pi} v(\pi, x),$$

$x \in X$. Es decir, si $V(\pi^*, x) = V\left(x\right)$.

2.1.4. Problema de Control Óptimo

Sea $(X, A, \{A(x) \mid x \in X\}, Q, r)$, el modelo de control de Markov, Π el conjunto de políticas y $v(\pi, x)$ un criterio de rendimiento, el **problema de control óptimo** consiste en encontrar una política óptima, es decir, maximizar la función $\pi \rightarrow v(\pi, x)$ sobre Π, para toda $x \in X$. Implícitamente se asume que el controlador desea elegir una política óptima para todos los posibles estados iniciales del sistema. En la práctica, basta con encontrar una política óptima para algún estado inicial dado.

2.2. Problemas con Horizonte Finito

Considere el criterio de rendimiento de recompensa total acumulada con horizonte finito, es decir,

$$v(\pi, x) := E_x^{\pi} \left[\sum_{t=0}^{N-1} r(x_t, a_t) + r_N(x_N) \right].$$

Una técnica de solución para el problema de control óptimo, conocido como Programación Dinámica permite encontrar tanto a la función de valor óptimo $V(x)$, como la política óptima π^*.

2.2.1. Programación Dinámica

Teorema 2.2.1 *Sean $V_0, V_1, ..., V_N$ funciones sobre X definidas por*

$$V_N(x) := r_N(x), \tag{2.5}$$

$x \in X$ y para cada $t = N-1, ..., 1, 0,$

$$V_t(x) := \max_{a \in A(x)} \left[r(x, a) + \int_X V_{t+1}(y) Q(dy \mid x, a) \right], \tag{2.6}$$

$x \in X$. Supongamos que estas funciones son medibles y que para cada $t = N-1, ..., 1, 0,$ existe un selector $f_t(x) \in \mathbb{F}$ con $f_t(x) \in A(x)$, tal que

$$V_t(x) = r(x, f_t(x)) + \int_X V_{t+1}(y) Q(dy \mid x, f_t(x)),$$

$x \in X$. Entonces, la política determinista de Markov $\pi^* = \{f_0, f_1, ..., f_{N-1}\}$ es óptima y la función de valor óptimo $V(x)$ es $V_0(x)$, es decir, para $x \in X$ tenemos que $V(x) = v(\pi^*, x) = V_0(x)$.

Observación 2.2.2 *La relación (2.6) junto con su condición inicial (2.5) es conocida como Ecuación de Programación Dinámica (EPD).*

Demostración. Sea $\pi = \{\pi_t\}$ una política arbitraria, y sea

$$R_t(\pi, x) := E^{\pi}\left[\sum_{n=t}^{N-1} r(x_n, a_n) + r_N(x_N) \mid x_t = x\right],$$

$$R_N(\pi, x) := r_N(x),$$

$t = 0, 1, 2, ..., N-1$; $R_t(\pi, x)$ es llamada la recompensa total del instante t a $N-1$ cuando se usa la política π y el sistema se encuentra en el estado $x_t = x$. En particular, véase que

$$v(\pi, x) = R_0(\pi, x).$$

Para probar este teorema se mostrará que para todo $x \in X$ y $t = 0, 1, 2, ..., N-1$, se cumple que

$$R_t(\pi, x) \leq V_t(x),$$

con igualdad cuando $\pi = \pi^*$, es decir,

$$R_t(\pi^*, x) = V_t(x).$$

En particular, si $t = 0$ se tiene que

$$R_0(\pi, x) = v(\pi, x) \leq V_0(x)$$

y si utilizamos a π^*:

$$R_0(\pi^*, x) = v(\pi^*, x) = V_0(x) = V(x).$$

Para probar las relaciones anteriores, obsérvese que

$$R_N(\pi, x) = V_N(x) = r_N(x),$$

$x \in X$. Ahora, supóngase que para alguna $t \in \{0, 1, 2, ..., N-1\}$ se tiene que

$$R_{t+1}(\pi, x) \leq V_{t+1}(x), \ \ x \in X,$$

entonces por (2.2) y (2.3),

$$\begin{aligned}
R_t(\pi, x) &= E^{\pi}\left[\sum_{n=t}^{N-1} r(x_n, a_n) + r_N(x_N) \mid x_t = x\right] \\
&= E^{\pi}\left[r(x_t, a_t) + \sum_{n=t+1}^{N-1} r(x_n, a_n) + r_N(x_N) \mid x_t = x\right] \\
&= E^{\pi}\left[r(x_t, a_t) \mid x_t = x\right] + E^{\pi}\left[\sum_{n=t+1}^{N-1} r(x_n, a_n) + r_N(x_N) \mid x_t = x\right] \\
&= \int_A r(x, a)\pi(da \mid x) + E^{\pi}\left[\sum_{n=t+1}^{N-1} r(x_n, a_n) + r_N(x_N) \mid x_t = x\right],
\end{aligned}$$

por propiedades de esperanza condicional se llega a que

$$E^{\pi}\left[\sum_{n=t+1}^{N-1} r(x_n, a_n) + r_N(x_N) \mid x_t = x\right] =$$

$$E^{\pi}\left[E^{\pi}\left(\sum_{n=t+1}^{N-1} r(x_n, a_n) + r_N(x_N) \mid x_{t+1} = y\right) \mid x_t = x\right].$$

Por otro lado, obsérvese que

$$E^{\pi}\left[\sum_{n=t+1}^{N-1} r(x_n, a_n) + r_N(x_N) \mid x_{t+1} = y\right] = R_{t+1}(\pi, y).$$

Así,

$$\begin{aligned}
R_t(\pi, x) &= \int_A \left[r(x, a) + \int_X R_{t+1}(\pi, y) Q(dy \mid x, a)\right] \pi(da \mid x) \\
&\leq \int_A \left[r(x, a) + \int_X V_{t+1}(y) Q(dy \mid x, a)\right] \pi(da \mid x) \\
&\leq \sup_{a \in A(x)} \left[r(x, a) + \int_X V_{t+1}(y) Q(dy \mid x, a)\right] = V_t(x),
\end{aligned}$$

con igualdad cuando $\pi = \pi^*$ pues, por hipótesis de inducción se tiene que

$$R_{t+1}(\pi^*, x) = V_{t+1}(x).$$

■

De lo anterior se muestra que $V_t(x)$ es el óptimo del problema desde el tiempo t a N, es decir,

$$V_t(x) = \sup_{\pi \in \Pi} R_t(\pi, x),$$

para toda $x \in X$ y $t = 0, 1, 2, ..., N$.

Así, se ha calculado para cada tiempo t la recompensa óptima de t en adelante, con esta interpretación de $V_t(x)$ es posible caracterizar a la ecuación de programación dinámica.

En efecto, para probar la EPD, sea $\pi = \{\pi_t, ..., \pi_{N-1}\}$ un política tal que $\pi_t = f \in \mathbb{F}$ es un selector arbitrario y $\{\pi_{t+1}, ..., \pi_{N-1}\}$ es una política óptima para el problema de $t+1$ hasta N, entonces

$$\begin{aligned} R_t(\pi, x) &= r(x, f) + \int V_{t+1}(y) Q(dy \mid x, f) \\ &\leq \sup_{a \in A(x)} \left[r(x, a) + \int_X V_{t+1}(y) Q(dy \mid x, a) \right], \end{aligned}$$

para todo $x \in X$. De aquí, tenemos que

$$V_t(x) \leq \sup_{a \in A(x)} \left[r(x, a) + \int_X V_{t+1}(y) Q(dy \mid x, a) \right],$$

para ver la desigualdad inversa nótese que

$$V_t(x) \geq r(x, f) + \int V_{t+1}(y) Q(dy \mid x, f),$$

lo cual implica que

$$V_t(x) \geq \sup_{a\in A(x)} \left[r(x,a) + \int_X V_{t+1}(y)Q(dy \mid x,a) \right],$$

pues f es arbitrario, por lo tanto

$$V_t(x) := \max_{a\in A(x)} \left[r(x,a) + \int_X V_{t+1}(y)Q(dy \mid x,a) \right].$$

2.2.2. Condición de Selección Medible

El teorema de Programación Dinámica tiene como suposición la existencia de selectores $f(x) \in \mathbb{F}$, los cuales maximizan el lado derecho de la EPD en cada etapa. Este supuesto es referido como condición de selección medible.

Condición de Selección Medible

Considere un Modelo de Control de Markov y una función medible $u : X \to \mathbb{R}$ conocida. Entonces la función u^* definida para cada $x \in X$ como

$$u^*(x) := \sup_{a\in A(x)} \left\{ r(x,a) + \int_X u(y)Q(dy \mid x,a) \right\}, \tag{2.7}$$

es medible y existe una función medible $f(x) \in \mathbb{F}$ tal que la función entre llaves alcanza su máximo en $f(x) \in A(x)$ para toda $x \in X$, es decir,

$$u^*(x) = r(x, f(x)) + \int_X u(y)Q(dy \mid x, f(x)).$$

En conclusión, si este supuesto ocurre entonces se puede cambiar supremo por máximo en (2.7).

Definición 2.2.3 *Sean (X,d) un espacio métrico y $u : X \to \mathbb{R} \cup \{+\infty\}$ una función tal que $u(x) < \infty$ para al menos una*

$x \in X$. La función u es semicontinua superiormente (s.c.s.) en x, si

$$\limsup_{n\to\infty} u(x_n) \leq u(x),$$

para cualquier sucesión $\{x_n\}$ en X convergente a x.

Si u es s.c.s. para toda $x \in X$, entonces u es semicontinua superiormente (s.c.s.). La función u es semicontinua inferiormente (s.c.i.) si $-u$ es semicontinua superiormente.

Obsérvese que u es continua, si y sólo si, es semicontinua superior e inferiormente.

Definición 2.2.4 *Una función $\widehat{u} : \mathbb{K} \to \mathbb{R}$ se llama sup-compacta sobre $\mathbb{K}$, si para toda $x \in X$ y $\lambda \in \mathbb{R}$, el conjunto $\{a \in A(x) \mid \widehat{u}(x,a) \geq \lambda\}$ es compacto.*

Definición 2.2.5 *Sea $w(x,a) := \int_X v(y)Q(dy \mid x,a)$.*

a) *La ley de transición Q es débilmente continua si $w(x,a)$ es continua y acotada en $\mathbb{K}$ para cada función v continua y acotada en X.*

b) *La ley de transición Q es fuertemente continua si $w(x,a)$ es continua y acotada en $\mathbb{K}$ para cada función v medible y acotada en X.*

Hipótesis I

a) La función de recompensa r es s.c.s., acotada superiormente y sup-compacta en $\mathbb{K}$.

b) La ley de transición Q es débilmente continua, ó fuertemente continua.

Hipótesis II

a) $A(x)$ es compacto para todo $x \in X$.

b) La función de recompensa $r(x, \cdot)$ es s.c.s en $A(x)$ para cada $x \in X$.

c) La función $w(x, a) := \int_X v(y)Q(dy \mid x, a)$ definida en $\mathbb{K}$, satisface alguna de las siguientes condiciones:

- **c1)** $w(x, \cdot)$ es s.c.s en $A(x)$ para cada $x \in X$ y cada función v continua y acotada en X, ó
- **c2)** $w(x, \cdot)$ es s.c.s en $A(x)$ para cada $x \in X$ y cada función v medible y acotada en X.

Hipótesis III

a) $A(x)$ es compacto para toda $x \in X$.

b) La función de recompensa r es s.c.s y acotada superiormente.

c) La ley de transición Q es débilmente continua, ó fuertemente continua.

Teorema 2.2.6 *Bajo el bloque de Hipótesis I, II o III se tiene que, para cualquier función $u : X \to \mathbb{R}$ medible y no negativa, la condición de selección medible se cumple.*

Demostración. Véase [31], pp. 28-30. ■

Observación 2.2.7 *Un modelo de Markov que satisface el bloque de Hipótesis II ó III se le conoce como modelo semicontinuo, y si el modelo satisface la Hipótesis I se le conoce como modelo semicontinuo-semicompacto.*

Capítulo 3

Principio del Máximo

El Principio del Máximo (PM) es una técnica complementaria a Programación Dinámica que provee condiciones necesarias para determinar el control óptimo. El PM ha sido utilizado especialmente en problemas a tiempo continuo (véase Pontryagin et. al. [38]). En este capítulo se presentarán dos versiones del PM a tiempo discreto, una de ellas referente a problemas de control determinista y la otra para problemas de control estocástico.

3.1. Principio del Máximo Determinista

El Principio del Máximo fue inicialmente propuesto por Pontryagin y su grupo en la década de los 50 (véase Pontryagin et. al. [38]), dicho principio fue aplicado a problemas de control determinista a tiempo continuo. Una de las razones, como se menciona en Canon [16], para el estudio de problemas de control óptimo a tiempo discreto es debido a que los problemas de control óptimo a tiempo continuo pueden ser aproximados por estos. En cualquier cálculo llevado a cabo en un ordenador, lo mejor es obtener un conjunto finito de números reales (discretización). Sin embargo, una discretización sencilla requiere una gran capacidad de memoria y una gran cantidad de cálculos cuando son llevados a cabo en un ordenador. Además, en muchos casos debido a la naturaleza del problema es necesario observar el sistema de forma discreta. De acuerdo a lo anterior se considera de interés el estudio de problemas de control óptimo a tiempo discreto. En particular, en esta sección se presenta una técnica de solución para el problema de control óptimo a tiempo discreto conocida como Principio del Máximo.

Algunas referencias en donde se han estudiado distintas versiones del Principio del Máximo para modelos deterministas, estocásticos, a tiempo discreto y continuo son: Arkin & Evstigneev [4], Arkin & Krechetov [5], Basile & Mininni [8], Burmeister & Dobell [13], Cadenillas y Karatzas [14], Canon et. al. [15], Clark [18], Clarke & Vinter [19], Fan & Wang [23], Fattorini [24], Fleming & Rishel [25], Halkin [26], Hartl et. al. [29], Haussmann [30], Holtzman [33], Intriligator [34], Medhin [35], Peng [36], Pham [37], Shell [43], Tang & Hou [44], Urzula & Heinz [45], Xu [47] y Yong & Xu [48].

En particular, entre las referencias dadas en el párrafo anterior se hace un bosquejo del trabajo realizado por Canon et. al. ([15] y [16]).

Considere el Modelo de Control de Markov con algunos supuestos adicionales:

a) $X \subset \mathbb{R}^n$ $(n \geq 1)$, espacio de estados.

b) $A \subset \mathbb{R}^m$ $(m \geq 1)$, espacio de controles.

c) $A(x) = A$, $x \in X$.

d) La dinámica del sistema se describe por

$$x_{t+1} - x_t = \tau(x_t, a_t), \tag{3.1}$$

para $t = 0, 1, ..., N-1$ (con $N \in \mathbb{Z}^+$, $N > 1$), donde $\tau : X \times A \to X$ es una función conocida.

e) Criterio de rendimiento

$$\sum_{t=0}^{N-1} r(x_t, a_t), \tag{3.2}$$

donde $r : X \times A \to \mathbb{R}$, es la función de recompensa.

f) La maximización está sujeta a las siguientes restricciones,

$$x_t \in X_t = \{x \in X : q(x) \leq 0\},$$

$t = 0, 1, ..., N$, donde $q : \mathbb{R}^n \to \mathbb{R}^n$.

Observación 3.1.1 *Nótese que la dinámica del sistema puede ser escrita como:*

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

donde $\widehat{F}(x_t, a_t) = \tau(x_t, a_t) - x_t$*.*

El problema de control óptimo consiste en determinar una sucesión de controles $\pi^* = (a_0^*, a_1^*, ..., a_{N-1}^*)$ y su correspondiente trayectoria $x^* = (x_0^*, x_1^*, ..., x_N^*)$ tal que se maximice (3.2).

Canon et. al. (véase [16]) presenta un enfoque para problemas de maximización restringidos con recompensa total y horizonte finito. La idea del trabajo consiste en transladar el problema de control óptimo en un problema de optimización general, llamado problema base (PB). Para resolver el problema base se hace uso de algunos teoremas de separación de análisis convexo.

La notación usada en esta sección se puede consultar en el Apéndice A.2.

A continuación se define el PB, bajo las siguientes condiciones sobre el modelo de control:

Hipótesis V

1) Las funciones $\tau : X \times A \to X$ y $r : X \times A \to \mathbb{R}$ son continuamente diferenciables.

2) Para cada $a \in A$ existe una aproximación cónica $C(a, A)$ (véase Apéndice, Definición A.2.4).

Problema Base

Sean $l : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ funciones continuamente diferenciables y $\Omega \subset \mathbb{R}^n$. Considérese el siguiente problema de optimización: encontrar un vector $z^* \in \mathbb{R}^n$ tal que

a) $z^* \in \Omega$,

b) $g(z^*) = 0$ y
$$l(z^*) = \max_{z \in \mathbb{R}^n} l(z).$$

El problema de control óptimo, planteado al inicio de esta sección, se puede llevar al problema base (PB) de la siguiente manera.

Sean $z = (x_0, x_1, ..., x_N, a_0, a_1, ..., a_{N-1})$, la sucesión de control y su correspondiente trayectoria de estados en el sistema y $\Omega = \mathbb{R}^{(N+1)n} \times \mathbb{R}^{Nm}$. Sea $g : \Omega \to \mathbb{R}^{Nn}$ la función definida por

$$g\left(x_0, x_1, ..., x_N, a_0, a_1, ..., a_{N-1}\right) = \begin{bmatrix} x_1 - x_0 - \tau(x_0, a_0) \\ x_2 - x_1 - \tau(x_1, a_1) \\ \cdot \\ \cdot \\ \cdot \\ x_N - x_{N-1} - \tau(x_{N-1}, a_{N-1}) \end{bmatrix}, \tag{3.3}$$

y $l : \Omega \to \mathbb{R}$ dada por

$$l\left(x_0, x_1, ..., x_N, a_0, a_1, ..., a_{N-1}\right) = \sum_{t=0}^{N-1} r(x_t, a_t). \tag{3.4}$$

Las ecuaciones (3.3) y (3.4) permiten vincular al problema de control óptimo con el problema base. Análogamente es posible llevar el problema base al problema de control óptimo (véase Canon [16]). De lo anterior resulta que el problema base es equivalente al problema de control óptimo.

El siguiente resultado provee una condición necesaria de optimalidad para el PB a partir del cual se puede derivar condiciones necesarias para problema de control óptimo. Las pruebas de dichos teoremas se pueden encontrar en la literatura de problemas de optimización matemática, en particular en el libro de Canon et. al. (véase [15] y [16]).

Lema 3.1.2 *Si z^* es una solución óptima al PB y $C(z^*, \Omega)$ es una aproximación cónica a Ω en z^*, entonces existe un vector no nulo $\psi = (\psi^0, \psi^1, \psi^2, ..., \ \psi^m) \in \mathbb{R}^{m+1}$, con $\psi^0 \leq 0$ tal que para todo $z \in \overline{C}(z^*, \Omega)$,*

$$\left\langle \psi, \frac{\partial G(z^*)}{\partial z} z \right\rangle \leq 0, \tag{3.5}$$

donde $G(z) = (l(z), g(z))$, $G : \mathbb{R}^n \to \mathbb{R}^{m+1}$ (la función de recompensa y la restricción de igualdad del problema básico) es continuamente diferenciable.

A continuación se presenta un bosquejo de la demostración del lema anterior, para una prueba detallada se puede consultar Canon [16].

Demostración. La prueba consiste en establecer la separación (véase Apén
dice, Definición A.2.1) del cono

$$K(z^*) = \left\{ y : y = \frac{\partial G(z^*)}{\partial z} z, \quad z \in C(z^*, \Omega) \right\},$$

con el rayo

$$R = \{y : y = \beta(-1, 0, 0, ..., 0), \quad \beta > 0\},$$

para esto se hace uso del teorema de separación (véase Apéndice, Teorema A.2.11). Si $K(z^*)$ y R están separados, entonces debe existir un vector no nulo $\psi \in \mathbb{R}^{m+1}$ tal que

$$\langle \psi, y \rangle \leq 0, \tag{3.6}$$

para todo $y \in K(z^*)$ y

$$\langle \psi, y \rangle \geq 0, \tag{3.7}$$

para todo $y \in R$. Nótese que (3.6) es ligeramente más débil que (3.5), pues no considera la cerradura del cono, pero si (3.6) es válida, entonces

$$\langle \psi, y \rangle \leq 0,$$

para todo $y \in \overline{K}(z^*)$, donde $\overline{K}(z^*)$ es la cerradura de $K(z^*)$, lo cual implica (3.5). Finalmente usando (3.7) se tiene que $\psi^0 \leq 0$ como lo afirma el Lema 3.1.2. ■

Sea $z = (x_0, x_{1,}..., x_N, a_0, a_1, ..., a_{N-1})$, la sucesión de controles y su correspondiente trayectoria de estados en el sistema. Sean $\Omega = X^{(N+1)} \times A^N$, $g : \Omega \to \mathbb{R}^{Nn}$ la función definida por

$$g\left(x_0, x_{1,}..., x_N, a_0, a_1, ..., a_{N-1}\right) = \begin{bmatrix} x_1 - x_0 - \tau(x_0, a_0) \\ x_2 - x_1 - \tau(x_1, a_1) \\ \cdot \\ \cdot \\ \cdot \\ x_N - x_{N-1} - \tau(x_{N-1}, a_{N-1}) \end{bmatrix}, \tag{3.8}$$

y $l : \Omega \to \mathbb{R}$ dada por

$$l\left(x_0, x_{1,}..., x_N, a_0, a_1, ..., a_{N-1}\right) = \sum_{t=0}^{N-1} r(x_t, a_t). \tag{3.9}$$

Las funciones l y g así definidas satisfacen las propiedades de diferenciabilidad requeridas. Supóngase que $\widehat{z} \in \Omega$ es una solución óptima al problema de control óptimo. Considere el cono

$$\begin{aligned} C(\widehat{z},\Omega) &= IC(\widehat{x}_0,X)\times IC(\widehat{x}_1,X)\times\cdots\times IC(\widehat{x}_N,X) \\ &\quad \times C(\widehat{a}_0,A)\times C(\widehat{a}_1,A)\times\cdots\times C(\widehat{a}_{N-1},A), \end{aligned}$$

donde IC significa cono interno (véase Apéndice, Definición A.2.8) y C son las aproximaciones cónicas, las cuales existen por hipótesis. El cono $C(\widehat{z},\Omega)$ así definido es una aproximación cónica para el conjunto Ω en $\widehat{z}$.

Aplicando el Lema 3.1.2 con $G=(l(z),g(z))$ se afirma, para $\widehat{z}$ una solución óptima al problema de control, que existe un vector $\psi=(p^0,\pi)$, $\psi\in\mathbb{R}\times X$ donde $p^0\in\mathbb{R}$, $p^0\leq 0$ y $\pi=(-p_1,-p_2,...,-p_N)$ con $p_t\in X$, $t=1,2,3,...,N$. De esta manera se obtiene que

$$\left\langle \psi,\frac{\partial G\left(\widehat{z}\right)}{\partial z}z\right\rangle = p^0\left\langle \nabla l\left(\widehat{z}\right),z\right\rangle + \left\langle \pi,\frac{\partial g\left(\widehat{z}\right)}{\partial z}z\right\rangle \leq 0,$$

para todo $z\in\overline{C}(\widehat{z},\Omega)$. De las ecuaciones (3.8) y (3.9) se tiene que

$$\begin{aligned} \left\langle \psi,\frac{\partial G\left(\widehat{z}\right)}{\partial z}z\right\rangle &= p^0\left[\sum_{t=0}^{N-1}\left\langle \frac{\partial r(\widehat{x}_t,\widehat{a}_t)}{\partial x},x_t\right\rangle + \sum_{t=0}^{N-1}\left\langle \frac{\partial r(\widehat{x}_t,\widehat{a}_t)}{\partial a},a_t\right\rangle\right] \\ &\quad + \sum_{t=0}^{N-1}\left\langle -p_{t+1},\left[x_{t+1}-x_t-\frac{\partial\tau(\widehat{x}_t,\widehat{a}_t)}{\partial x}x_t-\frac{\partial\tau(\widehat{x}_t,\widehat{a}_t)}{\partial a}a_t\right]\right\rangle \\ &\leq 0, \end{aligned} \tag{3.10}$$

para todo $z=(x_0,x_1,...,x_N,a_0,a_1,...,a_{N-1})\in\overline{C}\left(\widehat{z},\Omega\right)$.

El resultado formal es el siguiente teorema el cual describe la condición (3.10) en términos de la función Hamiltoniano. Definimos la función Hamiltoniano, $H:\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^n\times\mathbb{R}\to\mathbb{R}$ de la siguiente manera,

$$H(x,a,p,p^0)=p^0r(x,a)+\langle p,\tau(x,a)\rangle\,.$$

Teorema 3.1.3 *(Principio del Máximo) Si*

$$z^* = \left(x_0^*, x_1^*, x_2^*, ..., x_N^*, a_0^*, a_1^*, a_2^*, ..., a_{N-1}^*\right)$$

es una solución óptima al problema de control, entonces existen vectores $p_0, p_1, ..., p_N$ *en* $\mathbb{R}^n$ *no todos nulos, y un escalar* $p^0 \leq 0$ *tal que se satisface la ecuación adjunta*

$$p_N = \left[\frac{\partial q\left(x_N^*\right)}{\partial x}\right]^T \lambda \tag{3.11}$$

con $\lambda \in \mathbb{R}^n$, $\lambda \leq 0$ *y para* $t = 0, 1, ..., N-1$,

$$p_t - p_{t+1} = \left[\frac{\partial \tau(x_t^*, a_t^*)}{\partial x}\right]^T p_{t+1} + p^0 \left[\frac{\partial r(x_t^*, a_t^*)}{\partial x}\right]^T + \left[\frac{\partial q\left(x_t^*\right)}{\partial x}\right]^T \lambda, \tag{3.12}$$

con $p_0 = 0$, *tal que se satisface la condición*

$$\left\langle \frac{\partial H(x_t^*, a_t^*, p_{t+1}, p^0)}{\partial a}, a_t \right\rangle \leq 0$$

o bien

$$p^0 \left\langle \frac{\partial r(x_t^*, a_t^*)}{\partial a}, a_t \right\rangle + \left\langle p_{t+1}, \frac{\partial \tau(x_t^*, a_t^*)}{\partial a} a_t \right\rangle \leq 0, \tag{3.13}$$

para todo $a_t \in \overline{C}(a_t^*, A)$.

Demostración. Inicialmente se probará (3.13) para ello sea

$$z = (0, 0, ..., 0, 0, a_t, 0, 0, ..., 0), \tag{3.14}$$

con $a_t \in \overline{C}\left(a_t^*, A\right)$. Al sustituir (3.14) en (3.10) se obtiene directamente (3.13).

Ahora, se probará que las ecuaciones adjuntas están dadas por las ecuaciones (3.11) y (3.12) para cada $t = N, N-1, ..., 0$. Sean $t = N$ y,

$$z = (0, 0, ..., 0, 0, x_N, 0, 0, ..., 0),$$

con $x_N \in \overline{IC}(x_N^*, X)$, al sustituir nuevamente en (3.10) se tiene que

$$\langle -p_N, x_N \rangle \leq 0. \tag{3.15}$$

Por otro lado, al estar $x_N \in \overline{IC}(x_N^*, X)$ la desigualdad:

$$\left\langle \frac{\partial q\left(x_N^*\right)}{\partial x}, x_N \right\rangle \leq 0, \tag{3.16}$$

es válida para todo x_N^* con $q\left(x_N^*\right) = 0$. Aplicando el Lema de Farkas (véase Apéndice, Lema A.2.12) se garantiza que existe $\lambda \in \mathbb{R}^n$ con $\lambda \leq 0$ tal que de (3.15) y (3.16) se tiene que

$$-p_N = -\left[\frac{\partial q\left(x_N^*\right)}{\partial x}\right]^T \lambda,$$

de donde

$$p_N = \left[\frac{\partial q\left(x_N^*\right)}{\partial x}\right]^T \lambda,$$

lo cual prueba la ecuación (3.11).

Siguiendo la misma idea, para $t = 0, 1, ..., N-1$ haciendo

$$z = (0, 0, ..., 0, 0, x_t, 0, 0, ..., 0)\,,$$

con $x_t \in \overline{IC}(x_t^*, X)$ y al sustituir en (3.10) el z anterior se obtiene que

$$p^0 \left\langle \frac{\partial r\left(x_t^*, a_t^*\right)}{\partial x}, x_t \right\rangle + \left\langle -p_{t+1}, -x_t - \frac{\partial \tau(x_t^*, a_t^*)}{\partial x} x_t \right\rangle + \langle -p_t, x_t \rangle \leq 0,$$

equivalentemente,

$$\left\langle p^0 \left[\frac{\partial r\left(x_t^*, a_t^*\right)}{\partial x}\right]^T + p_{t+1}\left[\frac{\partial \tau(x_t^*, a_t^*)}{\partial x}\right]^T + p_{t+1} - p_t, x_t \right\rangle \leq 0, \tag{3.17}$$

$x_t \in \overline{IC}\left(x_t^*, X\right)$. Además, para $x_t \in \overline{IC}\left(x_t^*, X\right)$ se cumple que

$$\left\langle \frac{\partial q\left(x_t^*\right)}{\partial x}, x_t \right\rangle \leq 0 \tag{3.18}$$

para todo x_t^* tal que $q\left(x_t^*\right) = 0$. El Lema de Farkas junto con las ecuaciones (3.17) y (3.18) proporcionan el vector $\lambda \in \mathbb{R}^n$ con $\lambda \leq 0$ de modo que

$$p^0 \left[\frac{\partial r\left(x_t^*, a_t^*\right)}{\partial x}\right]^T + p_{t+1} \left[\frac{\partial \tau(x_t^*, a_t^*)}{\partial x}\right]^T + p_{t+1} - p_t = - \left[\frac{\partial q\left(x_t^*\right)}{\partial x}\right]^T \lambda,$$

de donde

$$p_t - p_{t+1} = p^0 \left[\frac{\partial r\left(x_t^*, a_t^*\right)}{\partial x}\right]^T + p_{t+1} \left[\frac{\partial \tau(x_t^*, a_t^*)}{\partial x}\right]^T + \left[\frac{\partial q\left(x_t^*\right)}{\partial x}\right]^T \lambda$$

es decir (3.12) es válida. ■

3.2. Principio del Máximo Estocástico

En esta sección se presenta la versión del Principio del Máximo para procesos de control de Markov estocásticos con horizonte finito, pero antes de establecer tal resultado se da la teoría necesaria para asegurar que la función de valor y la política óptima sean diferenciables. Esto se deriva de aplicar el Teorema de la Envolvente (Teorema 3.2.2) al método de Programación Dinámica, bajo supuestos de diferenciabilidad en la función de recompensa y dinámica, así como de la interioridad de la política óptima.

3.2.1. Diferenciabilidad

Sean p y m enteros positivos fijos. Sean $X \subseteq \mathbb{R}^p$, $A \subseteq \mathbb{R}^m$, con A convexo, y $A(x) \subset A$, $x \in X$ son conjuntos medibles y no vacíos (con respecto a la $\sigma-$álgebra de Borel de A).

Sea $G : \mathbb{K} \to \mathbb{R}$ una función medible donde $\mathbb{K} := \{(x, a) : x \in X, a \in A(x)\}$. Supóngase que existe una función medible $\widehat{H} : X \to \mathbb{R}$ tal que

$$G(x, a) \leq \widehat{H}(x),$$

para todo $x \in X$ y $a \in A(x)$. Defínase $g : X \to \mathbb{R}$ por

$$g(x) = \sup_{a \in A(x)} G(x,a), \tag{3.19}$$

$x \in X$.

Hipótesis VI

1. $G \in C^2(int(\mathbb{K}); \mathbb{R})$.

2. $G_{aa}(x, \cdot)$ es definida negativa, para cada $x \in X$.

3. Existe una función $h : X \to A$ tal que $h(x) \in int\ (A(x))$ y $g(x) = G(x, h(x))$, para cada $x \in X$.

Observación 3.2.1 *Para cada $x \in X$, la condición de primer orden para la optimalidad de $G(x, \cdot)$ está definida por medio de los puntos $a^* \in int\,(A\,(x))$ los cuales satisfacen $G_a\,(x, a^*) = 0$.*
El punto 2 de la Hipótesis VI implica que $G(x, \cdot)$ es una función estrictamente cóncava para cada $x \in X$.
En el punto 3 de la Hipótesis VI, si $A\,(x)$ es abierto para cada $x \in X$ entonces el maximizador $h\,(x)$ pertenece al interior de $A\,(x)\,,\ x \in X$.

Teorema 3.2.2 *Supóngase que las condiciones anteriores se satisfacen. Entonces $h \in C^1(int(X); A)$, $g \in C^2(int(X); \mathbb{R})$ y la siguiente fórmula*

$$g'(x) = G_x(x, h(x)), \tag{3.20}$$

es válida para todo $x \in int(X)$.

Demostración. Sea $x \in X$ fijo. Como el maximizador $h\,(x)$ está en el interior de $A\,(x)$, entonces de la condición de optimalidad de primer orden de G, resulta que $G_a\,(x, h\,(x)) = 0$.

Las Hipótesis VI 1 y 2 implican que $G_{aa}\,(x, \cdot)$ tiene una inversa G_{aa}^{-1} y que h es único (lo cual es consecuencia de la concavidad estricta de $G(x,\cdot)$. Entonces usando el Teorema de la Función Implícita, se tiene que $h \in C^1\left(int\left(X\right);A\right)$ y

$$h'\left(x\right) = -G_{aa}^{-1}\left(x,h(x)\right)G_{ax}\left(x,h(x)\right).$$

Por otro lado de la Hipótesis VI 3 se sigue que

$$g'(x) = G_x\left(x,h(x)\right) + G_a\left(x,h\left(x\right)\right)h'(x).$$

Como $G_a\left(x,h\left(x\right)\right) = 0$ entonces $g'(x) = G_x\left(x,h(x)\right)$, es decir, (3.20) se cumple. En consecuencia,

$$g''(x) = G_{xx}\left(x,h(x)\right) + G_{xa}\left(x,h\left(x\right)\right)h'(x).$$

De la Hipótesis VI 1 y del hecho de que $h \in C^1(Int\left(X\right);A)$, se sigue que $g \in C^2\left(int\left(X\right);\mathbb{R}\right)$, ya que x es arbitrario. ■

En el desarrollo subsecuente se considera que la ley de transición es inducida por la siguiente ecuación en diferencias

$$x_{t+1} = L(F(x_t,a_t),\xi_t), \tag{3.21}$$

donde $L : X \times S \to X$, es medible y conocida, $t = 0,1,2...N-1$, y $x_0 = x$. En este caso suponemos que $F : \mathbb{K} \to X$ es una función medible conocida y $\{\xi_t\}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas con función de distribución μ.

Observación 3.2.3 *La ecuación en diferencias dada en (3.21) incluye el caso en que se tiene ruido aditivo y multiplicativo, es decir, L puede tomar la forma*

$$L(F(x_t,a_t),\xi_t) = F(x_t,a_t) + \xi_t,$$

o bien,

$$L(F(x_t,a_t),\xi_t) = F(x_t,a_t)\xi_t,$$

con $x \in X$, $a \in A\left(x\right)$, $s \in S$ donde $F \in C^2\left(int\left(\mathbb{K}\right);X\right)$.

Sea

$$G^t(x,a) = r(x,a) + \alpha \int V_{t+1}(L\left(F(x,a),s\right))d\mu(s), \qquad (3.22)$$

$(x,a) \in \mathbb{K}$ y $t = N-1,...,2,1,0$ con $V_N = 0$.

Se tendrán los siguientes supuestos en el modelo de control de Markov.

Hipótesis VII

1. $r \in C^2\left(int\left(\mathbb{K}\right);\mathbb{R}\right)$.
2. $x_{t+1} = L(F(x_t,a_t),\xi_t)$ para $t = N-1,...,2,1,0,$ con $x_0 = x$ y $L\left(\cdot,s\right) \in C^2\left(X;S\right)$ para cada $s \in S$.
3. $f_t(x) \in int(A(x))$ para cada $x \in X$, $t = N-1,...,2,1,0$.
4. $G^t_{aa}(x,\cdot)$ es definida negativa en A para cada $x \in X$ y $t = N-1,...,2,1,0$.
5. El intercambio entre derivadas e integrales es válido.

Observación 3.2.4 *La Hipótesis VII 4 implica que* $G^t_{aa}(x,\cdot)$ *es una función estrictamente cóncava para cada* $x \in X$ *y* $t = 0,1,2,...,N-1$*, además,* f_t *es única para cada* $t = N-1,...,2,1,0$.

Observación 3.2.5 *En el Lema 3.2.6 la Hipótesis VII 5 garantiza la diferenciabilidad de segundo orden de la integral*

$$\int V_t(L\left(F(x,a),s\right))d\mu(s),$$

con respecto a x *y* a*. En la práctica esta condición puede ser verificada cuando las derivadas parciales de* $V_t(L\left(F(x,a),s\right))$ *cumplen ciertas condiciones de acotabilidad, véase Aliprantis [1].*

Lema 3.2.6 $V_t \in C^2(int(X),\mathbb{R})$ *para cada* $t = N-1,...,2,1,0,$ *con* $V'_N = 0$ *y*

$$V'_t(x) = G^t_x(x, f_t(x)),$$

con $x \in X$. $f_t \in C^1\left(int\left(X\right);A\right)$ *donde* f_t *son los minimizadores del algoritmo de programación dinámica.*

Demostración. La prueba se hará por inducción. Considere $t = N - 1$, en este caso

$$V_{N-1}(x) := \max_{a \in A(x)} r(x, a),$$

$x \in X$. Como $f_{N-1} \in int\,(A\,(x))$, $x \in X$, el minimizador f_{N-1} satisface la condición de primer orden para la optimalidad de r, es decir, $r_a\,(x, f_{N-1}\,(x)) = 0$, $x \in X$. Usando el Teorema de la Función Implícita y la Hipótesis VII 1 y 4 se tiene que $f_{N-1} \in C^1\,(int\,(X)\,; A)$. Como $V_{N-1}(x) = r\,(x, f_{N-1}\,(x))$, $x \in X$, entonces

$$V'_{N-1}(x) = r_x\,(x, f_{N-1}\,(x)) + r_a\,(x, f_{N-1}\,(x))\, f'_{N-1}\,(x)\,,$$

$x \in X$, usando la condición de primer orden para la optimalidad de r se obtiene que

$$V'_{N-1}(x) = r_x\,(x, f_{N-1}\,(x))\,,$$

$x \in X$. Así, como $r \in C^2\,(int\,(\mathbb{K})\,; \mathbb{R})$ y $f_{N-1} \in C^1\,(int\,(X)\,; A)$ entonces $V_{N-1} \in C^2\,(int\,(X)\,; \mathbb{R})$. Ahora supóngase que $V_t \in C^2\,(int\,(X)\,; \mathbb{R})$ para $t < N-1$. Se probará que $V_{t-1} \in C^2\,(int\,(X)\,; \mathbb{R})$. De la Hipótesis VII 1 y VII 5 se tiene que $G^{t-1}(x, a) \in C^2\,(int\,(\mathbb{K})\,; \mathbb{R})$, luego, de la condición de primer orden para la optimalidad de $G^{t-1}(x, a)$ se tiene que

$$G_a^{t-1}(x, f_{t-1}\,(x)) = 0,$$

$x \in X$. Nuevamente, del Teorema de la Función Implícita y de la Hipótesis VII 4 se obtiene que $f_{t-1} \in C^1\,(int\,(X)\,; A)$. Entonces,

$$V'_{t-1}\,(x) = G_x^{t-1}(x, f_{t-1}\,(x)) + G_a^{t-1}(x, f_{t-1}\,(x)) f'_{t-1}\,(x)\,,$$

$x \in X$. Usando la condición de primer orden en la igualdad anterior resulta que $V'_{t-1}\,(x) = G_x^{t-1}(x, f_{t-1}\,(x))$, $x \in X$. Finalmente, como $G^{t-1}(x, a) \in C^2\,(int\,(\mathbb{K})\,; \mathbb{R})$ y $f_{t-1} \in C^1\,(int\,(X)\,; A)$ entonces $V_{t-1} \in C^2\,(int\,(X)\,; \mathbb{R})$. ∎

A continuación, se presenta una versión del Principio del Máximo en el contexto de Procesos de Control de Markov con horizonte finito.

Teorema 3.2.7 *Supóngase que $\pi^* = \{f_0, f_1, ..., f_{N-1}\}$ es una política óptima para el problema descrito. Entonces existen funciones ϕ_t con $t = 0, 1, ..., N$, tales que satisfacen las siguientes relaciones:*

$$\phi_t(x) = r_x(x, f_t(x)) + \alpha E\left[\phi_{t+1}(L(u,\xi))L_u(u,\xi)\right] F_x(x, f_t(x)), \tag{3.23}$$

$t = 0, ..., N-1$, y

$$\phi_N(x) = 0,$$

$x \in X$. Además

$$0 = r_a(x, f_t(x)) + \alpha E\left[\phi_{t+1}(L(u,\xi))L_u(u,\xi)\right] F_a(x, f_t(x)), \tag{3.24}$$

para $t = 0, .., N-1$. Donde $u = F(x, f_t(x))$.

Demostración. La idea de la demostración es proponer a ϕ_t, $t \leq N-1$, como la derivada de la función V_t definida en el Teorema 2.2.1. Esto es posible debido a que el Lema 3.2.6 garantiza la diferenciabilidad de dicha función.

La prueba se hará por inducción. Sea $t = N-1$, con $V_N(x) = 0$, $x \in X$. En este caso se tiene $G^{N-1}(x,a) = r(x,a)$. Por hipótesis $r \in C^2$, de donde

$$\begin{aligned} V_{N-1}(x) &= \max_{a \in A(x)} r(x,a) \\ &= r(x, f_{N-1}(x)), \end{aligned}$$

$x \in X$. Del Lema 3.2.6 se tiene que $V_{N-1}(x)$ es diferenciable, así

$$V'_{N-1}(x) = r_x(x, f_{N-1}(x)) + r_a(x, f_{N-1}(x)) f'_{N-1}(x),$$

por otro lado

$$r_a(x, f_{N-1}(x)) f'_{N-1}(x) = 0,$$

$x \in X$, ya que representa la condición de primer orden del maximizador $f_{N-1}(x) \in int(A(x))$. Entonces,

$$\phi_{N-1}(x) = V'_{N-1}(x) = r_x(x, f_{N-1}(x)),$$

$x \in X$. Ahora, supóngase que el resultado es válido para $t < N-1$ entonces

$$G^{t}(x,a) = r(x,a) + \alpha \int V_{t+1}(L(F(x,a),s))\, d\mu(s),$$

y

$$V_{t}(x) = \max_{a \in A(x)} G^{t}(x,a),$$

de la condición de primer orden se tiene que

$$G_{a}^{t}(x, f_{t}(x)) = 0,$$

y

$$V_{t}'(x) = G_{x}^{t}(x, f_{t}(x)) + G_{a}^{t}(x, f_{t}(x))\, f_{t}'(x),$$

$x \in X$. Entonces para $x \in X$

$$\phi_{t}(x) = V_{t}'(x) = G_{x}^{t}(x, f_{t}(x)).$$

Se probará que el resultado es válido para $t-1$.

Sea

$$G^{t-1}(x,a) = r(x,a) + \alpha \int V_{t}(L(F(x,a),s))\, d\mu(s),$$

y

$$V_{t-1}(x) = \max_{a \in A(x)} G^{t-1}(x,a),$$

luego, bajo el supuesto de intercambio entre derivada e integral se tiene que la condición de primer orden esta dada por

$$\begin{aligned} 0 &= r_{a}(x, f_{t-1}(x))\, f_{t-1}'(x) \\ &\quad + \alpha \int V_{t}'(L(u,s))\, L_{u}(u,s)\, F_{a}(x, f_{t-1}(x))\, f_{t-1}'(x)\, d\mu(s), \end{aligned}$$

con $u = F(x,a)$. Por otro lado, para la ecuación adjunta

$$V_{t-1}'(x) = G_{x}^{t-1}(x,a),$$

es decir,

$$\begin{aligned} V_{t-1}'(x) &= r_x(x, f_{t-1}(x)) + r_a(x, f_{t-1}(x)) f_{t-1}'(x) \\ &+ \alpha \int V_t'(L(u,s)) [L_u(u,s) F_x(x, f_{t-1}'(x)) \\ &+ L_u(u,s) F_a(x, f_{t-1}(x)) f_{t-1}'(x)] d\mu(s) \end{aligned}$$

y de la condición de primer orden se obtiene que

$$\begin{aligned} V_{t-1}'(x) &= r_x(x, f_{t-1}(x)) \\ &+ \alpha \int V_t'(L(u,s)) L_u(u,s) F_x(x, f(x)) d\mu(s), \end{aligned}$$

como $\phi_{t-1}(x) = V_{t-1}'(x)$, entonces

$$\begin{aligned} \phi_{t-1}(x) &= r_x(x, f_{t-1}(x)) \\ &+ \alpha \int V_t'(L(u,s)) L_u(u,s) F_x(x, f_{t-1}'(x)) d\mu(s) \end{aligned}$$

lo cual muestra (3.23). ■

Las ecuaciones (3.23) y (3.24) se pueden escribir en términos de la función Hamiltoniano

$$H(x, a, \omega) = r(x, a) + \alpha \int \omega(y) Q(dy \mid x, a),$$

con $x \in X$, $a \in A(x)$, $\omega \in C^2(X)$ y $t = 0, 1, ..., N$. Es decir, la ecuación (3.23) y la ecuación de primer orden (3.24) se obtienen de la siguiente manera

$$\begin{aligned} \phi_t(x) &= \left. \frac{\partial H}{\partial x} \right|_{\substack{\omega' = \phi_{t+1} \\ a = f_t}}, \\ 0 &= \left. \frac{\partial H}{\partial a} \right|_{\substack{\omega' = \phi_{t+1} \\ a = f_t}}, \end{aligned}$$

respectivamente, donde

$$\frac{\partial H}{\partial x} = r_x(x, f_t(x)) + \alpha E\left[\phi_{t+1}(L(u,\xi)) L_u(u,\xi)\right] F_x(x, f_t(x)),$$

y

$$\frac{\partial H}{\partial a} = r_a(x, f_t(x)) + \alpha E\left[\phi_{t+1}(L(u,\xi))L_u\left(u,\xi\right)\right] F_a(x, f_t(x)).$$

Observación 3.2.8 *En el caso en que se considere recompensa final* $r^N(x)$*, se puede tomar* $V_N(x) = r^N(x)$.

Observación 3.2.9 *La prueba dada en el Teorema 3.2.7 referente al Principio del Máximo está basada en Programación Dinámica. La idea de la prueba consiste en proponer al multiplicador* ϕ *como la derivada de las funciones* V_t *del algoritmo de Programación Dinámica. En otras palabras, en el contexto económico, se está proponiendo al multiplicador como el valor óptimo marginal del problema. En este contexto se les conoce a los multiplicadores como valores sombra (ver Caputo [17]).*

Capítulo 4

Aplicaciones del Principio del Máximo

En esta sección se presentarán algunos ejemplos de la teoría de control estocástico. Dichos ejemplos serán resueltos usando el Teorema 3.2.7 enunciado en el capítulo anterior.

4.1. Modelo Lineal-Cuadrático

El caso en que la dinámica del sistema se describe por una ecuación en diferencias lineal y el costo es descrito por un funcional cuadrático se le llama problema Lineal Cuadrático (LQ). Ésta es una formulación de un problema de regulación en el cual se pretende que el estado del sistema permanezca lo más cerca posible a un punto fijo, por ejemplo, mantener la temperatura de una calefacción, el rumbo de una avión o la velocidad de un automóvil en un valor establecido. Usar la función de costo cuadrático, es en muchas ocasiones razonable, puesto que induce una alta penalidad para grandes desviaciones del estado al punto fijo, pero penalidades relativamente pequeñas para desviaciones pequeñas. Además, bajo ciertas condiciones en el modelo de control es posible encontrar aproximaciones de problemas de control de Markov por problemas lineales cuadráticos (véase por ejemplo

Santos [42])

Ejemplo 4.1.1 *Sean* $X = \mathbb{R}^n$*,* $A = A(x) = \mathbb{R}^n$*,*

$$x_{t+1} = Ax_t + Ba_t,$$

$t = 0, 1, 2, ..., N-1$ *y la función de costo dado por*

$$c(x_t, a_t) = x_t' Q x_t + a_t' R a_t + x_N' Q x_N,$$

donde A*,* B*,* Q*,* R *son matrices cuadradas,* Q *es simétrica y semidefinida positiva,* R *es simétrica y definida positiva. Para poder aplicar el Principio del Máximo considere el problema de maximizar* $-c(x,a)$*:*

$$v(\pi, x) = \sum_{t=0}^{N-1} -\left(x_t' Q x_t + a_t' R a_t + x_N' Q x_N\right), \qquad (4.1)$$

$x \in X$ *y* $\pi \in \Pi$*.*

El siguiente lema garantiza la existencia de una política óptima como una consecuencia del Teorema 2.2.5, bajo el bloque de Hipótesis I.

Lema 4.1.2
a) La función de recompensa, $r(x,a) = -c\,(x,a)$*,* $(x,a) \in \mathbb{K}$*, es semicontinua superiormente, acotada superiormente y sup-compacta.*
b) La ley de transición es débilmente continua.

Demostración. Para el inciso a) se probará el equivalente, es decir, $c(x,a)$ es semicontinua inferiormente, acotada inferiormente e inf-compacta.

Nótese que la función de costo es continua por lo que sólo falta probar que es inf-compacta en $\mathbb{K}$, es decir, para todo $x \in X$ y $\lambda \in \mathbb{R}$, el conjunto

$$A_\lambda\,(x) = \{a \in A(x) \mid x'Qx + a'Ra + x_N' Q x_N \leq \lambda\},$$

es compacto. Sean $x \in X$ y $\lambda \in \mathbb{R}$, afirmamos que $A_\lambda(x)$ es acotado, de lo contrario, existe una sucesión $\{\widehat{a}_n\}$ en $A_\lambda(x)$, tal que $\widehat{a}_n \to \infty$, entonces,

$$\lim_{n\to\infty} c(x, \widehat{a}_n) = \infty,$$

esto implica que $\infty \leq \lambda$, lo cual es una contradicción. Ahora, sea $\{\widetilde{a}_n\}$ una sucesión en $A_\lambda(x)$, tal que $\widetilde{a}_n \to a \in A$, como $0 \leq c(x, \widetilde{a}_n) \leq \lambda$ y c es continua, se tiene que $0 \leq c(x,a) \leq \lambda$, entonces $a \in A_\lambda(x)$, por lo que se concluye que $A_\lambda(x)$ es cerrado. Por lo tanto $A_\lambda(x)$ es compacto y como λ y $x \in X$ son arbitrarios tenemos que c es ínf-compacto en $\mathbb{K}$.

b) Sea $W \in B(X)$, entonces

$$\begin{aligned} P(W \quad | \quad x) &= P(x_{t+1} \in W \mid x_t = x, a_t = a) \\ &= P\left(Ax + Ba \in W\right) \\ &= I_W\left(Ax + Ba\right), \end{aligned}$$

en consecuencia la ley de transición es débilmente continua. ■

Observación 4.1.3 *Es claro que se satisfacen los supuestos de diferenciabilidad en la función de recompensa y transición del sistema. Además, el minimizador $f_t(x)$ es único para cada t, debido a que se cumplen las condiciones C2 del artículo de Cruz-Suárez et. al., véase [22], además se alcanza en el interior para cada $x \in X$, $t = N-1, ..., 1, 0$, puesto que X es abierto.*

Lema 4.1.4 *La política óptima para el problema lineal cuadrático es $\pi^* = \{f_t\}_{t=0}^{N-1}$ donde*

$$f_t(x) = -[R + B'K_{t+1}B]^{-1}B'K_{t+1}Ax, \tag{4.2}$$

$x \in X$, y

$$K_t = Q + A'(K_{t+1} - K_{t+1}B[R + B'K_{t+1}B]^{-1}B'K_{t+1})A,$$

con $K_N = Q$.

Demostración. En $t = N$ se tiene $f_N(x) = 0$ y $\phi_N(x) = -2Qx_N$. Usando el Teorema 3.2.7, de la ecuación (3.23) se tiene para $t = N - 1$

$$\begin{aligned}\phi_{N-1}(x) &= -2Qx - 2Q[Ax + Bf_{N-1}(x)]A \\ &= -2Qx - 2A'QAx - 2A'QBf_{N-1}(x) \\ &= -2[Qx + A'QAx + A'QBf_{N-1}(x)],\end{aligned}$$

luego de la ecuación de primer orden (3.24),

$$\begin{aligned}0 &= -2Rf_{N-1}(x) + [-2Q(Ax + Bf_{N-1}(x))B] \\ 0 &= -2Rf_{N-1}(x) - 2B'QAx - 2B'QBf_{N-1}(x) \\ 0 &= [R + B'QB]f_{N-1}(x) + B'QAx,\end{aligned}$$

de donde

$$f_{N-1}(x) = -[R + B'QB]^{-1}B'QAx,\ x \in X, \tag{4.3}$$

evaluando para $\phi_{N-1}(x)$,

$$\begin{aligned}\phi_{N-1}(x) &= -2[Qx + A'QAx + A'QB(-[R + B'QB]^{-1}B'QAx)] \\ &= -2[Q + A'QA - A'QB[R + B'QB]^{-1}B'QA]x \\ &= -2[Q + A'(Q - QB[R + B'QB]^{-1}B'Q)A]x \\ &= -2K_{N-1}x,\ \ x \in X\end{aligned}$$

con

$$K_{N-1} = [Q + A'(Q - QB[R + B'QB]^{-1}B'Q)A].$$

El valor obtenido en (4.3) verifica que (4.2) es válido para $t = N - 1$. Supóngase que (4.2) se cumple para $t < N - 1$. Es decir,

$$\phi_t(x) = -2K_t x,\ \ x \in X,$$
$$f_t(x) = -[R + B'K_{t+1}B]^{-1}B'K_{t+1}Ax,\ \ x \in X,$$

Se demostrará que (4.2) es válida para $t-1$. En este caso se tiene que

$$\begin{aligned}\phi_{t-1}(x) &= -2Qx - 2K_t[Ax + Bf_{t-1}(x)] \\ &= -2Qx - 2A'K_tAx - 2A'K_tBf_{t-1}(x) \\ &= -2[Qx + A'K_tAx + A'K_tBf_{t-1}(x)],\end{aligned}$$

y

$$\begin{aligned}
0 &= -2Rf_{t-1}(x) + [-2K_t(Ax + Bf_{t-1}(x))B] \\
0 &= -2Rf_{t-1}(x) - 2B'K_tAx - 2B'K_tBf_{t-1}(x) \\
0 &= [R + B'K_tB]f_{t-1}(x) + B'K_tAx,
\end{aligned}$$

así

$$f_{t-1}(x) = -[R + B'K_tB]^{-1}B'K_tAx, \;\; x \in X,$$

con lo cual

$$\begin{aligned}
\phi_{t-1}(x) &= -2[Qx + A'K_tAx + A'K_tB(-[R + B'K_tB]^{-1}B'K_tAx)] \\
&= -2[Q + A'K_tA - A'K_tB[R + B'K_tB]^{-1}B'K_tA]x \\
&= -2[Q + A'(K_t - K_tB[R + B'K_tB]^{-1}B'K_t)A]x \\
&= -2K_{t-1}x,
\end{aligned}$$

con $x \in X$, donde

$$K_{t-1} = [Q + A'(K_t - K_tB[R + B'K_tB]^{-1}B'K_t)A],$$

de acuerdo a lo anterior el lema está probado. ■

La función de valor óptimo se determina por evaluación directa del control óptimo $f_t(x)$ en (4.1).

Ejemplo 4.1.5 *Sea $X = A = S = \mathbb{R}$, $A(x) = [x-1, x+1]$, $x \in X$. La función de costo y la dinámica del sistema están dados por*

$$\begin{aligned}
c(x_t, a_t) &= (x_t - a_t)^2 + T, \\
x_{t+1} &= a_t + \xi_t,
\end{aligned}$$

respectivamente, $t = 0, 1, 2, ..., N-1$ y T un número real fijo. Las variables aleatorias ξ_t, $t = 0, 1, 2, ..., N-1$, se suponen independientes e idénticamente distribuidas tomando valores en el espacio $S = \mathbb{R}$ y con densidad común

$$\Delta(y) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2},$$

$y \in \mathbb{R}$.

Considere el criterio de rendimiento $\alpha-$*descontado, con* $0 < \alpha < 1$,

$$v(\pi, x) = -\left[E \sum_{t=0}^{N-1} \alpha^t (x_t - a_t)^2 + \alpha^t T \right]. \tag{4.4}$$

Lema 4.1.6 *Existe una política óptima.*

Demostración. Para probar la existencia de una política óptima se deben verificar las condiciones (véase hipótesis de selección medible)

a) $A(x)$ es compacto para toda $x \in X$.

b) La función de costo es semicontinua inferiormente en $A(x)$.

c) La ley de transición Q es fuertemente continua.

Para esto nótese que sólo falta verificar c). En este caso la ley de transición Q está dada por:

$$\begin{aligned} Q(B \quad | \quad & x, a) = \Pr(x_{t+1} \in B \mid x_t = x, a_t = a) \\ = \quad & \Pr(F(x_t, a_t, \xi_t) \in B \mid x_t = x, a_t = a) \\ = \quad & \Pr(a_t + \xi_t \in B \mid x_t = x, a_t = a) \\ = \quad & \Delta\{s \mid a + s \in B\} \\ = \quad & \int\limits_S I_B[a+s]\Delta(s)ds, \end{aligned}$$

haciendo un cambio de variable se tiene que

$$Q(B \mid x, a) = \int\limits_S I_B[u]\Delta(u-a)ds,$$

de aquí, como la densidad Δ es continua se garantiza que Q es fuertemente continua. ■

Lema 4.1.7 *Existe una única política óptima* $\pi^* = \{f_t\}_{t=0}^{N-1}$ *tal que* $f_t(x) \in int(A(x)),\ \ x \in X$.

Demostración. La existencia de la política óptima se garantiza en el Lema 4.1.6. La unicidad de la política se tiene puesto que las componentes del modelo satisfacen las condiciones C2 del artículo [22]. Ahora, de

$$v(\pi, x) = -\left[E \sum_{t=0}^{N-1} \alpha^t \left[(x_t - a_t)^2 + T\right]\right],$$

evaluando en la política $g(x) = x$, $x \in X$, se tiene que

$$\begin{aligned} V(g(x), x) &= -\left[E \sum_{t=0}^{N-1} \alpha^t T\right] \\ &= -T\frac{\alpha^{N-1}}{\alpha - 1}, \end{aligned}$$

y en la frontera de $A(x)$ se tiene para $g_1(x) = x - 1$, $x \in X$,

$$\begin{aligned} V(g_1(x), x) &= -\left[E \sum_{t=0}^{N-1} \alpha^t \left[1 + T\right]\right] \\ &= -[T+1]\frac{\alpha^{N-1}}{\alpha - 1}, \end{aligned}$$

y para $g_2(x) = x + 1$ con $x \in X$,

$$\begin{aligned} V(g_2(x), x) &= -\left[E \sum_{t=0}^{N-1} \alpha^t \left[1 + T\right]\right] \\ &= -[T+1]\frac{\alpha^{N-1}}{\alpha - 1}. \end{aligned}$$

Usando las relaciones anteriores se tiene que

$$V\left(g(x), x\right) > V\left(g_1\left(x\right), x\right) = V\left(g_2(x), x\right),$$

$x \in X$. Por lo tanto el óptimo se alcanza en el interior de $A\left(x\right)$.
■

Lema 4.1.8 *Para cada* $t = 0, 1, 2, ..., N-1$, $G^t \in C^2\left(int\left(\mathbb{K}\right); \mathbb{R}\right)$ *y* $G^t_{aa}(x, \cdot) > 0$.

Demostración. Sea t fijo, aplicando el Teorema de Cambio de Variable se tiene que

$$\begin{aligned} H\left(a\right) &= \int V_t\left(a+s\right)\Delta\left(s\right)ds = \int V_t\left(u\right)\Delta\left(u-a\right)du \\ &= \int \frac{V_t\left(u\right)}{\sqrt{2\pi}}e^{-1/2(u-a)^2}du \\ &= \int M^t\left(u,a\right)du, \end{aligned}$$

donde

$$M^t\left(u,a\right) = \frac{V_t\left(u\right)}{\sqrt{2\pi}}e^{-1/2(u-a)^2}, \quad u \in \mathbb{R} \text{ } y \text{ } a \in A\left(x\right).$$

Así,

$$M^t_a\left(u,a\right) = \frac{V_t\left(u\right)}{\sqrt{2\pi}}e^{-1/2(u-a)^2}\left(u-a\right), \quad u \in \mathbb{R} \text{ } y \text{ } a \in A\left(x\right).$$

Como $x - 1 \leq a \leq x + 1$ y puesto que

$$-T\frac{\alpha^{N-1}}{\alpha-1} = V_t\left(g(x), x\right) \leq V\left(x\right) \leq 0,\ x \in X,$$

donde $g(x) = x$, con $x \in X$, entonces los casos siguientes son posibles para determinar cotas para M^t_a.

Caso (1)

Supóngase que $u - a \geq 0$, entonces

$$\begin{aligned} \left|M^t_a\left(u,a\right)\right| &\leq \frac{V_t\left(u\right)}{\sqrt{2\pi}}e^{-\frac{1}{2}(u-a)^2}\left(1-x+u\right) \\ &\leq \frac{T\left(\alpha^{N-1}\right)}{\left(\alpha-1\right)\sqrt{2\pi}}e^{-\frac{1}{2}(u-a)^2}\left(1-x+u\right) \\ &\leq \theta_1\left(u,x\right),\ \ x \in X, \end{aligned}$$

donde

$$\begin{aligned}\theta_1(u,x) &= \frac{T\left(\alpha^{N-1}\right)(1-x+u)}{(\alpha-1)\sqrt{2\pi}} \\ &\quad \left[e^{-\frac{1}{2}u^2+u(x+1)}I_{[0,\infty)}(u)+e^{-\frac{1}{2}u^2+u(x-1)}I_{(-\infty,0]}(u)\right],\end{aligned}$$

con $I_{[\cdot]}$ denotando la función indicadora del conjunto $[\cdot]$.

Caso (2)

Ahora, suponga que $u-a<0$, entonces

$$\begin{aligned}\left|M_a^t(u,a)\right| &\leq \frac{V_t(u)}{\sqrt{2\pi}}e^{-\frac{1}{2}(u-a)^2}(1+x-u) \\ &\leq \frac{T\left(\alpha^{N-1}\right)}{(\alpha-1)\sqrt{2\pi}}e^{-\frac{1}{2}(u-a)^2}(1+x-u) \\ &\leq \theta_2(u,x),\ \ x\in X,\end{aligned}$$

donde

$$\begin{aligned}\theta_2(u,x) &= \frac{T\left(\alpha^{N-1}\right)(1+x-u)}{(\alpha-1)\sqrt{2\pi}} \\ &\quad \left[e^{-\frac{1}{2}u^2+u(x+1)}I_{[0,\infty)}(u)+e^{-\frac{1}{2}u^2+u(x-1)}I_{(-\infty,0]}(u)\right].\end{aligned}$$

Como x es arbitrario, se sigue que $\theta=\theta_1+\theta_2$ es una cota para $|M_a^t|$. Más aún, se puede verificar que esta cota es integrable, es decir, $\int\theta(u,x)\,du<\infty$ para todo $x\in X$. La cota para M_{aa} se puede obtener de forma similar. Entonces $H\in C^2(A;\mathbb{R})$. Como $c\in C^2(int(\mathbb{K});\mathbb{R})$ entonces $G_t\in C^2(\mathbb{K};\mathbb{R})$, $t=0,1,2,...,N-1$. Además $G_{aa}^t>0$, $t=0,1,2,...,N-1$ pues $c_{aa}>0$ y $H_{aa}\geq 0$, esto en consecuencia de la convexidad y diferenciabilidad de segundo orden de H. ■

Lema 4.1.9 *La política óptima para el Ejemplo 4.1.4 es* $\pi^*=\{f_t\}_{t=0}^{N-1}$ *donde*

$$f_t(x)=x,$$

$x \in X$ y la función de valor óptimo es

$$V(x) = -\frac{\alpha^N - 1}{\alpha - 1} T,$$

$x \in X$.

Demostración. En este caso se considera costo terminal 0, y por lo tanto $\phi_N(x) = 0$, $x \in X$.

De la ecuación (3.23) para $t = N - 1$, se obtiene que

$$\phi_{N-1}(x) = 2(x - f_{N-1}(x))$$

$x \in X$ y de la ecuación de primer orden (3.24), se tiene que

$$0 = -2(x - f_{N-1}(x))$$

de donde $f_{N-1}(x) = x$, $x \in X$ y por lo tanto $\phi_{N-1}(x) = 0$.

Ahora, supóngase que

$$\begin{aligned} \phi_t(x) &= 0, \\ f_t(x) &= x, \ \ x \in X. \end{aligned}$$

Entonces, nuevamente de las ecuaciones (3.23) y (3.24) se obtiene que

$$\begin{aligned} \phi_{t+1}(x) &= 2(x - f_{t+1}(x)) \\ 0 &= -2(x - f_{t+1}(x)), \end{aligned}$$

con lo cual $\phi_{t+1}(x) = 0$ y $f_{t+1}(x) = x$, $x \in X$.

Finalmente, para obtener la función de valores óptimos se evalúa el control óptimo en (4.4), es decir,

$$\begin{aligned} V(x) &= -\left[E\sum_{t=0}^{N-1} \alpha^t (x_t - a_t)^2 + \alpha^t T\right] \\ &= -\left[E\sum_{t=0}^{N-1} \alpha^t T\right] \\ &= -T\frac{\alpha^{N-1}}{\alpha - 1}, \end{aligned}$$

$x \in X$. ■

4.2. Aplicaciones en Economía

En esta sección se presentarán dos ejemplos en el contexto económico. El primero de ellos se encuentra relacionado con los problemas de Crecimiento Económico. El Crecimiento Económico se puede entender como el aumento de la cantidad de trabajo, la renta o valor de bienes y servicios producidos por una economía. El crecimiento así definido se ha considerado deseable, ya que guarda una relación con la cantidad de bienes materiales y por ende una cierta mejora del nivel de vida de las personas. Robert Lucas (premio Nobel de economía en 1995) declara que la oportunidad de un país para emprender su crecimiento depende de los ingresos promedio a nivel mundial en esos momentos. Cuanto más rico el mundo en general, mayor la oportunidad de cualquier país preindustrializado para empezar a crecer. En general, un modelo de crecimiento económico presenta la siguiente forma:

a) $\widehat{X} \subset \mathbb{R}^p$.

b) Una multifunción $\Gamma : X \rightarrow X$ es decir, para cada $x \in X$ existe un subconjunto no vacío $\Gamma(x) \in X$, a Γ se le conoce como correspondencia tecnológica.

c) $U : Graph(\Gamma) \rightarrow \mathbb{R}$ es la función de recompensa, se asume que $U \leq 0$ o $U \leq \varepsilon$, donde ε es un número real fijo y $Graph(\Gamma) := \{(x, y) \in X \times X : \ x \in X, y \in \Gamma(x)\}$.

d) $\alpha \in (0, 1)$ un factor de descuento.

Entonces el problema de maximización es

$$\max_{\{x_{t+1}\}} \sum_{t=0}^{\infty} \alpha^t U(x_t, x_{t+1})$$

sujeto a $x_{t+1} \in \Gamma(x_t)$, $t = 0, 1, ..., x_0 \in \widehat{X}$ fijo.

Es posible establecer este problema de maximización en términos de un proceso de control de Markov descontado, haciendo $X = A$, $A(x) = \Gamma(x)$, $x \in X$, $x_{t+1} = F(x_t, a_t) = a_t$ y $c(x_t, a_t) = -U(x_t, x_{t+1})$, $t = 0, 1, 2, ...$

Ejemplo 4.2.1 *Sea* $X = A = [0,1)$, $A(0) = \{0\}$; $A(x) = (0, x^\delta]$, $x \in X$, $x \neq 0$, *donde* $\delta \in (0,1)$ *es un número fijo. La ley de transición esta dado por* $x_{t+1} = (x_t^\delta - a_t)$, $t = 0, 1, ..., N-1$. *Con función de recompensa definido por* $r(x,a) = \ln(a)$, *si* $x \in (0,1)$, $a \in (0, x^\delta]$, *y* $r(0,0) = -\infty$. *El criterio de rendimiento* α*-descontado se escribe como:,*

$$v(\pi, x) = \sum_{t=0}^{N-1} \alpha^t \ln(a_t),$$

donde $0 < \alpha < 1$, *con recompensa final* 0.

Lema 4.2.2

a) *r es continua, acotada superiormente y sup-compacta en* $\mathbb{K}$.

b) *La ley de transición es débilmente continua.*

Demostración. a) La función logaritmo es continua, además, como $0 < a \leq x^\delta$ y $0 \leq x < 1$ se tiene que la recompensa esta acotada superiormente por 0. Para ver que es sup-compacta, se probará que el conjunto $A_\lambda(x) = \{a \in A(x) \mid -c(x,a) \leq \lambda\}$, es compacto con $\lambda \in \mathbb{R}$ y $x \in X$. Para ello nótese que

$$\begin{aligned} A_\lambda(x) &= \{a \in A(x) \mid -\ln a \leq \lambda\} \\ &= \{a \in A(x) \mid a \geq e^{-\lambda}\}. \end{aligned}$$

Como $0 < a \leq x^\delta$ entonces $A_\lambda(x) = \left[e^{-\lambda}, x^\delta\right]$. Por lo tanto r es sup-compacta.

b) Se tiene la ley de transición

$$\begin{aligned} Q(B \quad | \quad & x, a) = P(x_{t+1} \in B \mid x_t = x, a_t = a) \\ = \quad & P\left(x^\delta - a \in B\right) \\ = \quad & I_B\left(x^\delta - a\right), \end{aligned}$$

por lo tanto la ley de transición es débilmente continua. ■

Observación 4.2.3 *El Ejemplo 4.2.1 satisface las hipótesis de diferenciabilidad en la función de transición y recompensa. Además* $f_t(x)$ *está en el interior para cada* $x \in X$, $t = N-1, ..., 1, 0$, *puesto que, para cada* $x \in (0,1)$

$$\begin{aligned} V_{t+1}(x) &= \max_{a \in (0, x^\delta]} \left[\ln(a) + \alpha V_t \left(x^\delta - a \right) \right] \\ &= \max_{a \in (0, x^\delta)} \left[\ln(a) + \alpha V_t \left(x^\delta - a \right) \right], \end{aligned}$$

$x \in (0,1)$, *pues* $V_t \left(x^\delta - x^\delta \right) = -\infty$.

Lema 4.2.4 *La política óptima para el Ejemplo 4.2.1 está dado por* $\pi^* = \{f_t\}_{t=0}^{N-1}$ *donde*

$$f_t(x) = \frac{\alpha\delta - 1}{(\alpha\delta)^{N-t} - 1} x^\delta, \quad x \in X, \tag{4.5}$$

y la función de valor óptimo es

$$\begin{aligned} V(x) &= \left[1 - \frac{\left(\alpha - \alpha^N\right)}{\alpha - 1} \right] \delta \ln(x) - \left[\frac{\left(\alpha - \alpha^N\right)}{\alpha - 1} \right] \ln\left[1 - \alpha\delta\right] \\ &\quad - \sum \alpha^t \ln\left[1 - (\alpha\delta)^{t+1} \right], \end{aligned} \tag{4.6}$$

$x \in X$.

Demostración. En este caso, como no se considera recompensa terminal se tiene $\phi_N(x) = 0$. El proceso de optimización comienza en $t = N-1$, en éste paso, $x_{N-1} = x$ y $f_{N-1}(x) = x^\delta$ pues el máximo se alcanza en un extremo, entonces $\phi_{N-1}(x) = \frac{\delta}{x}$, $x \neq 0$, ya que el valor óptimo en $t = N-1$ es $V'_{N-1} = \ln(x^\delta)$, y dado que $\phi_{N-1}(x) = V'_{N-1}(x)$, $x \in X$. Con lo anterior se probará (4.5) para $t = 1, 2, ..., N-1$. Primero se probará que

$$f_t(x) = \frac{x^\delta}{K_t}, \quad t = 0, 1, 2, 3, ..., N-2,$$

y

$$\phi_t(x) = \frac{\delta}{x} K_t, \quad t = 0, 1, 2, 3, ..., N-2,$$

para $x \in X/\{0\}$, donde

$$K_t = \sum_{s=0}^{N-(t+1)} (\alpha\delta)^s, \quad t = 1, 2, 3, ..., N-2.$$

Nótese que $K_N = 0$ y $K_{N-1} = 1$. Sea $t = N-2$, de las ecuaciones (3.23) y (3.24) se sigue que

$$\phi_{N-2}(x) = \alpha \frac{\delta^2 x^{\delta-1}}{x^\delta - f_{N-2}(x)},$$

y,

$$0 = \frac{1}{f_{N-2}(x)} - \alpha \frac{\delta}{x^\delta - f_{N-2}(x)},$$

de donde

$$f_{N-2}(x) = \frac{x^\delta}{K_{N-2}}, \tag{4.7}$$

y

$$\phi_{N-2}(x) = \frac{\delta}{x} K_{N-2},$$

$x \in X/\{0\}$ con $K_{N-2} = 1 + \alpha\delta$. Ahora, supóngase que se cumple para $t < N-2$, esto es, se tiene que para $x \in X$

$$f_t(x) = \frac{x^\delta}{K_t},$$

y

$$\phi_t(x) = \frac{\delta}{x} K_t,$$

$x \in X/\{0\}$, donde

$$K_t = \sum_{s=0}^{N-(t+1)} (\alpha\delta)^s.$$

Se probará que se cumple para $t-1$. Nuevamente, de las ecuaciones (3.23) y (3.24) se sigue que

$$\phi_{t-1}(x) = \alpha \frac{\delta^2 x^{\delta-1}}{x^\delta - f_{t-1}(x)} K_t, \ x \in X,$$

y

$$0=\frac{1}{f_{t-1}(x)}-\alpha\frac{\delta K_t}{x^{\delta}-f_{t-1}(x)},$$

a partir de estas ecuaciones se obtiene el valor de $f_{t-1}(x)$, dado por

$$\begin{aligned} f_{t-1}(x) &= \frac{x^{\delta}}{1+\alpha\delta K_t}, \\ f_{t-1}(x) &= \frac{x^{\delta}}{K_{t-1}},\ x\in X \end{aligned}$$

donde

$$K_{t-1}=\sum_{s=0}^{N-t}(\alpha\delta)^s$$

con lo cual

$$\begin{aligned} \phi_{t-1}(x) &= \frac{\delta}{x}\frac{\alpha\delta K_t}{K_{t-1}-1}K_{t-1} \\ &= \frac{\delta}{x}K_{t-1}, \end{aligned}$$

$x\in X/\{0\}$. Como

$$\sum_{s=0}^{N-(t+1)}(\alpha\delta)^s=\frac{(\alpha\delta)^{N-t}-1}{\alpha\delta-1},$$

pues $\alpha\delta<1$. Entonces

$$f_t(x)=\frac{\alpha\delta-1}{(\alpha\delta)^{N-t}-1}x^{\delta},$$

$t=1,2,...,N-1$, $x\in X$. Así, se ha probado (4.5).

La función de valor óptimo se obtiene por evaluación directa de la política óptima $\pi^*=\{f_t\}$ en el criterio de rendimiento, obteniendo

$$\begin{aligned}
V(x) &= \sum_{t=0}^{N-1} \alpha^t \ln(a_t) \\
&= \delta \ln(x) + \sum_{t=1}^{N-1} \alpha^t \ln\left[\frac{\alpha\delta - 1}{(\alpha\delta)^{N-t} - 1} x^{\delta}\right] \\
&= \delta \ln(x) + \sum_{t=1}^{N-1} \alpha^t \left[\ln\left[\frac{1-\alpha\delta}{1-(\alpha\delta)^{N-t}}\right] + \delta \ln(x)\right] \\
&= \delta \ln(x) + \delta \ln(x) \sum_{t=1}^{N-1} \alpha^t + \sum_{t=1}^{N-1} \alpha^t \ln\left[\frac{1-\alpha\delta}{1-(\alpha\delta)^{N-t}}\right] \\
&= \delta \ln(x) \left[1 - \frac{\left(\alpha - \alpha^N\right)}{\alpha - 1}\right] + \sum_{t=1}^{N-1} \alpha^t \ln\left[1 - \alpha\delta\right] \\
&\quad - \sum_{t=1}^{N-1} \alpha^t \ln\left[1 - (\alpha\delta)^{N-t}\right],
\end{aligned}$$

finalmente

$$\begin{aligned}
V(x) &= \left[1 - \frac{\left(\alpha - \alpha^N\right)}{\alpha - 1}\right] \delta \ln(x) - \left[\frac{\left(\alpha - \alpha^N\right)}{\alpha - 1}\right] \ln\left[1 - \alpha\delta\right] \\
&\quad - \sum_{t=1}^{N-1} \alpha^t \ln\left[1 - (\alpha\delta)^{t+1}\right],
\end{aligned}$$

con $x \in X$. ■

Otro tipo de problemas referentes a economía son los de Consumo Inversión. Un consumidor trataría de maximizar su utilidad a lo largo de toda su vida, por lo que determinaría en base a esto, el consumo y el ahorro en cada instante del tiempo, controlaría dinámicamente qué parte de su presupuesto iría destinada al consumo y qué parte a ahorro. A nivel de política macroeconómica, por ejemplo, un gobierno se plantearía el control de la tasa de inflación y/o de la tasa de desempleo, a fin de conducirlas lo más próximo a unas tasas consideradas como óptimas (mínimas, dada la coyuntura de la economía internacional), las cuales también podrían cambiar con el paso del tiempo, todo ello sin que se disparara el déficit público y/o el déficit comercial de la balanza de pagos. En particular, se presenta a continuación un ejemplo de la literatura de problemas de consumo e inversión resuelto mediante el Principio del Máximo.

Ejemplo 4.2.5 (Consumo-Inversión) *Un inversionista desea asignar su riqueza presente x_t entre la inversión a_t y el consumo $(x_t - a_t)$ en cada período $t = 0, 1, 2, ..., N-1$. Suponemos que no se permite pedir préstamos y que la inversión está restringida al conjunto $A(x) = [0, x]$, $x \in [0, \infty) = X$. La conexión entre decisiones de inversión y capital acumulado esta dado por*

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

$t = 0, 1, 2, ..., N-1$*, de modo que la riqueza en el período $t+1$ sea proporcional a lo invertido en el período t. En este caso $\{\xi_t\}$ son v.a.i.i.d., independientes de la riqueza inicial x_0 y con densidad continua y acotada Δ. Además, para asegurar que la reinversión de verdad puede ser provechosa, se asume que el valor esperado de la tasa de interés arbitraria ξ_t es finito y mayor que 1, es decir, $m := E(\xi_t) > 1$. Supóngase que el inversionista está interesado en maximizar su riqueza, bajo la suposición de que su utilidad de consumo está dada por la siguiente función:*

$$r(x_t, a_t) = \frac{b}{\gamma}(x_t - a_t)^{\gamma},$$

$t = 0, 1, 2, ..., N-1$*, donde $0 < \gamma < 1$ y $b > 0$. Se desea maximizar la utilidad descontada total esperada de consumo:*

$$v(\pi, x) = E\left[\sum_{t=0}^{N-1} \alpha^t \frac{b}{\gamma}(x_t - a_t)^{\gamma}\right].$$

Nótese que, para cada $t = 0, 1, 2, ..., N-1$, $V_t(0) = 0$, $f_t(0) = 0$, $V(0) = 0$.

Lema 4.2.6

a) *$A(x)$ es compacto para todo $x \in X$.*

b) *La función de recompensa $r(x,a)$ es semicontinua superiormente en $A(x)$ para cada $x \in X$.*

c) *$w(x,a) := \int_X v(y)Q(dy \mid x,a)$ es continua en $A(x)$ para cada $x \in X$ y cada función v continua y acotada en X.*

Demostración. Claramente a) de este lema es válido ya que $A(x) = [0,x]$, $x \in X$. Por otro lado b) se cumple ya que la función

$$r(x,a) = \frac{b}{\gamma}(x-a)^\gamma,$$

es continua en $\mathbb{K}$.

Finalmente, para verificar c) nótese que la ley de transición Q está dada por,

$$\begin{aligned} w(x,a) &= \int_X v(y)Q\left(dy \mid x,a\right) \\ &= \int_0^\infty v\left(sa\right)\Delta(s)ds, \end{aligned}$$

haciendo un cambio de variable se tiene que

$$w(x,a) = \frac{1}{a}\int_0^\infty v\left(u\right)\Delta(\frac{u}{a})du,$$

para $x \in X$ y $a \in (0,\infty)$. También, nótese que $w(x,0) = v\left(0\right), x \in X$. Como

$$\begin{aligned} \lim_{a\to 0^+} w(x,a) &= \lim_{a\to 0^+}\int_0^\infty v\left(a-s\right)\Delta(s)ds \\ &= \int_0^\infty v\left(0\right)\Delta(s)ds = v(0) = w(x,0), \end{aligned}$$

concluimos que $w(x,a)$ es continua en $A(x)$ para cada $x \in X$, debido a que v y Δ son continuas y acotadas. ■

Lema 4.2.7 *En el ejemplo de consumo inversión existe una única política óptima* $\pi^* = \{f_t\}_{t=0}^{N-1}$ *tal que* $f_t(x) \in int(A(x))$, $x \in (0,\infty)$. *Tal política está dada de la siguiente forma, en* $t = N-1$ *es*

$$f_{N-1}(x) = 0,$$

y para $t = N-2, N-3, ..., 0,$

$$f_t(x) = \frac{x}{1+\delta\kappa_{t+1}^{1/\gamma-1}},\ x \in (0,\infty), \tag{4.8}$$

donde

$$\kappa = (\alpha E(\xi^\gamma))^{1/\gamma-1},$$

$$K_t = \frac{\kappa^{\gamma-1}K_{t+1}}{\left[1+\kappa K_{t+1}^{1/\gamma-1}\right]^{\gamma-1}},$$

y $K_{N-1} = 1$.

Demostración. El Lema 4.2.6 garantiza la existencia de la política óptima debido a que se satisfacen las Hipótesis II. La unicidad de la política se garantizan debido al bloque de condiciones C1 del artículo [22]. A continuación se probará que la política óptima está dada por (4.8) y que ésta a su vez, se encuentra en el interior de $A(x)$, $x \in (0,\infty)$. Aplicando el Teorema 3.2.7 y considerando que el problema es de recompensa terminal cero, se tiene $\phi_N(x) = 0, x \in (0,\infty)$. Luego, en $t = N-1$ como $\phi_{N-1}(x) = V'_{N-1}(x)$, se tiene que $\phi_{N-1}(x) = bx^{\gamma-1}$, $x \in (0,\infty)$, pues en este primer paso el óptimo se alcanza en $f_{N-1}(x) = 0$. En $t = N-2$ se tiene que,

$$\phi_{N-2}(x) = b\,(x - f_{N-2}(x))^{\gamma-1}$$

y

$$\begin{aligned} 0 &= -b\,(x - f_{N-2}(x))^{\gamma-1} + \alpha E\left[b\,(\xi f_{N-2}(x))^{\gamma-1}\xi\right] \\ &= -b\,(x - f_{N-2}(x))^{\gamma-1} + \alpha b\,(f_{N-2}(x))^{\gamma-1} E\left[\xi^\gamma\right], \end{aligned}$$

$x \in (0, \infty)$, de donde

$$f_{N-2}(x) = \frac{x}{1+\kappa}, \tag{4.9}$$

$x \in (0, \infty)$, y

$$\phi_{N-2}(x) = bK_{N-2}x^{\gamma-1},$$

$x \in (0, \infty)$, con

$$K_{N-2} = \frac{\kappa^{\gamma-1}}{(1+\kappa)^{\gamma-1}}, \ \ \kappa = (\alpha E[\xi^\gamma])^{1/\gamma-1}.$$

Ahora se probará que la política (4.9) es óptima interior.

Se tiene que $\widehat{V}_{N-2}(x) = \frac{b}{\gamma}x^\gamma$, y $\widetilde{V}_{N-2}(x) = \frac{b}{\gamma}x^\gamma\alpha E[\xi^\gamma]$, donde $\widehat{V}_{N-2}(x)$ corresponde al valor obtenido en la etapa $N-2$ cuando se utiliza $a^* = 0$ y $\widetilde{V}_{N-2}(x)$ corresponde al valor obtenido en la etapa $N-2$ cuando se utiliza $a^{**} = x$. Comparando las relaciones anteriores con el valor obtenido en la etapa $N-2$ generado por (4.9), denotado por V_{N-2}, se tiene que

$$\widehat{V}_{N-2}(x) < V_{N-2}(x), \text{ y } \widetilde{V}_{N-2}(x) < V_{N-2}(x)$$

$x \in (0, \infty)$. Por lo tanto el maximizador óptimo es $f_{N-2}(x)$.

Supóngase que para $t < N-2$ es válido (4.8), esto es, se cumple

$$\begin{aligned} f_t(x) &= \frac{x}{1+\kappa K_{t+1}^{1/\gamma-1}}, \ x \in (0, \infty), \qquad (4.10) \\ \phi_t(x) &= bK_t x^{\gamma-1}, \ x \in (0, \infty), \end{aligned}$$

donde

$$\begin{aligned} K_t &= \frac{\kappa^{\gamma-1}K_{t+1}}{\left[1+\kappa K_{t+1}^{1/\gamma-1}\right]^{\gamma-1}}, \\ \kappa &= (\alpha E(\xi^\gamma))^{1/\gamma-1}. \end{aligned}$$

Con (4.10) óptimo interior, esto es $\widehat{V}_t(x) < V_t(x)$, y $\widetilde{V}_t(x) < V_t(x)$

Se probará que es cierto también para $t-1$. De las ecuaciones (3.23) y (3.24) se obtiene que

$$\phi_{t-1}(x) = b\left(x - f_{t-1}(x)\right)^{\gamma-1}$$

y

$$\begin{aligned} 0 &= -b\left(x - f_{t-1}(x)\right)^{\gamma-1} + \alpha E\left[bK_t\left(\xi f_{t-1}(x)\right)^{\gamma-1}\xi\right] \\ &= -b\left(x - f_{t-1}(x)\right)^{\gamma-1} + \alpha bK_t\left(f_{t-1}(x)\right)^{\gamma-1}E\left[\xi^{\gamma}\right], \end{aligned}$$

de donde

$$f_{t-1}(x) = \frac{x}{\left[1+\kappa K_t^{1/\gamma-1}\right]},\ x \in (0,\infty) \tag{4.11}$$

y

$$\begin{aligned} \phi_{t-1}(x) &= bx^{\gamma-1}\frac{\kappa^{\gamma-1}K_t}{\left[1+\kappa K_t^{1/\gamma-1}\right]^{\gamma-1}} \\ &= bK_{t-1}x^{\gamma-1}, \end{aligned}$$

$x \in (0,\infty)$, con

$$\begin{aligned} K_{t-1} &= \frac{\kappa^{\gamma-1}K_t}{\left[1+\kappa K_t^{1/\gamma-1}\right]^{\gamma-1}}, \\ \kappa &= \left(\alpha E\left[\xi^{\gamma}\right]\right)^{1/\gamma-1}. \end{aligned}$$

Por otro lado, el valor obtenido con $a^* = 0$ es $\widehat{V}_{t-1}(x) = \frac{b}{\gamma}x^{\gamma}$, con $a^{**} = x$ es $\widetilde{V}_{t-1}(x) = \frac{b}{\gamma}x^{\gamma}K_t\alpha E\left[\xi^{\gamma}\right]$ y con (4.11) es

$$V_{t-1}(x) = \frac{b}{\gamma}x^{\gamma}\frac{\kappa^{\gamma-1}K_t}{\left[1+\kappa K_t^{1/\gamma-1}\right]^{\gamma-1}},$$

$x \in (0,\infty)$. Comparando las relaciones anteriores se tiene que

$$\widehat{V}_{t-1}(x) < V_{t-1}(x)\,,\ \text{y}\ \widetilde{V}_{t-1}(x) < V_{t-1}(x)$$

con $x \in (0,\infty)$. De lo anterior se concluye que (4.11) es óptimo interior. ■

Observación 4.2.8 *La función de valor óptimo se obtiene de sustituir la política óptima en el criterio de rendimiento, como se describe a continua-
ción,*

$$\begin{aligned} V(x) &= \frac{b}{\gamma} E \sum_{t=0}^{N-1} \alpha^t \left[x_t - \frac{x}{1 + \kappa K_{t+1}^{1/\gamma - 1}} \right]^\gamma \\ &= \frac{b}{\gamma} E \sum_{t=0}^{N-1} \alpha^t \left[\frac{x_t + x_t \kappa K_{t+1}^{1/\gamma-1} - x}{1 + \kappa K_{t+1}^{1/\gamma-1}} \right]^\gamma , \end{aligned}$$

con $0 < \gamma < 1$ *y* $b > 0$*.*

Capítulo 5

Conclusiones

En la tesis se trató la teoría de Procesos de Control de Markov (PCMs). En particular, se consideraron problemas a tiempo discreto con horizonte finito. Para evaluar la calidad de las políticas se utilizó el criterio de rendimiento (o función objetivo) de recompensa total descontada esperada.

La teoría de PCMs es presentada en el Capítulo 2, dando los resultados necesarios para el desarrollo subsecuente de la tesis. En éste capítulo se presentó la teoría básica y las hipótesis en el modelo de control necesarias para la existencia de políticas óptimas. También, se presenta el Teorema de Programación Dinámica en el contexto de problemas de control con horizonte finito. En el Capítulo 3 se aborda el problema de control a tiempo discreto mediante el Principio del Máximo (PM). Primeramente se da una versión de este principio para problemas deterministas siguiendo las ideas de Canon et. al. (véase [15] y [16]). Después se trabajó el PM para problemas de control estocástico, en este caso se da una prueba del PM basado en Programación Dinámica (PD), más específicamente, en la diferenciabilidad de las funciones V_t del algoritmo de Programación Dinámica, para ello se estudio inicialmente éste problema. Es importante mencionar que la prueba vía PD ha sido utilizada en distintos contextos para problemas de control, en muchas de ellas se supone la diferenciabilidad de la función de valor óptimo (véase Arkin ([4],

[5]), Clarke [19] y Fleming [25]). A diferencia de estas referencias en este trabajo de tesis se proveen condiciones que garantizan la diferenciabilidad de la función de valor. Finalmente en el Capítulo 4 se proporciona algunos ejemplos donde se muestra el uso del PM propuesto en esta tesis. Estos ejemplos se encuentran relacionados con problemas económicos y lineales cuadráticos.

En resumen los resultados de la tesis se pueden dar en los siguientes puntos:

1. Se han establecido condiciones necesarias de diferenciabilidad en la función de recompensa y la dinámica del sistema.

2. Se ha proporcionado un método complementario a programación dinámica para procesos de control de Markov estocástico con horizonte finito.

3. Se escribió una versión del Principio del Máximo en el contexto de Procesos de Control de Markov considerando un proceso a tiempo discreto, estocástico y con horizonte finito.

4. Se resolvieron algunos ejemplos de la teoría de PCMs mediante el Principio del Máximo.

En consecuencia, los resultados de la presente tesis proporcionan herramientas complementarias a PD para la solución de problemas de control óptimo a tiempo discreto. Las hipótesis que se presentan en el modelo de control de Markov piden condiciones de diferenciabilidad, así para poder
usar el PM se deben verifica dichos supuestos. Existen algunos casos en donde no es posible aplicar el PM, por ejemplo, cuando la dinámica y/o la función de costos no son diferenciables, o bien, cuando no se cumplen las condiciones de interioridad (véase Hernández-Lerma [31] y Zacarías [49]). Sin embargo, existen ejemplos interesantes en la literatura de PCMs donde los supuestos del modelo se verifican (véase Capítulo 4).

Los problemas que se derivan de la tesis son los siguientes:

1. El estudio del PM en problemas estocásticos a tiempo discreto con horizonte infinito.
2. Estudiar relaciones del PM con la Ecuación de Euler.
3. Proveer algoritmos numéricos para la solución del problema de control óptimo a tiempo discreto mediante el PM.

Apéndice A

Apéndice

A.1. Kérneles Estocásticos

Sean X y Y espacios de Borel.

Definición A.1.1 *Un kérnel estocástico en X dado Y, es una función $P(\cdot \mid \cdot)$ tal que:*

a) $P(\cdot \mid y)$ es una medida de probabilidad en X para cada $y \in Y$.

b) $P(B \mid \cdot)$ es una función medible en Y para cada $B \in \mathcal{B}(X)$.

En adelante $\mathcal{P}(X \mid Y)$ denotará el conjunto de todos los kérneles estocásticos en X dado Y, $M(X)$ denota el espacio de funciones medibles en X, $L(X)$ es la subclase de funciones que son semicontinuas y acotadas inferiormente, $M_b(X)$ es el espacio de funciones medibles y acotadas y $C_b(X)$ es el subespacio de funciones continuas y acotadas en X. Sea $P \in \mathcal{P}(X \mid Y)$ un kérnel estocástico dado, entonces se tienen los siguientes resultados, para la demostración de estos véase Bertsekas & Shreve [11], Ash [7].

Proposición A.1.2 *$v \in L(X)$ si y sólo si existe una sucesión de funciones continuas y acotadas v_n en X tal que $v_n \uparrow v$.*

Proposición A.1.3 *Si $v \in M_b(X \times Y)$, entonces la función*

$$g(y) := \int_X v(x, y) P(dx \mid y) \in M_b(Y).$$

Definición A.1.4 *El kérnel estocástico $P \in \mathcal{P}(X \mid Y)$ es*

a) Débilmente continuo si la función $y \to \int_X v(x) P(dx \mid y) \in C_b(Y)$ para cualquier $v \in C_b(X)$.

b) Fuertemente continuo si la función $y \to \int_X v(x) P(dx \mid y) \in C_b(Y)$ para cualquier $v \in M_b(X)$.

Es claro que fuertemente continuo implica débilmente continuo.

Proposición A.1.5 *Las siguientes proposiciones son equivalentes:*

a) P es fuertemente continuo;

b) La función $y \to \int_X v(x) P(dx \mid y)$ es semicontinua inferior, para cada $v \in M_b(X)$.

c) $P(B \mid \cdot)$ es continua sobre Y, para todo $B \in \mathcal{B}(X)$.

Además, las siguientes proposiciones son equivalentes

d) P es débilmente continuo;

e) La función $y \to \int_X v(x) P(dx \mid y)$ es semicontinua inferior, para cada $v \in L(X)$.

La demostración de (d) implica (e) es directa de la Definición A.1.4.

Sea $P \in \mathcal{P}(X \mid Y)$ y supóngase que existe una medida σ-finita m en X tal que, para cada $y \in Y$, $P(\cdot \mid y)$ tiene una densidad $p(\mid y)$ con respecto a m, esto es,

$$P(B \mid y) = \int_B p(x \mid y) m\,(dx)\,,$$

para todo $B \in \mathcal{B}(X)$, $y \in Y$. Si $p\,(x \mid)$ es continua sobre Y para cada $x \in X$, entonces P es fuertemente continua.

Sea $F : X \times Y \rightarrow X$ es una función medible y $\{y_t\}$ una sucesión de variables aleatorias independientes con distribución común μ. Entonces el proceso de Markov $\{x_t\}$ definido por

$$x_{t+1} = F\left(x_t, y_t\right),$$

$t = 0, 1, ...$, con condición inicial x_0 conocido e independiente de $\{y_t\}$ tiene la ley de transición

$$\begin{aligned} P(B \quad | \quad x) &= P(x_{t+1} \in B \mid x_t = x) \\ &= \mu\left(\{y \in Y \mid F\left(x, y\right) \in B\}\right) \\ &= \int_Y I_B\left[F\left(x, y\right)\right] \mu(dy). \end{aligned}$$

Si $F\left(\cdot \mid y\right)$ es continua en X para cada $y \in Y$, entonces P es débilmente continua.

Proposición A.1.6 *Si $P \in P(X \mid Y)$ es débilmente continuo y $v \in C_b(X \times Y)$ (respectivamente semicontinua y acotada inferiormente), entonces la función $y \rightarrow \int v(x)P(dx \mid y)$ es continua y acotada (resp. es l.s.c. y acotada inferiormente) sobre Y.*

Proposición A.1.7 ***(Teorema de Ionescu-Tulcea).*** *Sea $X_0, X_1, ...,$ una sucesión de espacios de Borel y, para $n = 0, 1, ...,$ definimos $Y_n := X_0 \times X_1 \times ... \times X_n$ y $Y := \prod_{n=0}^{\infty} X_n$. Sea v una medida de probabilidad arbitraria sobre X_0 y, para cada $n = 0, 1, ...,$ $P_n(dx_{n+1} \mid y_n)$ es un kérnel estocástico sobre X_{n+1} dado Y_n. Entonces existe una única medida de probabilidad P_v sobre Y tal que, para cada rectángulo medible $B_0 \times ... \times B_n$ en Y_n,*

$$\begin{aligned} P_v(B_0 \times ... \times B_n) \quad &= \quad \int_{B_0} v(dx_0) \int_{B_1} P_0(dx_1 \mid x_0) \int_{B_2} P_1(dx_2 \mid x_0, x_1) \\ &\quad ... \int_{Bn} P_{n-1}(dx_n \quad | \quad x_0, ..., x_{n-1}). \end{aligned}$$

Además, para cualquier función u *medible y no negativa sobre* Y*, la función*

$$x \to \int u(y) P_x(dy)$$

es medible en X_0*, donde* P_x *representa a* P_v *cuando* v *es la probabilidad concentrada en* $x \in X_0$*.*

A.2. Algunos Resultados sobre Análisis Convexo

Notación:

$\overline{A}$	Cerradura de $A \subset \mathbb{R}^n$
coA	cubierta convexa de A
$\overline{co}A$	cubierta convexa cerrada de A
$\langle \cdot, \cdot \rangle$	producto escalar
$[\cdot]^T$	Transpuesta
C	Cono
Ω	Conjunto de restricciones
$C(\widehat{z}, \Omega)$	Aproximación cónica a Ω en $\widehat{z}$
$IC(\widehat{z}, \Omega)$	Cono interno a Ω en $\widehat{z}$
$rel\ int\ A$	Interior relativo de A

Definición A.2.1 *Dos conjuntos $A_1, A_2 \subset \mathbb{R}^n$ son separados si existe un hiperplano P tal que A_1 está contenido en uno de los semiespacios, considerando a este cerrado, determinado por P y A_2 está contenido en el otro semiespacio, también cerrado, determinado por P.*

Definición A.2.2 *La cubierta convexa (coA) de un número finito de puntos $x_1, x_2, ..., x_N$ en $\mathbb{R}^n$ está definido como el conjunto*

$$\left\{ x : x = \sum_{i=1}^{N} \mu_i x_i, \quad \mu_i \geq 0, \quad \sum_{i=1}^{N} \mu_i = 1 \right\}.$$

Tal conjunto es llamado poliedro convexo.

Es también posible definir la cerradura o cubierta convexa de un subconjunto arbitrario de $\mathbb{R}^n$.

Definición A.2.3 *La cubierta convexa de $B \subset \mathbb{R}^n$ es la intersección de todos los conjuntos convexos que contienen a B.*

Definición A.2.4 *Un cono C en $\mathbb{R}^n$ es un conjunto de puntos tales que si $x \in C$, entonces $\alpha x \in C$ para todo $\alpha > 0$.*

Definición A.2.5 *Un cono C es llamado un cono convexo si C es un conjunto convexo.*

Teorema A.2.6 *Un cono C es convexo si y sólo si $x_1 + x_2 \in C$ siempre que $x_1, x_2 \in C$.*

Definición A.2.7 *Un cono convexo $C(z, \Omega) \subset \mathbb{R}^n$ será llamada una*
aproximación cónica al conjunto Ω en el punto $z \in \Omega$ si para cualquier colección $\{z_1, z_2, ..., z_N\}$ de vectores linealmente independientes en $C(z, \Omega)$ existe un $\epsilon > 0$ posiblemente dependiendo de z y $\{z_1, z_2, ..., z_N\}$ tal que la cubierta (cerradura) convexa de $\{z, z + \epsilon z_1, z + \epsilon z_2, ..., z + \epsilon z_N\}$ está completamente contenida en Ω.

Definición A.2.8 *Sea $\Omega = \{z : q(z) \leq 0\}$ donde $q : \mathbb{R}^n \to \mathbb{R}^k$ es una función continuamente diferenciable. Para cualquier $\widehat{z} \in \Omega$ el cono interno a Ω en $\widehat{z}$, denotado por $IC(\widehat{z}, \Omega)$, se define por*

$$IC(\widehat{z}, \Omega) = \{z' : \langle \nabla q(\widehat{z}), z' \rangle \leq 0 \ \text{tal que} \ q(\widehat{z}) = 0\} .$$

Lema A.2.9 *Sea $\Omega = \{z : q(z) \leq 0\}$ donde $q : \mathbb{R}^n \to \mathbb{R}^k$ es una función continuamente diferenciable. Para cualquier $\widehat{z} \in \Omega$ distinto al origen, el cono interno $IC(\widehat{z}, \Omega)$ es una aproximación cónica.*

Definición A.2.10 *Sea $X \subset \mathbb{R}^n$ un subconjunto convexo k-dimensional. El interior relativo (rel int A) de X se define como el interior de X relativo a la variedad lineal k-dimensional $\zeta(X)$ que lo contiene, es decir, x pertenece al interior relativo de X si existe una bola abierta centrada en x, $B(x)$ tal que $\zeta(X) \cap B(x) \subset X$.*

El interior relativo de cualquier conjunto convexo siempre es no vacío y por lo tanto convexo.

Teorema A.2.11 *Considere los conos convexos* $K_1, \ K_2 \subset \mathbb{R}^n$*. Si*

a) $K_1 \cup K_2$ *puede estar contenido en un hiperplano* $(n-1)$ *dimensional,*

b) *rel int* $K_1 \cap$ *rel int* $K_2 = \varnothing$,

entonces K_1 *y* K_2 *son separados.*

Lema A.2.12 ***Lema de Farkas.*** *Si* $a_1, ..., a_k$ *y* b *es un conjunto finito de vectores en* $\mathbb{R}^n$*, entonces*

$$\langle b, x\rangle \leq 0$$

para todo $x \in \mathbb{R}^n$ *que satisface*

$$\langle a_i, x\rangle \leq 0,$$

$i = 1, 2, ..., k$ *si y sólo si*

$$b = \sum_{i=1}^{k} \mu^i a_i,$$

con $\mu^i \geq 0$.

Bibliografía

[1] Aliprantis CD, Burkinshaw O., Principles of Real Analysis. academic, Inc. Third Edition, 1998.

[2] Araujo A. & Scheinkman J. A., Maximum principle and transversality condition for concave infinite horizon economic models, Journal of Economic Theory, vol. 30, pp. 1-16, 1983.

[3] Araujo A. & Sheinkman JA Smoothness, comparative dynamics, and the turnpike property. Econometrica, vol. 45, no. 3, pp. 601-620, 1977.

[4] Arkin V. I. & Evstigneev I. V., Stochastic Models of Control and Economic Dynamics, Academic Press Inc., ISBN 978-0120620807, 1987.

[5] Arkin V. I. & Krechetov L. M., A Stochastic maximum principle in control problems with discrete time, Lectures Notes in Mathematics (In Proceedings of the third Japan-USSR Symposium on Probability Theory), no. 550, pp. 692-712, Springer Verlag, 1976.

[6] Arutyunov A. V., Pontryagin's maximum principle in optimal control theory, Journal of Mathematical Sciences, vol. 94, no. 3, pp. 1311-1364, 1999.

[7] Ash R. B., Real Analysis and Probability, Academic Press, 1972.

[8] Basile N. & Mininni M., An Extension of the maximum principle for a class of control problems in infinite-dimensional spaces, SIAM Journal on Control and Optimization, vol. 28, pp. 1113-1135, 1990.

[9] Benveniste L. M. & Scheinkman J. A., On the differentiability of the value function in dynamic models of economics, Econometrica, vol. 47, no. 3, pp. 727-732, 1979.

[10] Bertsekas D. P., Dynamic Programming and Optimal Control, Athena Scientific, Belmont, Massachusetts, vol. 1, ISBN 1-886-52913-2, 1995.

[11] Bertsekas D. P. & Shreve S.E, Stochastic Optimal Control: The Discrete Time Case, Academic Press, New York, ISBN 1-886529-03-5, 1978.

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

[13] Burmeister E. & Dobell A. R., Mathematical Theories of Economic Growth, Macmillan Publishing Co., Inc., 1970.

[14] Cadenillas A. & Karatzas I., The stochastic maximum principle for linear, convex optimal control with random coefficients, SIAM Journal on Control and Optimization, vol. 33, no. 2, pp. 590-624, 1995.

[15] Canon M. D., Cullum C. D. & Pollak E., Constrained minimization problems in finite-dimensional spaces, SIAM Journal on Control and Optimization, vol. 4, no.3, pp. 528-547, 1966.

[16] Canon M. D., Cullum C. D. & Pollak E., Theory of Optimal Control and Mathematical Programming, Mcgraw-Hill, ISBN 978-0070097605, 1970.

[17] Caputo M. R., Foundations of Dynamic Economic Analysis, Cambridge University Press, 2005.

[18] Clark C. W, Mathematical Bioeconomics, John Wiley & Sons, Inc., ISBN 978-0471751526, 1990.

[19] Clarke F. H. & Vinter R. B., The relationship between the maximum principle and dynamic programming, SIAM Journal on Control and Optimization, vol. 25, pp. 1291-1311, 1987.

[20] Cruz-Suárez H. & Montes-de-Oca R., Discounted Markov control processes induced by deterministic systems. Kybernetika, vol. 42, no. 6, pp. 647-665, 2006.

[21] Cruz-Suárez H. & Montes-de-Oca R., An envelope theorem and some applications to discounted Markov decision processes, Mathematical Methods of Operations Research, Springer Verlag, vol. 67, no. 2, pp. 299-321, 2008.

[22] Cruz-Suárez D., Montes-de-Oca R. & Salem-Silva F., Conditions for the uniqueness of optimal policies of discounted Markov decision processes, Mathematical Methods of Operations Research, Springer Verlag, pp. 415-436, 2004.

[23] Fan L.-T. & Wang C.-S., The Discrete Maximum Principle, New York: Wiley, 1974.

[24] Fattorini H. O., Existence theory and the maximum principle for relaxed infinite-dimensional optimal control problems, SIAM Journal on Control and Optimization, vol. 32, no. 2, pp. 311-331, 1994.

[25] Fleming W. H. & Rishel R. W., Deterministic and Stochastic Optimal Control, Springer Verlag, 1975.

[26] Halkin H., A maximum principle of Pontriagyn type for systems described by nonlinear difference equations, SIAM

Journal on Control and Optimization, vol. 4, pp. 90-111, 1966.

[27] Halkin H., Necessary conditions for optimal control problems with infinite horizons, Econometrica, vol. 42, no. 2, pp. 267-272, 1974.

[28] Haurie A. & Sethi S. P., Decision and forecast horizons, agreeable plans and the maximum principle for infinite horizon control problems, Operations Research Letters, vol. 3, no. 5, pp. 261-265, 1985.

[29] Hartl R. F., Sethi S. P. & Vickson R. G., A survey of the maximum principle for optimal control problems with state constraints, SIAM Review, vol. 37, no. 2, pp. 181-218, 1995.

[30] Haussmann U. G., Some examples of optimal stochastic controls OR: The stochastic maximum principle at work, SIAM Review, vol. 23, no. 3, pp. 292-307, 1981.

[31] Hernández-Lerma O. & Lasserre J. B., Discrete Time Markov Control Processes. Springer Verlag, ISBN 0-387-94579-2, 1996.

[32] Holtzman J. M., Convexity and the maximum principle for discrete systems, IEEE Transactions on Automatic Control, vol. 4, pp. 30-35, 1966.

[33] Holtzman J. M., On the maximum principle for nonlinear discrete time systems, IEEE Transactions on Automatic Control, vol. 4, pp. 273-274, 1966.

[34] Intriligator M. D., Mathematical Optimization and Economic Theory, Prentice Hall, 1971.

[35] Medhin N. G., Stochastic maximum principle a minimizing sequence approach, Stochastic Analysis and Applications, vol. 16, no. 1, pp.107-118, 1998.

[36] Peng S., A general stochastic maximum principle for optimal control problems, SIAM Journal on Control and Optimization, vol. 28, no. 4, pp. 966-979, 1990.

[37] Pham T. N., Feedback under control-dependent Markov disturbances: A discrete maximum principle, Journal of Mathematical Analysis and Applications, vol. 197, pp. 325-340, 1996.

[38] Pontryagin L. S., Boltyanskii V. G., Gamkrelidze R. V. & Mishchenko E. F., The mathematical theory of optimal processes, Interscience Publishers, 1962.

[39] Puterman M. L., Markov Decision Processes: Discrete Stochastic Dynamic Programming, Wiley, New York, ISBN 978-0471727828, 1994.

[40] Santos M. S., Smoothness of the policy function in discrete time economic models, Econometrica, vol. 59, no. 5, pp. 1365-1382, 1991.

[41] Santos M. S., Smooth dynamics and computation in models of economic growth, Journal of economic dynamics and control, vol. 18, pp. 879-895, 1994.

[42] Santos M. S., Numerical solution of dynamic economic models, Handbook of Macroeconomic, vol. 1 (Editors J. B. Taylor and M. Woodford), Elsevier Science B. V., 1999.

[43] Shell K., Applications of Pontryagin's Maximum Principle to Economics, Lectures Notes in Operations Research and Mathematical Economics vol. 1, (Editors H. N. Kuhn and G. P. Szegö), Springer Verlag, 1969.

[44] Tang S. & Hou S., Optimal control of point processes with noisy observations: The Maximum Principle, Applied Mathematics and Optimization vol. 45, pp. 185-212, 2002.

[45] Urzula L. & Heinz S., A high-order generalized local maximum principle, SIAM J. Control and Optimization, vol. 38, no. 3, pp. 823-854, 2000.

[46] Van Long N. & Shimomura K., A new proof of maximum principle, Economic Theory, vol. 22, pp. 671-674, 2003.

[47] Xu W. S., Maximum principle for a stochastic optimal control problem and application to portfolio/consumption choice, J. Optimization Theory Applications, vol. 98, pp. 719-731, 1998.

[48] Yong J. & Xu Y. Z., Stochastic Controls (Hamiltonian Systems and HJB Equations), Springer Verlag, 1999.

[49] Zacarías G. & González de la Palma L., Un ejemplo de control de inventarios, Aportaciones y Aplicaciones de la Probabilidad y la Estadística, Benemérita Universidad Autónoma de Puebla, Dirección de Fomento Editorial, 2006.