UNIVERSIDAD AUTÓNOMA DE CHIAPAS
FACULTAD DE CIENCIAS EN FÍSICA Y MATEMÁTICAS

Control óptimo de procesos de Markov a tiempo discreto con restricciones. Del caso descontado al caso promedio.

TESIS

que para obtener el grado de Maestro en Ciencias en la especialidad de Matemáticas, presenta

Omar Antonio de la Cruz Courtois
Enero 25, 2016

Asesores
DR. ARMANDO FELIPE MENDOZA PÉREZ
DR. HÉCTOR JASSO FUENTES

Universidad Autónoma de Chiapas

Facultad de Ciencias en Física y Matemáticas

Dirección

Tuxtla Gutiérrez, Chiapas
15 de enero de 2016
Oficio No. FCFM/0018/16

Dr. Armando Felipe Mendoza Pérez
Presidente y Director de Tesis
Presente

Por este medio me permito informarle que una vez efectuada la revisión de la tesis denominada:

"Control óptimo de procesos de Markov a tiempo discreto con restricciones. Del caso descontado al caso promedio".

Ha sido aceptada para sustentar el Examen de Grado de Maestro en Ciencias Matemáticas del **C. Omar Antonio De la Cruz Courtois** con matrícula escolar: 14117001

Se autoriza su impresión en virtud de cumplir con los requisitos correspondientes.

Atentamente
"Por la conciencia de la necesidad de servir"

Dr. Sendic Estrada Jiménez
Director

DIRECCIÓN
FCFM

C.c.p. Dr. Florencio Corona Vázquez, Secretario Académico de la FCFM.
Lic. Ana Gabriel Aguilar Avendaño. Encargada de Servicios Escolares de la FCFM
Archivo / Minutario
SEJ /agv

Por que en la mucha sabiduría
hay mucha molestia;
y quien añade ciencia,
añade dolor.

Eclesiastés, *1.18.*

Contenido

Dedicatoria II

Abreviaciones VI

Notación VII

1. Introducción **1**
1.1. Introducción 1
1.2. Resumen 3
1.3. Preliminares 4
1.4. La construcción canónica 6
1.5. Normas de peso W 8

2. Control descontado con restricciones de costo **10**
2.1. Introducción 10
2.2. El modelo de control descontado con restricciones 10
2.3. La ecuación de Poisson α-descontada 13
2.4. El problema descontado con restricciones (PDR) 17

3. Aproximación por multiplicadores de Lagrange **19**
3.1. Introducción 19
3.2. El problema descontado sin restricciones (PDSR) 20
3.3. Familia paramétrica de α-EGDO y el PDR 21
3.4. Existencia de políticas óptimas para el PDR 24

4. El caso promedio con restricciones **26**
4.1. Introducción 26
4.2. El problema de control de markoviano esperado con restricciones 26
4.3. El problema esperado con restricciones generalizado 29

4.4. La aproximación descontada desvanescente 33

5. Ejemplo: Sistema cuadrático-lineal LQ **43**
5.1. Hipótesis y resultados importantes del sistema LQ con restricciones 43
5.2. Ejemplo numérico 48

6. Conclusiones **51**

References **52**

Index **56**

Abreviaciones

i.i.d.	independiente e idénticamente distribuida
s.c.i.	semicontinua inferior
s.c.s.	semicontinua superior
PCMs	Procesos de Control de Markov
MCM	Modelo de control de Markov
PR	problema con restricciones
PER	Problema esperado con restricciones
PDR	Problema descontado con restricciones
PDSR	Problema descontado sin restricciones
α-EGDO	Ecuación de ganancia α-descontada óptima
E.P.	Ecuación de Poisson
LQ	Sistema cuadrático lineal

Notación

$\blacksquare$	fin de la demostración
$:=$	igual por definición
1_B	la función indicadora de un conjunto B
$\mathbb{N}$	el conjunto de números naturales $\{1, 2, \cdots\}$
$\mathbb{R}$	el conjunto de números reales
$\mathbb{K}$	el conjunto de pares de estado-accciones admisibles
φ	política aleatorizada estacionaria
Φ_s	conjunto de políticas aleatorizadas estacionarias
$\mathbb{F}$	conjunto de funciones selectoras
X	el espacio de Borel (Estados)
$\mathcal{B}(X)$	la σ-algebra de Borel de subconjuntos de X
$C_b(X)$	El espacio de Banach de funciones continuas acotadas sobre $\mathbf{X}$
$B_b(X)$	El espacio de Banach de funciones medibles, acotadas sobre $\mathcal{B}(\mathbf{X})$
$B_W(X)$	El espacio de Banach de las funciones medibles W-acotadas

Capítulo 1

Introducción

1.1. Introducción

Dentro del control estocástico hay dos criterios a horizonte infinito en el que se enmarcan los problemas de control óptimo: Estos son por un lado los criterios descontados, y por el otro, los criterios de promedio ergódico. Está ampliamente reconocido que estos dos criterios se comportan de manera diferente aunque ambos son complementarios; por ejemplo los criterios descontados concentran su desempeño en los períodos de tiempo tempranos, ya que la ganancia (o el costo) disminuye a largos períodos de tiempo; en el caso de los criterios de promedio ergódico estos se determinan por el comportamiento asintótico que no toma en cuenta intervalos finitos de tiempo.

La relación entre criterios descontados y promedios ha sido estudiada principalmente a través de los multicitados teoremas Abelianos, relacionando un funcional promedio con su correspondiente transformada de Laplace. Un problema que resolvemos en esta tesis es el de proporcionar condiciones que garanticen la aproximación a través de ganancias descontadas con restricciones en costos descontados, hacia la contraparte formada por la ganancia promedio con restricciones en costos promedios. Este método de aproximación se conoce como *método descontado desvanecente* y ha sido aplicado para estudiar una amplia variedad de sistemas controlados (véase las referencias que se citan en [22]).

El presente trabajo de tesis trata sobre los procesos de control markoviano a tiempo discreto con restricciones en espacios Borel medibles. El criterio para optimizar estos procesos es a través de la ganancia esperada promedio y la ganancia descontada promedio sujeto a restricciones sobre un número finito de costos esperados promedios y costos descontados promedios, respectivamente. Estos problemas aparecen en diversas ramas de las matemáticas y también forman una clase importante de problemas en control estocástico con aplicaciones en distintas áreas, incluyendo

economía, líneas de espera, procesos epidemiológicos, etc., para ello se puede consultar [4, 10, 16, 28], así como los libros [1, 2]. El artículo de Dufour y Stockbridge [8] considera problemas de control con restricciones para una clase de problemas de control markoviano a tiempo continuo con el criterio de costos descontados. Debemos mencionar también que Chen y Feinberg [6], Chang [5] y Lyer y Hamachandra [23] han realizado estudios sobre lo anteriormente descrito. Una característica común de todos estos trabajos es que todo ellos conciernen a los Problemas de Control Markoviano con criterios de descuento y en estados de espacio finitos.

Hay varias técnicas para analizar los problemas con restricciones. La más común es la técnica conocida como *método directo* (veáse, entre otros, [1, 4, 11, 17, 18, 24, 27]). En este método la idea es usar medidas de ocupación para transformar el problema con restricciones original en un problema de optimización equivalente, definido sobre un espacio de medidas adecuado, para así poder usar el resultado conocido de que una función semicontinua inferior (resp. semicontinua superior) sobre un espacio topológico compacto, alcanza su valor mínimo (resp. valor máximo). Un segundo método se basa en el uso de técnicas de la *programación lineal* (por ejemplo, se pueden citar [1, 14, 17, 27]), en donde se introducen espacios vectoriales de medidas y funciones sobre los cuales el problema original se puede expresar como un programa lineal. Más aún, podemos reescribir el problema con restricciones como un programa convexo, o bien, como un problema de control de Markov sin restricciones usando métodos analítico-convexos o por multiplicadores de Lagrange (ver, por ejemplo, [3, 13, 26, 27, 33]). Nuestro trabajo de tesis utilizará en parte esta técnica como describimos más abajo.

Por otro lado, las técnicas de *vanishing* (o desvanecimiento) descontadas es un procedimiento general para resolver los problemas de control óptimo de ganancia y costos esperados promedio a través de problemas α-descontados cuando el factor de descuento α tiende a uno. En este caso, bajo las hipótesis adecuadas, se puede probar que el límite de las políticas óptimas α-descontadas son también óptimas para el caso promedio esperado. La aproximación descontada desvanecente ha sido ampliamente aplicada para los modelos de control a tiempo discreto; podemos mencionar los artículos [7, 9, 10, 15, 32]. Para problemas con restricciones a tiempo continuo se puede consultar [15, 29].

El aporte de esta tesis consiste en introducir dos métodos para calcular políticas óptimas con restricciones para el problema promedio descontado y esperado, es decir, cuando la función de costo es dominada por otra función (en particular una constante). Para deducir los resultados de optimalidad para el caso descontado, usamos técnicas de multiplicadores de Lagrange que conducen a ciertas familias paramétricas de ecuaciones de optimalidad descontadas sin restricciones. Posteriormente, basados en la teoría de programación dinámica así como el uso del cálculo diferencial elemental, obtendremos un multiplicador de lagrange que es punto crítico de una función

de optimalidad descontada sin restricciones, así como una familia de políticas óptimas de control asociadas a esta función de optimalidad en dicho punto crítico, que en consecuencia, resultarán ser políticas óptimas para nuestro problema original descontado con restricciones. Posteriormente, la aproximación descontada desvanecente nos proveerá una prueba elemental de la existencia de políticas óptimas para el problema con restricciones promedio esperado. De aquí, obtenemos la solución al problema con restricciones esperado (PRE). Cabe mencionar que el presente trabajo es una extensión para el caso de cadenas de Markov a tiempo discreto, de los resultados obtenidos en [22] para difusiones controladas con restricciones. En dicho trabajo, los autores utilizan las propiedades del espacio topológico compacto de las políticas aleatorizadas como lo es la convergencia de políticas así como propiedades de elipticidad uniforme, condiciones de tipo Lipschitz, etc. que determinarán propiedades de regularidad del modelo. En nuestro trabajo de tesis prescindimos de estas propiedades, sin embargo mantenemos la W-ergodicidad geométrica. También hacemos notar que en esta tesis se resuelve el problema con restricciones promedio esperadas imponiendo hipótesis más generales y menos restrictivas que las dadas en los trabajos previos [24, 25], y además, de acuerdo a nuestro conocimiento, el análisis propuesto en esta tesis es original y no ha sido previamente estudiado en el contexto de los problemas de control de Markov a tiempo discreto.

1.2. Resumen

El material en esta tesis está organizado como sigue: En la continuación de este capítulo introducimos todos los conceptos esenciales para los modelos de control de Markov (Sección 1.3 y sección 1.4).
En el capítulo 2 enunciamos algunas de nuestras principales hipótesis con las que trabajaremos el resto de la tesis, así como los criterios de optimalidad y estableceremos el problema de control descontado con restricciones de costo. En el capítulo 3 damos un método basado en los multiplicadores de Lagrange bajo el cual las ecuaciones de optimalidad para el problema de control descontado con restricciones son equivalentes a las ecuaciones de optimalidad para el problema sin restricciones.
En el capítulo 4 consideramos el estudio de las ecuaciones de optimalidad para el problema promedio esperado con restricciones. Para este propósito, aplicamos la aproximación descontada desvanecente al problema descontado sin restricciones asociado (el cual depende por supuesto, de los multiplicadores de Lagrange); de aquí, haciendo tender ciertos factores de descuento a uno, obtendremos que el límite de una sucesión de "multiplicadores deseado" resolverán nuestro problema ergódico promedio con restricciones.
Finalmente, en el capítulo 5 ilustramos los resultados mediante un sistema cuadrático lineal LQ.

1.3. Preliminares

Definición 1.3.1 *Un modelo de control de Markov con restricciones es un sexteto*

$$(X, A, \{A(x) : x \in X\}, Q, c, r),$$

donde X es el espacio de estados y A es el espacio de control o conjunto de acciones, siendo ambos espacios medibles de Borel polacos, con $\mathcal{B}(X)$ y $\mathcal{B}(A)$ las σ-álgebras de Borel correspondientes. $\{A(x) : x \in X\}$ es una familia no vacía de subconjuntos $A(x) \in \mathcal{B}(A)$, donde $A(x)$ es el conjunto de acciones admisibles dado el estado $x \in X$. Además $Q = \{Q(B|x, a) : B \in \mathcal{B}(X), (x, a) \in X \times A\}$, la ley de transición entre estados es un kernel estocástico en X dado (X, A). Las funciones $r : X \times A \to \mathbb{R}$ y $c : X \times A \to \mathbb{R}$ son medibles y corresponderán a las llamadas función de ganancia y función de costo, respectivamente.

Denotamos por

$$\mathbb{K} := \{(x, a)|\ x \in X,\ a \in A(x)\}, \tag{1.3.1}$$

como el conjunto de pares de estados y acciones admisibles, el cual es un subconjunto medible de $X \times A$. Con esta notación, tenemos que la función de ganancia y costo serán funciones medibles $r : \mathbb{K} \to \mathbb{R}$ y $c : \mathbb{K} \to \mathbb{R}$, respectivamente.

Definición 1.3.2 *Consideremos el modelo de control de Markov en la Definición 1.3.1. Para cada $f = 0, 1, \ldots$, definimos el espacio H_t de historias admisibles al tiempo t como $H_0 := X$, y*

$$H_t := \mathbb{K}^t \times X = \mathbb{K} \times H_{t-1}, \quad t = 1, 2, \ldots,$$

donde $\mathbb{K}$ es el conjunto definido en (1.3.1). Los elementos h_t de H_t se llaman t-historias admisibles, o simplemente, t-historias, y constituyen arreglos de la forma

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

donde $(x_i, a_i) \in \mathbb{K}$ para $i = 0, \ldots, t-1$, $x_t \in X$. Para cada t, H_t es un subconjunto de

$$\overline{H_t} := (X \times A)^t \times X = (X \times A) \times \overline{H}_{t-1} \quad t = 1, 2, \ldots,$$

con $\overline{H}_0 := H_0 = X$. Además denotaremos por $\overline{H}_\infty := (X \times A)^\infty$.

Definición 1.3.3 *Sea $\mathbb{F}$ el conjunto de todas las funciones de decisión o funciones selectoras, es decir, $f : X \to A$ satisface $f(x) \in A(x)$ para toda $x \in X$. También denotaremos por Φ_s al conjunto de kerneles estocásticos φ sobre A dado X, para los cuales $\varphi(A(x)|x) = 1$. Una función selectora $f \in \mathbb{F}$ puede ser identificada con el kernel estocástico $\varphi \in \Phi_s$ para el cual $\varphi(\cdot|x)$ es la medida de Dirac concentrada en $f(x)$ para toda $x \in X$. De aquí, al hacer esta identificación, tenemos que $\mathbb{F} \subset \Phi_s$.*

Asumiremos que $\mathbb{F}$ es no vacío, o equivalentemente, que el conjunto $\mathbb{K}$ contiene la gráfica de una función medible de X a A. Esta hipótesis permitirá que el conjunto de políticas de control definidas abajo sea no vacío.

Definición 1.3.4 *Una política de control (aleatorizada) es una sucesión $\pi = \{\pi_t\}$ de kerneles estocásticos π_t sobre A dado H_t tal que*

$$\pi_t(A(x_t)|h_t) = 1,$$

para cada t-historia $h_t = (x_0, a_0, \ldots, x_{t-1}, a_{t-1}, x_t)$ en H_t. Denotaremos por Π al conjunto de todas las políticas de control. Más aún, una política de control $\pi = \{\pi_t\}$ se dice ser una

(a) *política Markov aleatorizada si existe una sucesión $\{\varphi_t\}$ de kerneles estocásticos $\varphi_t \in \Phi_s$ tal que*

$$\pi_t(\cdot|h_t) = \varphi(\cdot|x_t) \quad \forall h_t \in H_t, \quad t \in \mathbb{N}_0;$$

donde $\mathbb{N}_0 := \{0, 1, 2, 3 \ldots\}$.

(b) *política estacionaria (aleatorizada) si existe un kernel estocástico $\varphi \in \Phi_s$ tal que*

$$\pi_t(\cdot|h_t) = \varphi(\cdot|x_t) \quad \forall h_t \in H_t, \quad t \in \mathbb{N}_0;$$

(c) *política determinista estacionaria si existe una función selectora $f \in \mathbb{F}$ tal que $\pi(\cdot|h_t)$ es la medida de Dirac en $f(x_t) \in A(x_t)$ para cada $h_t \in H_t$, $t \in \mathbb{N}_0$.*

El conjunto de todas las políticas de Markov aleatorizadas se denotará por Π_{RM}. El conjunto de políticas estacionarias es denotado por Φ_s. También $\mathbb{F}$ denotará el conjunto de políticas estacionarias deterministas. Notemos que $\mathbb{F} \subset \Phi_s \subset \Pi_{RM} \subset \Pi$. Si $\pi = \{\varphi\}$ es una política estacionaria, la denotamos por $\pi = \varphi$.

1.4. La construcción canónica

Sea $(\Omega, \mathcal{F})$ el espacio medible (canónico) que consiste del espacio $\Omega := (X \times A)^\infty$ y $\mathcal{F}$ la σ-álgebra producto correspondiente. Los elementos de Ω son sucesiones de la forma $\omega = (x_0, a_0, x_1, a_1, \ldots)$ con x_n en X y a_n en A para toda $n = 0, 1, \ldots$; las proyecciones x_n y a_n de Ω a los conjuntos X y A son llamados las variables de estado y control (o acciones), respectivamente. Observemos que Ω contiene al espacio $H_\infty := \mathbb{K}^\infty$ de historias admisibles $(x_0, a_0, x_1, a_1, \ldots)$ con $(x_n, a_n) \in \mathbb{K}$ para cada $n \in \mathbb{N}_0$. Además, dada una política arbitraria en Π, podemos construir un espacio de probabilidad $(\Omega, \mathcal{F}, P_\nu^\pi)$, así como un proceso $(\{(x_t, a_t)\})$ safisfaciendo el siguiente teorema.

Teorema 1.4.1 *Sea $(\Omega, \mathcal{F})$ un espacio medible con $\Omega = (X \times A)^\infty$ donde $\mathcal{F}$ la σ-álgebra producto correspondiente. Sea $\pi = \{\pi_t\}$ una política de control, ν una medida de probabilidad sobre $(X, \mathcal{B}(X))$, llamada la distribución inicial. Entonces existe una única medida de probabilidad P_ν^π sobre $(\Omega, \mathcal{F})$, así como sendos procesos estocásticos $\{x_t\}$, $\{a_t\}$ (el proceso de estados y el proceso de acciones o control, respectivamente) tal que $P_\nu^\pi(H_\infty) = 1$ y además para toda $B \in \mathcal{B}(X), C \in \mathcal{B}(A)$ y $h_t \in H_t, t = 0, 1, \ldots$*

i) $P_\nu^\pi(x_0 \in B) = \nu(B)$

ii) $P_\nu^\pi(a_t \in C | h_t) = \pi_t(C|h_t) \quad P_\nu^\pi$ *c.s.*

iii) $P_\nu^\pi(x_{t+1} \in B | h_t, a_t) = Q(B|x_t, a_t) \quad P_\nu^\pi$ *c.s.*

Demostración Véase [19, Proposición C.10 y observación C.11]. ∎

Observación 1.4.1 *La siguiente notación es estándar y la emplearemos en el resto de la tesis.*

(a) El operador esperanza con respecto a P_ν^π es denotado por E_ν^π. Si ν se concentra en el estado inicial $x \in X$, luego escribimos P_ν^π y E_ν^π como P_x^π y P_x^π, respectivamente. Más aún, si $\pi = \varphi$ es una política estacionaria, luego denotaremos P_ν^π y E_ν^π como P_ν^φ y E_ν^φ, respectivamente.

(b) Sea $\varphi \in \Phi_S$ un kernel estocástico sobre A dado X, c y r las funciones costo y de ganancia, y Q la ley de transición. Luego definimos, para cada $x \in X$,

$$c_\varphi(x) := \int_A c(x, a)\varphi(da|x), \; r_\varphi(x) := \int_A r(x, a)\varphi(da|x), \tag{1.4.1}$$

y

$$Q_\varphi(\cdot|x) := \int_A Q(\cdot|x,a)\varphi(da|x). \tag{1.4.2}$$

En particular, para una función $f \in \mathbb{F}$, las ecuaciones anteriores se convierten en

$$c_f(x) = c(x, f(x)),\ r_f(x) = r(x, f(x)), \quad y\ Q_f(B|x) = Q(B|x, f(x)). \tag{1.4.3}$$

(c) Sea $\pi = \varphi$ una política estacionaria. La probabilidad de transición en el n-ésimo paso será denotada por Q^n_φ, esto es

$$Q^n_\varphi(B|x) := P^\varphi_x(x_n \in B),\ n \in \mathbb{N}_0,\ B \in \mathcal{B}(X),\ x \in X,$$

con $Q^1_\varphi(\cdot|x) := Q_\varphi(\cdot|x)$, $Q^0_\varphi(\cdot|x) = \delta_x$ la medida de Dirac concentrada en el estado inicial x. Además, podemos escribir Q^n_φ recursivamente como

$$\begin{aligned} Q^n_\varphi(B|x) &= \int_X Q_\varphi(B|y)Q^{n-1}_\varphi(dy|x) \\ &= \int_X Q^{n-1}_\varphi(B|y)Q_\varphi(dy|x),\ n \geq 1. \end{aligned} \tag{1.4.4}$$

En la siguiente proposición, veremos que si se escoge una política de Markov aleatorizada entonces el proceso de estados es de Markov.

Proposición 1.4.1 *Sea ν una distribución inicial en $(X, \mathcal{B}(X))$. Si $\pi = \{\varphi_t\} \in \Pi_{RM}$ es una política de Markov aleatorizada, entonces el proceso $\{x_t\}_{t=0}^\infty$ de estados es un proceso de Markov no homogéneo con kerneles de transición $\{Q_{\varphi_t}(\cdot|\cdot)\}_{t=0}^\infty$, donde $Q_{\varphi_t}(B|x) = \int_A Q(B|x,a)\varphi_t(da|x)$, $\forall B \in \mathcal{B}(X)$, $x \in X$. Es decir*

$$P^{\varphi_t}_\nu(x_{t+1} \in B|x_0, x_1, \ldots, x_t) = P^{\varphi_t}_\nu(x_{t+1} \in B|x_t) = Q_{\varphi_t}(B|x_t).$$

Demostración Para una prueba de la proposición 1.4.1 véase [19, pág. 19-20]. ∎

En particular, para políticas estacionarias tenemos la siguiente propiedad.

Proposición 1.4.2 *Sea ν una distribución inicial en $(X, \mathcal{B}(X))$. Si $\pi = \{\varphi\} \in \Phi$ es una política de Markov estacionaria aleatorizada, entonces el proceso $\{x_t\}_{t=0}^{\infty}$ de estados es un proceso de Markov con kerneles de transición $\{Q_\varphi(\cdot|\cdot)\}_{t=0}^{\infty}$, donde $Q_\varphi(B|x) = \int_A Q(B|x,a)\varphi(da|x)$, $\forall B \in \mathcal{B}(X)$, $x \in X$. Es decir*

$$P_\nu^\varphi(x_{t+1} \in B|x_0, x_1, \ldots, x_t) = P_\nu^\varphi(x_{t+1} \in B|x_t) = Q_\varphi(B|x_t).$$

1.5. Normas de peso W

Sea X un espacio métrico, y sea $B_b(X)$ el espacio de Banach de las funciones medibles, acotadas y real-valuadas u en X, con la norma del supremo

$$\|u\| := \sup_{x \in X} |u(x)|.$$

Sea además $C_b(X)$ el subespacio cerrado de $B_b(X)$ de todas las funciones continuas acotadas en X.

Supongamos también que $W : X \to [\theta, \infty)$ denota una función medible dada, la cual será referida como la *función de peso*, donde $\theta > 0$. Si u es una función real valuada en X, definimos su W-norma como

$$\|u\|_W := \sup_{x \in X} |u(x)|/W(x).$$

Notemos que si W es la función constante 1, $W(\cdot) \equiv 1$, luego la W-norma coincide con la norma del supremo.

Una función u se dice ser *acotada* si $\|u\| < \infty$ y W-acotada si $\|u\|_W < \infty$. En general, la función de peso W será no acotada, aunque evidentemente es W-acotada, ya que $\|W\|_W = 1$. Por otro lado, si u es una función acotada, luego esta es W-acotada pues como $W \geq \theta$ luego

$$\|u\|_W \leq \frac{1}{\theta}\|u\| < \infty \quad \forall u \in B_b(X). \tag{1.5.1}$$

Sea $B_W(X)$ el espacio lineal normado de funciones medibles W-acotadas u sobre X. Este espacio también es de Banach ya que si $\{u_n\}$ es una sucesión de Cauchy con la W-norma, luego $\{u_n/W\}$ es una sucesión de Cauchy con la norma del supremo; de aquí, como $B_b(X)$ es un espacio de Banach, uno puede deducir la existencia de una función u en $B_W(X)$ que sea el W-límite de $\{u_n\}$. Combinando este hecho y 1.5.1 obtenemos:

Proposición 1.5.1 $(B_W(X), \|\cdot\|_W)$ *es un espacio de Banach.*

Observación 1.5.1 *Notemos que el espacio normado* $(B_b(X), \|\cdot\|)$ *está inmerso continuamente en* $(B_W(X), \|\cdot\|_W)$.

Capítulo 2

Control descontado con restricciones de costo

2.1. Introducción

En esta capítulo probamos la existencia de políticas óptimas para el problema descontado con restricciones, que estableceremos en la definición 2.4.2 más adelante. El método que utilizaremos será la aplicación de la técnica de los multiplicadores de Lagrange, que permitirá obtener una familia paramétrica de ecuaciones de optimalidad asociadas a problemas de control descontado sin restricciones. Posteriormente, mediante técnicas de programación dinámica y argumentos de cálculo elemental obtendremos políticas óptimas de control para los problemas sin restricciones, las cuales resolverán el problema original de control con restricciones. Precisando, encontraremos un multiplicador adecuado λ^* y resolveremos el problema de control óptimo asociado al problema sin restricciones indizado por el multiplicador λ^*, de tal forma que la política óptima que resuelve este problema sin restricciones también resolverá nuestro problema inicial con restricciones. A continuación introduciremos nuestra hipótesis fundamentales, y definiremos las funciones de ganancia y costo descontados.

2.2. El modelo de control descontado con restricciones

Vamos a introducir los siguientes conjuntos de hipótesis. La primera, hipótesis 2.2.1, consiste en condiciones estándares de continuidad y compacidad. La segunda, hipótesis 2.2.2, usa una función de peso W para imponer condiciones de crecimiento sobre las funciones de costo y ganancia (véase [12, 20, 30]).

Hipótesis 2.2.1 *Para cada estado $x \in X$:*

a) $A(x)$ es un subconjunto compacto de A;

b) La función de costo $c(x,a)$ y la función de ganancia $r(x,a)$ son continuas en $a \in A(x)$;

c) La ley de transición Q es continua en $a \in A(x)$, esto es, el mapeo

$$a \longmapsto \int_X v(y)Q(dy|x,a)$$

es continuo en $A(x)$ para cada función medible acotada sobre X.

Hipótesis 2.2.2 *Existe una función medible $W \geq 1$ en X, una función medible acotada $b(\cdot) \geq 0$, constantes no negativas M, β, con $\beta < 1$, tales que:*

(a) $|r(x,a)| \leq MW(x), \quad |c(x,a)| \leq MW(x)$ para cada $(x,a) \in \mathbb{K}$;

(b) $\int_X W(y)Q(dy|x,a)$ es continua en $a \in A(x)$; y

(c) $\int_X W(y)Q(dy|x,a) \leq \beta W(x) + b(x)$ para cada $x \in X$.

Definición 2.2.1 *Sea $0 < \alpha < 1$ un factor de descuento. La ganacia promedio descontada y el costo promedio descontado dada una política $\pi \in \Pi$ y el estado inicial $x \in X$ se definen respectivamente por*

$$V_\alpha(x,\pi,r) := E_x^\pi \left[\sum_{t=0}^{\infty} \alpha^t r(x_t,a_t)\right], \quad y \quad V_\alpha(x,\pi,c) := E_x^\pi \left[\sum_{t=0}^{\infty} \alpha^t c(x_t,a_t)\right]. \tag{2.2.1}$$

.

Para probar que las funciones descontadas $V_\alpha(\cdot,\pi,r)$ y $V_\alpha(\cdot,\pi,c)$ pertenecen al espacio $B_W(X)$, requerimos del siguiente lema.

Lema 2.2.1 *Para cada política $\pi \in \Pi$, estado inicial $x \in X$, $t = 0, 1, \ldots$, se tiene que*

$$E_x^\pi W(x_t) \leq \beta^t W(x) + \frac{b}{1-\beta}, \tag{2.2.2}$$

donde $b := \sup_{x \in X} b(x)$.

Demostración Por la condición (iii) del teorema 1.4.1 y la condición (c) de la hipótesis 2.2.2, se tiene que para cada $t = 0, 1, \ldots,$

$$E_x^\pi[W(x_{t+1})|x_t, a_t] = \int_X W(y)Q(dy|x_t, a_t) \leq \beta W(x_t) + b. \tag{2.2.3}$$

Tomando esperanzas en ambos lados de la igualdad (2.2.3) se tiene que

$$E_x^\pi W(x_{t+1}) \leq \beta E_x^\pi W(x_t) + b. \tag{2.2.4}$$

Por un proceso inductivo, la desigualdad (2.2.4) implica necesariamente (2.2.2). ■

Proposición 2.2.1 *Las funciones $V_\alpha(\cdot, \pi, r)$ y $V_\alpha(\cdot, \pi, c)$ pertenecen al espacio $B_W(X)$ para cada $\pi \in \Pi$; de hecho, para cada estado inicial $x \in X$,*

$$\sup_{\pi \in \Pi} |V_\alpha(x, \pi, r)| \leq M(\alpha)W(x), \tag{2.2.5}$$

y

$$\sup_{\pi \in \Pi} |V_\alpha(x, \pi, c)| \leq M(\alpha)W(x), \tag{2.2.6}$$

con $M(\alpha) := M\left[\frac{1}{1-\alpha\beta} + \frac{b}{(1-\alpha)(1-\beta)}\right]$, donde $M, \beta, b(\cdot)$ son las constantes y la función que aparecen en la hipótesis 2.2.2, donde $b := \sup_{x \in X} b(x)$.

Demostración Probemos el resultado para $V_\alpha(x, \pi, r)$. De la hipótesis 2.2.2 y del lema 2.2.1 se tiene

$$|V_\alpha(x, \pi, r)| \leq \sum_{t=0}^{\infty} \alpha^t M E_x^\pi W(x_t) \leq M \sum_{t=0}^{\infty} \alpha^t \left(\beta^t W(x) + \frac{b}{1-\beta}\right) =$$

$$= M\left[\frac{W(x)}{1-\alpha\beta} + \frac{b}{(1-\alpha)(1-\beta)}\right] \leq M(\alpha)W(x),$$

con

$$M(\alpha) := M\left[\frac{1}{1-\alpha\beta} + \frac{b}{(1-\alpha)(1-\beta)}\right].$$

■

2.3. La ecuación de Poisson α-descontada

La siguiente proposición da una caracterización para la ganancia y el costo α-descontado, respectivamente, cuando consideramos políticas estacionarias aleatorizadas $\varphi \in \Phi$, y que en el marco de esta tesis, diremos que se satisface la *ecuación de Poisson α-descontada* (2.3.1).

Proposición 2.3.1 *Sea $v : \mathbb{K} \to \mathbb{R}$ una función medible que satisface condiciones similares a las dadas para r y c en las hipótesis 2.2.1 y 2.2.2. Luego, para cada política estacionaria aleatorizada $\varphi \in \Phi$ la función α-descontada asociada $V_\alpha(\cdot, \varphi, v)$ pertenece al espacio normado $B_W(X)$, y satisface la ecuación*

$$h(x) = v_\varphi(x) + \alpha \int_X h(y) Q_\varphi(dy|x) \text{ para cada } x \in X. \tag{2.3.1}$$

Recíprocamente, si alguna función $h \in B_W(X)$ satisface (2.3.1), entonces necesariamente $h(x) = V_\alpha(x, \varphi, v)$ para cada $x \in X$.

Antes de probar la proposición demostraremos el siguiente lema de carácter técnico.

Lema 2.3.1 *Para cada función h en $B_W(X)$ y para cada $t = 0, 1, \ldots,$ se tiene*

$$E_x^\varphi h(x_{t+1}) = \int_X E_z^\varphi h(x_t) Q_\varphi(dz|x). \tag{2.3.2}$$

Demostración Usando el Teorema de Fubini y la relación (1.4.4), obtenemos

$$E_x^\varphi h(x_{t+1}) = \int_X h(y) Q_\varphi^{t+1}(dy|x) = \int_X h(y) \int_X Q_\varphi^t(dy|z) Q_\varphi(dz|x)$$

$$= \int_X \left[\int_X h(y) Q_\varphi^t(dy|x) \right] Q_\varphi(dz|x) = \int_X \left[E_z^\varphi h(x_t) \right] Q_\varphi(dz|x).$$

■

A continuación demostramos la proposición 2.3.1

Demostración Sea $\varphi \in \Phi$ una política estacionaria aleatorizada. Por definición de la función descontada $V_\alpha(x, \varphi, v)$

$$\begin{aligned} V_\alpha(x,\varphi,v) &= E_x^\varphi\left[\sum_{t=0}^{\infty}\alpha^t v(x_t,\varphi)\right] = v_\varphi(x) + E_x^\varphi\left[\sum_{t=1}^{\infty}\alpha^t v_\varphi(x_t)\right] = \\ &= v_\varphi(x) + \alpha\left[E_x^\varphi\sum_{t=1}^{\infty}\alpha^{t-1}v_\varphi(x_t))\right], \end{aligned} \tag{2.3.3}$$

con

$$v_\varphi(x) = \int_A v(x,a)\varphi(da|x).$$

Por el lema 2.3.1,

$$E_x^\varphi v_\varphi(x_t) = \int_X E_x^\varphi v_\varphi(x_{t-1})Q_\varphi(dz|x) \tag{2.3.4}$$

Sustituyendo (2.3.4) en (2.3.3) se tiene que

$$\begin{aligned} V_\alpha(x,\varphi,v) &= v_\varphi(x) + \alpha\sum_{t=0}^{\infty}\alpha^t E_x^\varphi v_\varphi(x_{t-1}) \\ &= v_\varphi(x) + \alpha\sum_{t=0}^{\infty}\alpha^t\int_X E_z^\varphi v_\varphi(x_t)Q_\varphi(dz|x) \\ &= v_\varphi(x) + \alpha\int_X\left[\sum_{t=0}^{\infty}\alpha^t E_z^\varphi v_\varphi(x_t)\right]Q_\varphi(dz|x) \\ &= v_\varphi(x) + \alpha\int_X E_z^\varphi\left[\sum_{t=0}^{\infty}\alpha^t v_\varphi(x_t)\right]Q_\varphi(dz|x) \\ &= v_\varphi(x) + \alpha\int_X V_\alpha(z,\varphi,v)Q_\varphi(dz|x). \end{aligned} \tag{2.3.5}$$

Así pues, $V_\alpha(x, \varphi, v)$ satisface la ecuación (2.3.1). Recíprocamente sea h una función en $B_W(X)$ satisfaciendo (2.3.1). Teniendo en mente la hipótesis 2.2.2, integrando (2.3.1) respecto a $Q_\varphi^t(dx|z)$, obtenemos

$$\int_X h(x)Q_\varphi^t(dx|z) = \int_X v_\varphi(x)Q_\varphi^t(dx|z) + \alpha\int_X\left[\int_X h(y)Q_\varphi(dy|x)\right]Q_\varphi^t(dx|z). \tag{2.3.6}$$

Por Fubini y por la relación (1.4.4), la ecuación (2.3.6) toma la forma

$$\int_X h(x)Q_\varphi^t(dx|z) = \int_X v_\varphi(x)Q_\varphi^t(dx|z) + \alpha \int_X h(y)Q_\varphi^{t+1}(dx|z),$$

lo que equivale

$$E_z^\varphi h(x_t) = E_z^\varphi v_\varphi(x_t) + \alpha E_z^\varphi h(x_{t+1}) \quad \text{para todo } t = 0, 1, \ldots. \tag{2.3.7}$$

Multiplicando por α^t a la ecuación (2.3.7), se tiene

$$\alpha^t E_z^\varphi h(x_t) = \alpha^t E_z^\varphi v_\varphi(x_t) + \alpha^{t+1} E_z^\varphi h(x_{t+1}) \quad \text{para todo } t = 0, 1, \ldots.$$

Sumando término a término estas ecuaciones desde $t = 0$ hasta $t = n-1$ con $n \geq 1$, y cambiando z por x obtenemos

$$E_x^\varphi \left[\sum_{t=0}^{n-1} \alpha^t h(x_t)\right] = E_x^\varphi \left[\sum_{t=0}^{n-1} \alpha^t v_\varphi(x_t)\right] + E_x^\varphi \left[\sum_{t=1}^{n} +\alpha^t h(x_t)\right],$$

que al reducir términos, toma la forma

$$h(x) = E_x^\varphi \left[\sum_{t=0}^{n-1} \alpha^t v_\varphi(x_t)\right] + \alpha^n E_x^\varphi h(x_n). \tag{2.3.8}$$

Como

$$|E_x^\varphi h(x_n)| \leq \|h\|_W E_x^\varphi W(x_n) \leq \|h\|_W \left(\beta^n W(x) + \frac{b}{1-\beta}\right)$$

se tiene que

$$|\alpha^n E_x^\varphi h(x_n)| \leq \alpha^n \|h\|_W \left(W(x) + \frac{b}{1-\beta}\right) \to 0$$

cuando $n \to \infty$. Haciendo tender $n \to \infty$ en (2.3.8) obtenemos

$$h(x) = \sum_{t=0}^{\infty} E_x^\varphi v_\varphi(x_t) = V_\alpha(x, \varphi, v) \quad \text{para todo } x \in X,$$

es decir $h(\cdot)$ coincide con $V_\alpha(\cdot, \varphi, v)$ tal y como queríamos probar. ∎

Hipótesis 2.3.1 *Para $0 < \alpha < 1$, consideremos una función medible $\theta_\alpha : X \to \mathbb{R}$, tal que $\theta_\alpha(\cdot) \in B_W(X)$, la cual llamaremos la* función de restricción, *es decir*

$$|\theta_\alpha(x)| \leq \|\theta_\alpha\|_W W(x) \tag{2.3.9}$$

para cada $x \in X$.

Definición 2.3.1 *Sea $\theta_\alpha(\cdot)$ una función de restricción satisfaciendo la hipótesis 2.3.1. La restricción* total esperada α-descontada *cuando el controlador usa la política $\pi \in \Pi$, dado el estado inicial $x \in X$, se define como*

$$\overline{\theta}_\alpha(x, \pi) = (1-\alpha)E_x^\pi \left[\sum_{t=0}^{\infty} \alpha^t \theta_\alpha(x_t)\right]. \tag{2.3.10}$$

Observación 2.3.1 *La función $\overline{\theta}_\alpha(\cdot, \pi) \in B_W(X)$ para cada $\pi \in \Pi$. Más aún, fijando $x \in X$ tenemos que*

$$\sup_{\pi \in \Pi} |\overline{\theta}_\alpha(x, \pi)| \leq \frac{(1-\alpha)\|\theta_\alpha\|_W M(\alpha)}{M} W(x). \tag{2.3.11}$$

donde $M(\alpha)$ es la constante dada en la proposición 2.2.1 y M es la constante dada en la hipótesis 2.2.2. En efecto, de (2.2.2) en el lema 2.2.1

$$|\overline{\theta}_\alpha(x, \pi)| = \left|(1-\alpha)E_x^\pi \sum_{t=0}^{\infty} \alpha^t \theta_\alpha(x_t)\right| \leq (1-\alpha)\sum_{t=0}^{\infty} \|\theta_\alpha\|_W E_x^\pi W(x_t) \leq$$

$$\leq (1-\alpha)\|\theta_\alpha\|_W \sum_{t=0}^{\infty} \alpha^t \left(\beta^t W(x) + \frac{b}{1-\beta}\right) = (1-\alpha)\|\theta_\alpha\|_W \left[\frac{W(x)}{1-\alpha\beta} + \frac{b}{(1-\alpha)(1-\beta)}\right] \leq$$

$$\leq (1-\alpha)\|\theta_\alpha\|_W \left[\frac{1}{1-\alpha\beta} + \frac{b}{(1-\alpha)(1-\beta)}\right] W(x) = \frac{(1-\alpha)\|\theta_\alpha\|_W M(\alpha)}{M} W(x).$$

Similarmente a como se demostró en la proposición 2.3.1, podemos probar la siguiente:

Proposición 2.3.2 *Sea $0 < \alpha < 1$ y $\theta_\alpha : X \to \mathbb{R}$ una función de restricción satisfaciendo la hipótesis 2.3.1. Entonces para cada política estacionaria aleatorizada $\varphi \in \Phi$ la restricción total esperada α-descontada $x \longmapsto \overline{\theta}_\alpha(x, \varphi)$ está en $B_W(X)$ y satisface la ecuación*

$$h(x) = (1-\alpha)\theta_\alpha(x) + \alpha \int_X h(y) Q_\varphi(dy|x) \text{ para cada } x \in X. \tag{2.3.12}$$

Recíprocamente, si alguna función $h \in B_W(X)$ verifica (2.3.12), entonces necesariamente $h(x) = \overline{\theta}_\alpha(x, \varphi)$ para cada $x \in X$.

2.4. El problema descontado con restricciones (PDR)

Definición 2.4.1 *Para cada factor de descuento $0 < \alpha < 1$, supongamos que se tiene un estado inicial $x \in X$ y una función de restricción $\theta_\alpha(\cdot) \in B_W(X)$. Definimos el conjunto*

$$\mathcal{F}^x_{\theta_\alpha} := \{\pi \in \Pi : V_\alpha(x, \pi, c) \leq \overline{\theta}_\alpha(x, \pi)\}. \tag{2.4.1}$$

Observación 2.4.1 *Dado un punto fijo $x \in X$, podemos suponer que $\mathcal{F}^x_{\theta_\alpha}$ es no vacío para evitar situaciones triviales. Un caso muy particular consiste en considerar a $\theta_\alpha(\cdot)$ una función constante, es decir, $\theta_\alpha(y) = \theta_\alpha$ para cada $y \in X$, donde θ_α se puede elegir de tal forma que*

$$\theta_{\alpha,\text{mín}}(x) := \inf_{\pi\in\Pi} V_\alpha(x, \pi, c) < \theta_\alpha < \theta_{\alpha,\text{máx}}(x) := \sup_{\pi\in\Pi} V_\alpha(x, \pi, c). \tag{2.4.2}$$

En este caso, $\overline{\theta}_\alpha(y, \pi) = \theta_\alpha$ para cada $y \in X$, y el conjunto $\mathcal{F}^x_{\theta_\alpha}$ es no vacío, con

$$\mathcal{F}^x_{\theta_\alpha} = \{\pi \in \Pi : V_\alpha(x, \pi, c) \leq \theta_\alpha\}.$$

Bajo nuestras hipótesis, podemos probar que las funciones $\theta_{\alpha,\text{mín}}(\cdot)$, $\theta_{\alpha,\text{máx}}(\cdot) \in B_W(X)$, son medibles y satisfacen las ecuaciones de optimalidad

$$\theta_{\alpha,\text{mín}}(y) = \min_{a\in A(y)} \left[c(y, a) + \alpha \int_X \theta_{\alpha,\text{mín}}(z) Q(dz|y, a) \right], \tag{2.4.3}$$

y además

$$\theta_{\alpha,\text{máx}}(y) = \max_{a\in A(y)} \left[c(y, a) + \alpha \int_X \theta_{\alpha,\text{máx}}(z) Q(dz|y, a) \right], \tag{2.4.4}$$

para cada $y \in X$ (véase, por ejemplo [20, Sección 8.3]).

A continuación establecemos el problema descontado con restricciones (PDR).

Definición 2.4.2 (El problema descontado con restricciones (PDR).) *Decimos que una política* $\pi^* \in \Pi$ *es* α-óptima para el PDR *con estado inicial* $x \in X$*, si* $\pi^* \in \mathcal{F}^x_{\theta_\alpha}$*, y además*

$$V_\alpha(x, \pi^*, r) = \sup_{\pi \in \mathcal{F}^x_{\theta_\alpha}} V_\alpha(x, \pi, r), \text{ para cada } x \in X. \tag{2.4.5}$$

En este caso, $V^*_\alpha(x, r) = V_\alpha(x, \pi^*, r)$ *se llama la ganancia* α*-descontada óptima para el PDR con estado inicial* x*.*

Capítulo 3

Aproximación por multiplicadores de Lagrange

3.1. Introducción

En este capítulo usamos técnicas de multiplicadores de Lagrange y de programación dinámica para transformar el PDR original en un problema descontado sin restricciones parametrizado por una familia de multiplicadores de Lagrange. Para este propósito, tomemos $\lambda \leq 0$ y consideremos la función de ganancia generalizada

$$r_\alpha^\lambda(x,a) := r(x,a) + \lambda[c(x,a) - (1-\alpha)\theta_\alpha(x)]. \tag{3.1.1}$$

Notemos que $r_\alpha^\lambda(\cdot,\cdot)$ satisface condiciones similares a aquellas que satisfacen las funciones de costo $c(\cdot,\cdot)$ y la función de ganancia $r(\cdot,\cdot)$ en las hipótesis 2.2.1 y 2.2.2. Usando la misma notación que en (1.4.1) de la observación 1.4.1, podemos escribir para cada política $\varphi \in \Phi$ aleatorizada y estacionaria

$$r_{\alpha,\varphi}^\lambda(x) := r_\varphi(x) + \lambda[c_\varphi(x) - (1-\alpha)\theta_\alpha(x)]. \tag{3.1.2}$$

Además, observemos también que $r_{\alpha,\varphi}^\lambda(\cdot) \in B_W(X)$. En efecto,

$$\begin{aligned} |r_{\alpha,\varphi}^\lambda(x)| &\leq |r_\varphi(x)| + |\lambda\|c_\varphi(x)| + (1-\alpha)|\lambda\|\theta_\alpha(x)| \\ &\leq MW(x) + M|\lambda|W(x) + (1-\alpha)|\lambda\|\theta_\alpha\|_W W(x) \\ &\leq (M + M|\lambda| + (1-\alpha)|\lambda\|\theta_\alpha\|_W)W(x) = N_\alpha^\lambda W(x), \end{aligned} \tag{3.1.3}$$

donde $N_\alpha^\lambda := M + M|\lambda| + (1-\alpha)|\lambda\|\theta_\alpha\|_W$, y M es la constante que aparece en la hipótesis 2.2.1.

3.2. El problema descontado sin restricciones (PDSR)

De manera semejante a como se definieron las ganacias y costos descontados, definimos para cada $x \in X$, y para cada política $\pi \in \Pi$,

$$V_\alpha(x, \pi, r_\alpha^\lambda) := E_x^\pi \left[\sum_{t=0}^{\infty} \alpha^t r_\alpha^\lambda(x_t, a_t) \right]. \tag{3.2.1}$$

Observación 3.2.1 *Notemos que*

$$V_\alpha(x, \pi, r_\alpha^\lambda) = V_\alpha(x, \pi, r) + \lambda \left[V_\alpha(x, \pi, c) - \overline{\theta}_\alpha(x, \pi) \right] \quad \textit{para todo } x \in X,\ \pi \in \Pi. \tag{3.2.2}$$

De aquí, por la proposición 2.2.1 y por la relación (2.3.11) en la observación 2.3.1, obtenemos

$$\sup_{\pi \in \Pi} \left| V_\alpha(x, \pi, r_\alpha^\lambda) \right| \leq M_\alpha^\lambda W(x), \tag{3.2.3}$$

con $M_\alpha^\lambda := M(\alpha)[1 + |\lambda| + (1-\alpha)|\lambda| \|\theta_\alpha\|_W / M]$, *donde* $M(\alpha)$ *es la constante definida en la proposición 2.2.1. Así,* $V_\alpha(\cdot, \pi, r_\alpha^\lambda)$ *está en* $B_W(X)$.

El problema descontado sin restricciones se define como sigue.

Definición 3.2.1 El problema descontado sin restricciones (PDSR). *Una política* $\pi^* \in \Pi$ *satisfaciendo*

$$V_\alpha(x, \pi^*, r_\alpha^\lambda) = \sup_{\pi \in \Pi} V_\alpha(x, \pi, r_\alpha^\lambda) =: V_\alpha^*(x, r_\alpha^\lambda) \quad \textit{para todo } x \in X, \tag{3.2.4}$$

se llama α-descontada óptima *para el PDSR, y* $V_\alpha^*(x, r_\alpha^\lambda)$ *será llamada como la* ganancia α-descontada óptima *para el PDSR.*

Nótese que por (3.2.3), $V_\alpha^*(\cdot, r_\alpha^\lambda)$ pertenece a $B_W(X)$.

Imitando esencialmente los pasos de la demostración de la proposición 2.3.1, podemos probar el siguiente resultado que provee una caracterización útil de $V_\alpha(\cdot, \varphi, r_\alpha^\lambda)$ para cada política aleatorizada $\varphi \in \Phi$.

Proposición 3.2.1 *Supongamos válidas las hipótesis 2.2.1, 2.2.2 y 2.3.1. Entonces para cada $\lambda \leq 0$ y para cada política estacionaria aleatorizada $\varphi \in \Phi$, la función esperada α-descontada $V_\alpha(\cdot, \varphi, r_\alpha^\lambda)$ está en $B_W(X)$ y satisface la ecuación*

$$h(x) = r_{\alpha,\varphi}^\lambda(x) + \alpha \int_X h(y) Q_\varphi(dy|x) \quad \text{para cada } x \in X. \tag{3.2.5}$$

Recíprocamente, si alguna función $h \in B_W(X)$ satisface la ecuación (3.2.5), necesariamente se tendrá que

$$h(x) = V_\alpha(x, \varphi, r_\alpha^\lambda) \text{ para cada } x \in X. \tag{3.2.6}$$

Además, si la igualdad en (3.2.5) es reemplazada por " $\leq$ " o " $\geq$ ", luego (3.2.6) se mantiene con la respectiva desigualdad.

3.3. Familia paramétrica de α-EGDO y el PDR

La siguiente proposición establece la existencia de políticas descontadas óptimas para el PDSR. Además también asegura que la ganancia α-descontada óptima para el PDSR verifica la ecuación de ganancia α-descontada óptima, o simplemente, la ecuación de optimalidad (α-EGDO). La demostración de este resultado es un caso particular de la prueba dada en [20, Teorema 8.3.6].

Proposición 3.3.1 *Supongamos válidas las hipótesis de la proposición 3.2.1. Luego:*

i) *Para cada factor de descuento $\alpha \in (0, 1)$, $\lambda \leq 0$, la ganancia α-descontada óptima $V_\alpha^*(\cdot, r_\alpha^\lambda)$, tal y como se definió en (3.2.4), es la única solución de la ecuación de optimalidad en el espacio $B_W(X)$; esto es*

$$h(x) = \max_{a \in A(x)} \left\{ r_\alpha^\lambda(x, a) + \alpha \int_X h(y) Q(dy|x, a) \right\} \text{ para cada } x \in X. \tag{3.3.1}$$

ii) *Existe una política $f^* \in \mathbb{F}$ (que depende de λ y α) que maximiza el lado derecho de (3.3.1); esto es*

$$V_\alpha^*(x, r_\alpha^\lambda) = r_{\alpha,f^*}^\lambda(x) + \alpha \int_X V_\alpha^*(y, r_\alpha^\lambda) Q_{f^*}(dy|x) \text{ para cada } x \in X. \tag{3.3.2}$$

Más aún, f^* es α-descontada óptima para el PDSR; recíprocamente, si $\tilde{f} \in \mathbb{F}$ es α-descontada óptima para el PDSR, luego ésta satisface (3.3.2). Además,

$$V_\alpha^*(x, r_\alpha^\lambda) = \sup_{f \in \mathbb{F}} V_\alpha(x, f, r_\alpha^\lambda) = \sup_{\varphi \in \Phi} V_\alpha(x, \varphi, r_\alpha^\lambda) \text{ para cada } x \in X. \tag{3.3.3}$$

iii) *Una política $\pi^* \in \Pi$ es α-descontada óptima para el PDSR si y sólo si la correspondiente función de descuento $V_\alpha(\cdot, \pi^*, r_\alpha^\lambda)$ satisface la ecuación de optimalidad (3.3.1).*

iv) *Si existe una política α-descontada óptima para el PDSR, luego existe una política estacionaria determinista la cual es α-descontada óptima para el PDSR.*

Observación 3.3.1 *Para cada $\lambda \leq 0$, y $\alpha \in (0,1)$ denotamos*

$$\Pi_\alpha^\lambda := \{\pi \in \Pi | \pi \text{es una política } \alpha\text{-descontada óptima para el PDSR}\}. \tag{3.3.4}$$

La proposición 3.3.1(ii) nos garantiza que el conjunto $\mathbb{F} \cap \Pi_\alpha^\lambda$ no es vacío.

Lema 3.3.1 *Supongamos válidas las hipótesis de la proposición 3.2.1, y sea $\alpha \in (0,1)$. Entonces para cada $x \in X$ fijo, $\lambda \leq 0$ y cada número real η tal que $\lambda + \eta \leq 0$, se tiene:*

a) *Para cada $\pi_\alpha^\lambda \in \Pi_\alpha^\lambda$, $\pi_\alpha^{\lambda+\eta} \in \Pi_\alpha^{\lambda+\eta}$, son válidas las siguientes desigualdades,*

$$\begin{aligned} \eta[V_\alpha(x, \pi_\alpha^\lambda, c) - \overline{\theta}_\alpha(x, \pi_\alpha^\lambda)] &\leq V_\alpha^*(x, r_\alpha^{\lambda+\eta}) - V_\alpha^*(x, r_\alpha^\lambda) \leq \\ &\leq \eta[V_\alpha(x, \pi_\alpha^{\lambda+\eta}, c) - \overline{\theta}_\alpha(x, \pi_\alpha^{\lambda+\eta})] \end{aligned} \tag{3.3.5}$$

,

b) *El mapeo $\lambda \longmapsto V_\alpha^*(x, r_\alpha^\lambda)$ es continuo en $(-\infty, 0]$,*

c) *Si $\lambda \longmapsto V_\alpha^*(x, r_\alpha^\lambda)$ es diferenciable en un punto $\Lambda < 0$, luego*

$$\left.\frac{\partial V_\alpha^*(x, r_\alpha^\lambda)}{\partial \lambda}\right|_{\lambda=\Lambda} = V_\alpha(x, \pi_\alpha^\Lambda, c) - \overline{\theta}_\alpha(x, \pi_\alpha^\Lambda) \text{ para cada } \pi_\alpha^\Lambda \in \Pi_\alpha^\Lambda. \tag{3.3.6}$$

Demostración (a) De las relaciones en (3.2.2) y (3.2.4) , obtenemos

$$V_\alpha^*(x, r_\alpha^{\lambda+\eta}) \geq V_\alpha(x, \pi_\alpha^\lambda, r_\alpha^{\lambda+\eta}) = V_\alpha(x, \pi_\alpha^\lambda, r) + (\lambda+\eta)\left[V_\alpha(x, \pi_\alpha^\lambda, c) - \overline{\theta}_\alpha(x, \pi_\alpha^\lambda)\right], \tag{3.3.7}$$

donde la última desigualdad se sigue de la definición de $r_\alpha^{\lambda+\eta}$. Ahora, de la definición de $\pi_\alpha^\lambda \in \Pi_\alpha^\lambda$ y de la relación (3.2.2), obtenemos la igualdad

$$V_\alpha^*(x, r_\alpha^\lambda) = V_\alpha(x, \pi_\alpha^\lambda, r_\alpha^\lambda) = V_\alpha(x, \pi_\alpha^\lambda, r) + \lambda\left[V_\alpha(x, \pi_\alpha^\lambda, c) - \overline{\theta}_\alpha(x, \pi_\alpha^\lambda)\right]. \tag{3.3.8}$$

Restando (3.3.8) y (3.3.7), obtenemos

$$V_\alpha^*(x, r_\alpha^{\lambda+\eta}) - V_\alpha^*(x, r_\alpha^\lambda) \geq \eta\left[V_\alpha(x, \pi_\alpha^\lambda, c) - \overline{\theta}_\alpha(x, \pi_\alpha^\lambda)\right]. \tag{3.3.9}$$

Repitiendo el procedimiento anterior, pero ahora considerando $V_\alpha^*(x, r_\alpha^\lambda)$ y la política $\pi_\alpha^{\lambda+\eta}$ obtenemos también

$$V_\alpha^*(x, r_\alpha^{\lambda+\eta}) - V_\alpha^*(x, r_\alpha^\lambda) \leq \eta\left[V_\alpha(x, \pi_\alpha^{\lambda+\eta}, c) - \overline{\theta}_\alpha(x, \pi_\alpha^{\lambda+\eta})\right]. \tag{3.3.10}$$

Así pues, la relación (3.3.5) se sigue combinando (3.3.9) y (3.3.10).

(b) De (2.2.6) en la proposición 2.2.1 y por (2.3.11), se tiene para un estado inicial $x \in X$ fijo

$$\left|V_\alpha(x, \pi_\alpha^{\lambda+\eta}, c) - \overline{\theta}_\alpha(x, \pi_\alpha^{\lambda+\eta})\right| \leq M(\alpha)\left[1 + \frac{(1-\alpha)\|\theta_\alpha\|_W}{M}\right] W(x).$$

De aquí, la continuidad de $\lambda \longmapsto V_\alpha^*(x, r_\alpha^\lambda)$ se sigue cuando $\eta \to 0$ en todos los términos de (3.3.5).

(c) Para un estado inicial $x \in X$, suponiendo que $\lambda \longmapsto V_\alpha^*(x, r_\alpha^\lambda)$ sea diferenciable en $\lambda = \Lambda < 0$, luego para cada $\pi_\alpha^\Lambda \in \Pi_\alpha^\Lambda$ fija, de la primera desigualdad en (3.3.5), obtenemos, para cada $\eta > 0$,

$$V_\alpha(x, \pi_\alpha^\Lambda, c) - \overline{\theta}_\alpha(x, \pi_\alpha^\Lambda) \leq \frac{V_\alpha^*(x, r_\alpha^{\Lambda+\eta}) - V_\alpha^*(x, r_\alpha^\Lambda)}{\eta},$$

y

$$V_\alpha(x, \pi_\alpha^\Lambda, c) - \overline{\theta}_\alpha(x, \pi_\alpha^\Lambda) \geq \frac{V_\alpha^*(x, r_\alpha^{\Lambda-\eta}) - V_\alpha^*(x, r_\alpha^\Lambda)}{-\eta}.$$

Tomando límites cuando $\eta \downarrow 0$ se sigue (3.3.6). ∎

3.4. Existencia de políticas óptimas para el PDR

Teorema 3.4.1 *Supongamos válidas las hipótesis 2.2.1, 2.2.2 y 2.3.1. Fijemos un factor de descuento $\alpha \in (0,1)$, así como un estado inicial $x \in X$. Entonces:*

a) *Supongamos que existe $\Lambda \leq 0$ y $\widehat{\pi} \in \Pi$ satisfaciendo*

$$V_\alpha(x, \widehat{\pi}, c) = \overline{\theta}_\alpha(x, \widehat{\pi}), \quad y \quad V_\alpha(x, \widehat{\pi}, r_\alpha^\Lambda) = V_\alpha^*(x, r_\alpha^\Lambda). \tag{3.4.1}$$

Por lo tanto $\widehat{\pi}$ es una política α-óptima para el PDR, y $V_\alpha^*(x, r_\alpha^\Lambda)$ es el valor α-óptimo para el PDR. Más aún,

$$V_\alpha^*(x, r_\alpha^\Lambda) = \inf_{\lambda \leq 0} V_\alpha^*(x, r_\alpha^\lambda). \tag{3.4.2}$$

b) *Si $\lambda^* < 0$ es un punto crítico de $\lambda \longmapsto V_\alpha^*(x, r_\alpha^\lambda)$, esto es, si la derivada en (3.3.6) es igual a cero en $\lambda = \lambda^*$, luego cada política $\pi_\alpha^{\lambda^*} \in \Pi_\alpha^{\lambda^*}$ es α-óptima para el PDR. Además se tiene $V_\alpha(x, \pi_\alpha^{\lambda^*}, c) = \overline{\theta}_\alpha(x, \pi_\alpha^{\lambda^*})$, siendo $V_\alpha^*(x, r_\alpha^{\lambda^*})$ el valor α-óptimo para el PDR, el cual coincide con $V_\alpha(x, \pi_\alpha^{\lambda^*}, r)$. También vale*

$$V_\alpha^*(x, r_\alpha^{\lambda^*}) = \inf_{\lambda \leq 0} V_\alpha^*(x, r_\alpha^{\lambda^*}). \tag{3.4.3}$$

c) *Caso $\lambda = 0$: Si $\pi_\alpha^0 \in \Pi_\alpha^0$ satisface $V_\alpha(x, \pi_\alpha^0, c) \leq \overline{\theta_\alpha}(x, \pi_\alpha^0)$; esto es, $\pi_\alpha^0 \in \mathcal{F}_{\theta_\alpha}^x$, luego π_α^0 es una política α-óptima para el PDR. En este caso, $V_\alpha^*(x, r_\alpha^0) = V_\alpha^*(x, r)$ es el valor α-óptimo para el PDR y coincide con $V_\alpha(x, \pi_\alpha^0, r)$. Además se tiene*

$$V_\alpha^*(x, r_\alpha^0) = \inf_{\lambda \leq 0} V_\alpha^*(x, r_\alpha^\lambda). \tag{3.4.4}$$

Demostración (a) Sea $\hat{\pi} \in \Pi$ satisfaciendo (3.4.1). Así, $\hat{\pi} \in \mathcal{F}_{\theta_\alpha}^x$, y por (3.2.2), tenemos

$$V_\alpha^*(x, r_\alpha^\Lambda) = V_\alpha(x, \hat{\pi}, r_\alpha^\Lambda) = V_\alpha(x, \hat{\pi}, r). \tag{3.4.5}$$

Por otro lado, para cada $\pi \in \mathcal{F}_{\theta_\alpha}^x$, se tiene que $V_\alpha(x, \pi, c) - \overline{\theta}_\alpha(x, \pi) \leq 0$, de donde $\Lambda[V_\alpha(x, \pi, c) - \overline{\theta}_\alpha(x, \pi)] \geq 0$. Así,

$$V_\alpha(x, \pi, r) \leq V_\alpha(x, \pi, r) + \Lambda[V_\alpha(x, \pi, c) - \overline{\theta}_\alpha(x, \pi)] = V_\alpha(x, \pi, r_\alpha^\Lambda) \quad \text{para toda } \pi \in \mathcal{F}_{\theta_\alpha}^x. \tag{3.4.6}$$

Combinando (3.4.5) y (3.4.6), obtenemos

$$V_\alpha(x, \hat{\pi}, r) = V_\alpha^*(x, r_\alpha^\Lambda) = \sup_{\tilde{\pi} \in \Pi} V_\alpha(x, \tilde{\pi}, r_\alpha^\Lambda) \geq \sup_{\pi \in \mathcal{F}_{\theta_\alpha}^x} V_\alpha(x, \pi, r).$$

Esto prueba que $\hat{\pi} \in \Pi$ es una política α-óptima para el PDR, y $V_\alpha^*(x, r_\alpha^\Lambda)$ es el valor α-óptimo para el PDR.

Ahora bien, para cada $\lambda \leq 0$, tenemos por (3.2.2)

$$V_\alpha^*(x, r_\alpha^\lambda) = \sup_{\tilde{\pi} \in \Pi} V_\alpha(x, \tilde{\pi}, r_\alpha^\lambda) \geq V_\alpha(x, \hat{\pi}, r_\alpha^\lambda) = V_\alpha(x, \hat{\pi}, r) = V_\alpha^*(x, r_\alpha^\Lambda).$$

Esta última desigualdad prueba (3.4.2).

(b) Dada que $\lambda^* < 0$ es un punto crítico de $\lambda \longmapsto V_\alpha^*(x, r_\alpha^\lambda)$, de (3.3.6) en el lema 3.3.1,

$$\left.\frac{\partial V_\alpha^*(x, r_\alpha^\lambda)}{\partial \lambda}\right|_{\lambda=\lambda^*} = V_\alpha(x, \pi_\alpha^{\lambda^*}, c) - \overline{\theta}_\alpha(x, \pi_\alpha^{\lambda^*}) = 0,$$

esto es, $V_\alpha(x, \pi_\alpha^{\lambda^*}, c) = \overline{\theta}_\alpha(x, \pi_\alpha^{\lambda^*})$. Ya que $V_\alpha^*(y, r_\alpha^{\lambda^*}) = V_\alpha(y, \pi_\alpha^{\lambda^*}, r_\alpha^{\lambda^*})$ para toda $y \in X$, en particular, tenemos que $V_\alpha^*(x, r_\alpha^{\lambda^*}) = V_\alpha(x, \pi_\alpha^{\lambda^*}, r_\alpha^{\lambda^*})$. El resto de la prueba es una consecuencia directa de (a).

(c) Es claro que $\lambda = 0$ implica que $r_\alpha^0(x, a) = r(x, a)$ para toda $(x, a) \in \mathbb{K}$. Ya que Π_α^0 es no vacío (ver observación 3.3.1), consideremos $\pi_\alpha^0 \in \Pi_\alpha^0$ tal que $V_\alpha(x, \pi_\alpha^0, c) \leq \overline{\theta}_\alpha(x, \pi_\alpha^0)$. Así $\pi_\alpha^0 \in \mathcal{F}_{\theta_\alpha}^x$. Por otro lado, π_α^0 es óptima para el PDSR ($\lambda = 0$). Luego π_α^0 es α-óptima para el PDR

$$V_\alpha(x, \pi_\alpha^0, r) = V_\alpha^*(x, r_\alpha^0) = \sup_{\pi \in \Pi} V_\alpha(x, \pi, r) \geq \sup_{\pi \in \mathcal{F}_{\theta_\alpha}^x} V_\alpha(x, \pi, r) \geq V_\alpha(x, \pi_\alpha^0, r),$$

esto es

$$V_\alpha(x, \pi_\alpha^0, r) = V_\alpha^*(x, r_\alpha^0) = \sup_{\pi \in \mathcal{F}_{\theta_\alpha}^x} V_\alpha(x, \pi, r).$$

Así π_α^0 es una política α-óptima para el PDR, y $V_\alpha^*(x, r_\alpha^0)$ es el valor α-óptimo para el PDR. Ahora, de (3.3.5) con $\lambda = 0$, y $\eta < 0$

$$0 \leq \eta \left[V_\alpha(x, \pi_\alpha^0, c) - \overline{\theta}_\alpha(x, \pi_\alpha^0)\right] \leq V_\alpha^*(x, r_\alpha^\eta) - V_\alpha^*(x, r_\alpha^0),$$

lo que implica (3.4.4). ∎

Capítulo 4

El caso promedio con restricciones

4.1. Introducción

En este capítulo estudiaremos los procesos de control Markoviano promedio con restricciones en costos promedios, los cuales bajo hipótesis adicionales, garantizan en el modelo un buen comportamiento "estable" y uniforme en Φ. Este comportamiento estable permitirá utilizar la técnica de aproximación desvanecente para establecer la existencia de políticas óptimas para el problema con restricciones promedio. Para esto computaremos las políticas óptimas que resuelven los $\alpha-$PDR a través de la técnicas por multiplicadores de Lagrange utilizadas en el capítulo 3. Posteriormente hacemos tender el factor de descuento α a 1, y en consecuencia se obtendrán las políticas óptimas para el problema con restricciones promedio.

4.2. El problema de control de markoviano esperado con restricciones

Dada $\pi \in \Pi$, $x \in X$, y $n = 1, 2, \ldots$, definimos la ganancia por trayectorias en n-etapas y la ganancia esperada en n-etapas como

$$S_n(x, \pi, r) := \sum_{k=0}^{n-1} r(x_k, a_k), \quad \text{y} \quad J_n(x, \pi, r) := E_x^{\pi}[S_n(x, \pi, r)] \tag{4.2.1}$$

respectivamente. Reemplazando la función ganancia r con la función costo c obtenemos la definición de el costo por trayectorias en n-etapas y el costo esperado en n-etapas

$$S_n(x,\pi,c) := \sum_{k=0}^{n-1} c(x_k, a_k), \quad \text{y} \quad J_n(x,\pi,c) := E_x^{\pi}[S_n(x,\pi,c)] \tag{4.2.2}$$

respectivamente.

Definición 4.2.1 *La ganancia promedio esperada ("long-run" en inglés) está dada por*

$$J(x,\pi,r) := \liminf_{n\to\infty} \frac{1}{n} J_n(x,\pi,r). \tag{4.2.3}$$

Similarmente, el costo promedio esperado está definido por

$$J(x,\pi,c) := \limsup_{n\to\infty} \frac{1}{n} J_n(x,\pi,c). \tag{4.2.4}$$

Observe que $J(x,\pi,r)$ está definido como un "lím inf", mientras que $J(x,\pi,c)$ es un "lím sup", esto debido a que por convenciones estándares, la función r se interpreta como una función de ganancia, mientras que la función c es una función de costo.

El siguiente conjunto de hipótesis garantiza que el proceso de control Markoviano tiene un comportamiento "estable" y "uniforme" en Φ.

Hipótesis 4.2.1 (W-ergodicidad geométrica) *. Para cada política estacionaria aleatorizada $\varphi \in \Phi$ existe una medida invariante de probabilidad (necesariamente única) μ_φ en X tal que (con Q_φ^t como en (1.4.2)-(1.4.4))*

$$\left| \int_X u(y) Q_\varphi^t(dy|x) - \mu_\varphi(u) \right| \leq \|u\|_W R\rho^t W(x), \tag{4.2.5}$$

para cada $t = 0, 1, \ldots, u \in B_W(X)$, y $x \in X$, donde $R > 0$ y $0 < \rho < 1$ son constantes independientes de φ.

Observación 4.2.1 *Condiciones suficientes para que la hipótesis 4.2.1 se cumpla en el caso de políticas deterministas estacionarias son bien conocidas, véase por ejemplo [31, Teorema 3.6], [20, Proposición 10.2.5], [12, Lemas 3.3 y 3.4]. Para el caso de políticas estacionarias aleatorizadas véase por ejemplo, [24, Lemas 4.8 y 4.9].*

En particular, por las hipótesis 4.2.1 y 2.2.2(c), tenemos que

$$\mu_\varphi(W) \leq b/(1-\beta) < \infty \quad \textit{para cada} \quad \varphi \in \Phi, \tag{4.2.6}$$

con $b = \sup_{x\in X} b(x)$. *Más aún, por (4.2.5),* $J(x,\varphi,r)$ *y* $J(x,\varphi,c)$ *en la definición 4.2.1 son constantes (esto es, no dependen del estado inicial* x*), y verifican que*

$$J(x,\varphi,r) = \lim_{n\to\infty} \frac{1}{n} E_x^\varphi \sum_{k=0}^{n-1} r(x_k,a_k) = \mu_\varphi(r_\varphi) =: g(\varphi,r), \tag{4.2.7}$$

(donde la letra g es una abreviación para la palabra ganacia, el cual es otro nombre para la "ganancia promedio"[28], [30]). De la misma forma

$$J(x,\varphi,c) = \lim_{n\to\infty} \frac{1}{n} E_x^\varphi \sum_{k=0}^{n-1} c(x_k,a_k) = \mu_\varphi(c_\varphi) =: g(\varphi,c). \tag{4.2.8}$$

Tomando en cuenta la hipótesis 4.2.1, así como la observación 4.2.1, definimos

$$\theta_{\text{mín}} := \min_{\varphi\in\Phi} \int_X c_\varphi(y)\mu_\varphi(dy) \quad \text{y} \quad \theta_{\text{máx}} := \max_{\varphi\in\Phi} \int_X c_\varphi(y)\mu_\varphi(dy), \tag{4.2.9}$$

los cuales son números finitos. Para evitar situaciones triviales, consideraremos una restricción a la constante θ de tal forma que

$$\theta_{\text{mín}} < \theta < \theta_{\text{máx}}. \tag{4.2.10}$$

Definición 4.2.2 (El problema esperado con restricciones (PER)) *Sea* θ *una constante como en (4.2.10). Definimos el problema esperado con restricciones (PER) como:*

$$\textit{Maximizar} \quad J(x,\pi,r) \tag{4.2.11}$$

$$\textit{sujeto a:} \quad \pi \in \Pi \quad y \quad J(x,\pi,c) \leq \theta \quad \textit{para cada} \quad x \in X. \tag{4.2.12}$$

Definición 4.2.3 *Una política $\pi \in \Pi$ se dice ser admisible para el PER si éste satisface la restricción en (4.2.12), esto es, $J(x,\pi,c) \leq \theta$ para cada x en X. Denotaremos el conjunto de políticas admisibles para el PER como*

$$\mathcal{F}_\theta := \{\pi \in \Pi : J(x,\pi,c) \leq \theta \quad \textit{para cada} \quad x \in X\}. \tag{4.2.13}$$

Más aún, una política admisible π^* es llamada óptima *para el PER (4.2.11)-(4.2.12) si $J(x,\pi,r) \leq J(x,\pi^*,r)$ para toda $x \in X$ y para cada política admisible π. De aquí, definimos la función de* valor óptimo *para el PER como:*

$$V_\theta^*(x) := \sup_{\pi \in \mathcal{F}_\theta} J(x,\pi,r) = J(x,\pi^*,r) \quad \textit{para toda } x \in X. \tag{4.2.14}$$

Para situar el PER (4.2.11)-(4.2.12) en un contexto más general, convenimos en establecer la siguiente notación.

Restricción esperada promedio. Sea $\theta : X \to \mathbb{R}$ un mapeo medible tal que $\theta(\cdot) \in B_W(X)$ llamada la *tasa de restricción*. La *restricción esperada promedio* cuando el controlador usa una política $\pi \in \Pi$, dado el estado inicial $x \in X$, es definido por

$$\underline{\theta}(x,\pi) := \liminf_{n\to\infty} \frac{1}{n} E_x^\pi \sum_{k=0}^{n-1} \theta(x_k). \tag{4.2.15}$$

Notemos que si $\varphi \in \Phi$ es una política estacionaria aleatorizada, luego $\underline{\theta}(x,\varphi) = \mu_\varphi(\theta)$, donde μ_φ es la medida de probabilidad invariante asociada a φ en la hipótesis 4.2.1.

4.3. El problema esperado con restricciones generalizado

Definición 4.3.1 (El problema esperado con restricciones generalizado (PERG)) *Sea $\theta(\cdot) \in B_W(X)$ una tasa de restricción. Definimos el* problema esperado con restricciones generalizado *(PERG) por:*

$$\textit{maximizar } J(x,\pi,r) \tag{4.3.1}$$

$$\textit{sujeto a:} \quad \pi \in \Pi \quad \text{y} \quad J(x,\pi,c) \leq \underline{\theta}(x,\pi) \quad \textit{para toda } x \in X. \tag{4.3.2}$$

Definición 4.3.2 *Una política $\pi \in \Pi$ se dice ser* admisible *para el PERG si esta satisface las restricciones en (4.3.2), esto es, $J(x,\pi,c) \leq \underline{\theta}(x,\pi)$ para toda x en X. Denotamos el conjunto de políticas admisibles para el PERG como*

$$\mathcal{F}_\theta := \{\pi \in \Pi : J(x,\pi,c) \leq \underline{\theta}(x,\pi) \quad \textit{para toda } x \in X\}. \tag{4.3.3}$$

Además, una política admisible π^* se llama óptima *para el PERG (4.3.1)-(4.3.2) si $J(x,\pi,r) \leq J(x,\pi^*,r)$ para toda $x \in X$ y para cada política admisible π. De aquí, definimos la función de valor óptimo para el PERG como*

$$V_\theta^*(x) := \sup_{\pi\in\mathcal{F}_\theta} J(x,\pi,r) = J(x,\pi^*,r) \quad \textit{para toda } x \in X. \tag{4.3.4}$$

Requerimos de los siguientes lemas.

Lema 4.3.1 *Supongamos válidas las hipótesis 2.2.1, 2.2.2, 2.3.1, y 4.2.1. Sea $z \in X$ un estado fijo, con $v(x,a) := r(x,a)$ o $c(x,a)$. Luego para cada $\varphi \in \Phi$, $x \in X$,*

$$|V_\alpha(x,\varphi,v) - V_\alpha(z,\varphi,v)| \leq MR(1-\rho)^{-1}[1+W(z)]W(x), \tag{4.3.5}$$

y

$$|\underline{\theta}_\alpha(x,\varphi) - \underline{\theta}_\alpha(z,\varphi)| \leq (1-\alpha)\|\theta_\alpha\|_W R(1-\rho)^{-1}[1+W(z)]W(x), \tag{4.3.6}$$

con R y ρ como en la hipótesis 4.2.1, M como en la hipótesis 2.2.2.

Demostración De la hipótesis 4.2.1, para cada $x \in X$, $\varphi \in \Phi$, y $t = 0, 1, \cdots$, tenemos que

$$\begin{aligned} |E_x^\varphi v_\varphi(x_t) - E_z^\varphi v_\varphi(x_t)| &\leq \left|\int_{\mathbf{X}} v_\varphi(y)Q_\varphi^t(dy|x) - \mu_\varphi(v_\varphi)\right| + \left|\int_{\mathbf{X}} v_\varphi(y)Q_\varphi^t(dy|z) - \mu_\varphi(v_\varphi)\right| \\ &\leq MR\rho^t[W(x)+W(z)] \leq MR\rho^t[1+W(z)]W(x), \end{aligned}$$

la última desigualdad se sigue ya que $W(x) \geq 1$. De esta última desigualdad obtenemos

$$\begin{aligned} |V_\alpha(x,\varphi,v) - V_\alpha(z,\varphi,v)| &\leq \sum_{t=0}^{\infty} \alpha^t |E_x^\varphi v_\varphi(x_t) - E_z^\varphi v_\varphi(x_t)| \\ &\leq MR[1+W(z)]W(x)\sum_{t=0}^{\infty}\alpha^t\rho^t \leq MR[1+W(z)]W(x)\sum_{t=0}^{\infty}\rho^t, \end{aligned}$$

De aquí, la relación (4.3.5) se sigue ya que $0 < \alpha < 1$ y $\sum_{t=0}^{\infty}\rho^t = (1-\rho)^{-1}$.

Similarmente, podemos probar (4.3.6). ∎

Para una prueba del siguiente lema, véase [20, Lemma 8.3.7].

Lema 4.3.2 *Supongamos válidas las hipótesis 2.2.1(c) y 2.2.2(b). Luego:*

(a) *La función $x \mapsto \int_X h(y)Q(dy|x,a)$ es continua en $a \in A(x)$ para cada $x \in X$ y cada función $h \in B_W(X)$.*

(b) *Sea $\{h_m(\cdot)\}$ una sucesión acotada en $B_W(X)$, es decir, existe una constante K tal que $\|h_m\|_W \leq K$ para toda m. Definamos*

$$\underline{h}(x) := \liminf_{m\to\infty} h_m(x), \quad y \quad \overline{h}(x) := \limsup_{m\to\infty} h_m(x).$$

Luego para cualquier estado $x \in X$ y cualquier sucesión $\{a_m\}$ in $A(x)$ tal que $a_m \to a$ en $A(x)$, tenemos

$$\liminf_{m\to\infty} \int_X h_m(y)Q(dy|x,a_m) \geq \int_X \underline{h}(y)Q(dy|x,a), \tag{4.3.7}$$

y

$$\limsup_{m\to\infty} \int_X h_m(y)Q(dy|x,a_m) \leq \int_X \overline{h}(y)Q(dy|x,a). \tag{4.3.8}$$

De aquí, si $h_m(x) \to h(x)$ para toda $x \in X$, (esto es, $h = \underline{h} = \overline{h}$), luego

$$\lim_{m\to\infty} \int_X h_m(y)Q(dy|x,a_m) = \int_X h(y)Q(dy|x,a). \tag{4.3.9}$$

La siguiente proposición caracteriza a la constante $g(\cdot,c)$ como el límite de $(1-\alpha)V_\alpha(\cdot,\cdot,c)$ cuando $\alpha \uparrow 1$.

Proposición 4.3.1 *Suponga válidas las hipótesis 2.2.1, 2.2.2, 2.3.1, y 4.2.1. Considere una sucesión de factores de descuento $\{\alpha_m\}_m$ tal que $\alpha_m \uparrow 1$ cuando $m \to \infty$, así como una sucesión de políticas de control deterministas $\{f_{\alpha_m}\}_m$. También $f \in \mathbb{F}$ de tal forma que $f_{\alpha_m}(x) \to f(x)$ para toda $x \in X$. Luego*

$$(1-\alpha_m)V_{\alpha_m}(x, f_{\alpha_m}, v) \to g(f,v) = \mu_f(v_f), \quad \textit{cuando } m \to \infty, \quad \textit{para toda } x \in X,$$

donde $v = r$ o c.

Demostración Sea $z \in X$ un estado inicial fijo. De la proposición 2.3.1, la función $x \mapsto V_\alpha(x, f_{\alpha_m}, v)$ satisface la ecuación

$$V_{\alpha_m}(x, f_{\alpha_m}, v) = v_{f_{\alpha_m}}(x) + \alpha_m \int_X V_{\alpha_m}(y, f_{\alpha_m}, v)\ Q_{f_{\alpha_m}}(dy|x) \quad \text{para toda } x \in X. \tag{4.3.10}$$

Definimos

$$h_{\alpha_m}(x) := V_{\alpha_m}(x, f_{\alpha_m}, v) - V_{\alpha_m}(z, f_{\alpha_m}, v) \quad \text{para toda } x \in X. \tag{4.3.11}$$

En términos de la función h_{α_m}, la ecuación (4.3.10) se convierte en

$$h_{\alpha_m}(x) + (1-\alpha_m)V_{\alpha_m}(z, f_{\alpha_m}, v) = v_{f_{\alpha_m}}(x) + \alpha_m \int_X h_{\alpha_m}(y)\, Q_{f_{\alpha_m}}(dy|x) \quad \text{para toda } x \in X. \tag{4.3.12}$$

Note que por (2.2.5) y (2.2.6) se tiene

$$|(1-\alpha_m)V_{\alpha_m}(z, f_{\alpha_m}, v)| \leq \frac{M(1+b)}{1-\beta} W(z) \quad \text{para toda } m. \tag{4.3.13}$$

Así, (4.3.13) implica que la sucesión $\{(1-\alpha_m)V_{\alpha_m}(z, f_{\alpha_m}, v)\}$ pertenece a un conjunto compacto, y en consecuencia existe una constante ρ_v y una subsucesión $\{\alpha_{m_k}\}_k$ (denotada de nuevo por $\{\alpha_m\}_m$) tal que

$$\lim_{m\to\infty} (1-\alpha_m)V_{\alpha_m}(z, f_{\alpha_m}, v) = \rho_v. \tag{4.3.14}$$

Por otro lado, de (4.3.11) y el lema 4.3.1, tenemos que $\{h_{\alpha_m}\}_m$ es una sucesión acotada en $B_W(X)$. Así, podemos definir

$$\underline{h}(x) := \liminf_{m\to\infty} h_{\alpha_m}(x), \quad \text{y} \quad \overline{h}(x) := \limsup_{m\to\infty} h_{\alpha_m}(x).$$

De aquí, tomando los límites apropiados en (4.3.12), y considerando la hipótesis 2.2.1(b), del lema 4.3.2(b), así como por (4.3.14), obtenemos

$$\underline{h}(x) + \rho_v \geq v_f(x) + \int_X \underline{h}(y) Q_f(dy|x), \quad \text{para toda } x \in X, \tag{4.3.15}$$

y

$$\overline{h}(x) + \rho_v \leq v_f(x) + \int_X \overline{h}(y) Q_f(dy|x), \quad \text{para toda } x \in X. \tag{4.3.16}$$

Integrando ambos lados de las desigualdades (4.3.15) y (4.3.16) con respecto a la medida de probabilidad invariante μ_f, obtenemos que $\rho_v = \mu_f(v_f)$, esto es, $\rho_v = g(f, v)$. De aquí,

$$g(f, v) = \lim_{m\to\infty} (1-\alpha_m)V_{\alpha_m}(z, f_{\alpha_m}, v). \tag{4.3.17}$$

Ya que $z \in \mathbf{X}$ fue escogido arbitrariamente, luego (4.3.17) prueba que este límite se mantiene para toda $x \in X$. ■

Similarmente a como se hizo en la proposición 4.3.1, podemos probar lo siguiente.

Proposición 4.3.2 *Supongamos que las hipótesis de la proposición 4.3.1 se satisfacen. Consideremos una sucesión de factores de descuento* $\{\alpha_m\}_m$ *tales que el límite* $\alpha_m \uparrow 1$ *se tiene cuando* $m \to \infty$*, una sucesión acotada* $\{\theta_{\alpha_m}\}_m$ *in* $B_W(X)$ *de funciones de restricción (esto es,* $\sup_m \|\theta_{\alpha_m}\|_W < \infty$*),* $\theta(\cdot) \in B_W(X)$ *tal que* $(1-\alpha_m)\theta_{\alpha_m}(x) \to \theta(x)$ *cuando* $m \to \infty$ *para toda* $x \in X$*, y una sucesión de políticas de control deterministas* $\{f_{\alpha_m}\}_m$*,* $f \in \mathbb{F}$ *tal que* $f_{\alpha_m}(x) \to f(x)$ *para toda* $x \in X$*. Luego*

$$(1-\alpha_m)\overline{\theta}_{\alpha_m}(x, f_{\alpha_m}) \to \mu_f(\theta) = \underline{\theta}(x, f), \quad \text{cuando } m \to \infty, \quad \text{para toda } x \in X,$$

donde μ_f *es la única medida de probabilidad invariante en la hipótesis 4.2.1.*

4.4. La aproximación descontada desvanescente

Sea $0 < \alpha < 1$, y $\lambda \leq 0$. Considere la función de ganancia óptima α-descontada para el PDSR definido en (3.2.4), $V^*_\alpha(\cdot, r^\lambda_\alpha)$. Elijamos arbitrariamente un estado $z \in X$, el cual permanecerá fijo por el momento. Definimos

$$h^\lambda_\alpha(x) := V^*_\alpha(x, r^\lambda_\alpha) - V^*_\alpha(z, r^\lambda_\alpha) \quad \text{para toda } x \in X. \tag{4.4.18}$$

Lema 4.4.1 *Supongamos que se satisfacen las hipótesis 2.2.1, 2.2.2, 2.3.1, y 4.2.1. Entonces las funciones* $h^\lambda_\alpha(\cdot)$ *definidas en (4.4.18) pertenecen a* $B_W(X)$*. Más aún*

$$\left|h^\lambda_\alpha(x)\right| \leq \hat{N}^\lambda_\alpha W(x) \quad \text{para toda } x \in X, \tag{4.4.19}$$

donde $\hat{N}^\lambda_\alpha := N^\lambda_\alpha R(1-\rho)^{-1}[1+W(z)]$*, con* N^λ_α *la constante dada en (3.1.3), esto es,* $N^\lambda_\alpha = M + |\lambda|M + (1-\alpha)|\lambda|\|\theta_\alpha\|_W$*.*

Demostración De la hipótesis 4.2.1 y de (3.1.3), para cada $x \in X$, $\varphi \in \Phi$, y $t = 0, 1, \cdots$, tenemos

$$\begin{aligned}
\left|E^\varphi_x r^\lambda_{\alpha\varphi}(x_t) - E^\varphi_z r^\lambda_{\alpha\varphi}(x_t)\right| &\leq \left|\int_X r^\lambda_{\alpha\varphi}(y) Q^t_\varphi(dy|x) - \mu_\varphi(r^\lambda_{\alpha\varphi})\right| + \left|\int_X r^\lambda_{\alpha\varphi}(y) Q^t_\varphi(dy|z) - \mu_\varphi(r^\lambda_{\alpha\varphi})\right| \\
&\leq N^\lambda_\alpha R\rho^t[W(x) + W(z)] \leq N^\lambda_\alpha R\rho^t[1 + W(z)]W(x),
\end{aligned}$$

donde la última desigualdad se sigue porque $W(x) \geq 1$. De aquí, notemos que

$$\begin{aligned}
\left|V_\alpha(x, \varphi, r^\lambda_\alpha) - V_\alpha(z, \varphi, r^\lambda_\alpha)\right| &\leq \sum_{t=0}^{\infty} \alpha^t \left|E^\varphi_x r^\lambda_{\alpha\varphi}(x_t) - E^\varphi_z r^\lambda_{\alpha\varphi}(x_t)\right| \\
&\leq N^\lambda_\alpha R[1 + W(z)]W(x) \sum_{t=0}^{\infty} \alpha^t \rho^t \leq \\
&\leq N^\lambda_\alpha R[1 + W(z)]W(x) \sum_{t=0}^{\infty} \rho^t = \hat{N}^\lambda_\alpha W(x). \qquad (4.4.20)
\end{aligned}$$

La última desigualdad en (4.4.20) se cumple ya que $0 < \alpha < 1$ y $\sum_{t=0}^{\infty} \rho^t = (1-\rho)^{-1}$. Finalmente, de la relación (3.3.3), $V_\alpha^*(x, r_\alpha^\lambda) = \sup_{\varphi\in\Phi} V_\alpha(x, \varphi, r_\alpha^\lambda)$ para toda $x \in X$, luego

$$\left|h_\alpha^\lambda(x)\right| = \left|\sup_{\varphi\in\Phi} V_\alpha(x, \varphi, r_\alpha^\lambda) - \sup_{\varphi\in\Phi} V_\alpha(z, \varphi, r_\alpha^\lambda)\right| \leq \sup_{\varphi\in\Phi} \left|V_\alpha(x, \varphi, r_\alpha^\lambda) - V_\alpha(z, \varphi, r_\alpha^\lambda)\right|,$$

y por (4.4.20) se sigue el resultado deseado. ■

Usaremos el siguiente resultado en el intercambio de límites y máximos (véase, para esto, [19, Lema 4.2.4, pág. 47]).

Lema 4.4.2 *Suponga válidas la hipótesis 2.2.1(a). Consideremos una sucesión decreciente* $\{v_m(x,a)\}_m$ *de funciones en* $\mathbb{K}$ *(el conjunto definido en (1.3.1)) tal que* $\{v_m\}_m$ *está acotada en* $B_W(\mathbb{K})$*, satisfaciendo el límite* $v_m(x,a) \downarrow v_0(x,a)$ *para toda* $(x,a) \in \mathbb{K}$*. Además suponemos que* $v_m(x,\cdot)$ *es semicontinua superior (s.c.s.) en* $A(x)$ *para toda* $x \in X$ *y* $m = 0, 1, \cdots$*. Entonces*

$$\lim_{m\to\infty} \sup_{a\in A(x)} v_m(x,a) = \sup_{a\in A(x)} v_0(x,a), \quad \textit{para toda } x \in X.$$

Demostración Ya que $\{v_m\}_m$ está acotada en $B_W(\mathbb{K})$, existe una constante $K > 0$ tal que $|v_m(x,a)| \leq KW(x)$ para toda $(x,a) \in \mathbb{K}$. Ahora, para cada $m = 0, 1, \cdots$, definimos $v_m^*(x) := \sup_{a\in A(x)} v_m(x,a)$ para toda $x \in X$. Así, $\{v_m^*\}_m$ es una sucesión acotada en $B_W(X)$, decreciente, de tal forma que

$$-kW(x) \leq v_0^*(x) \leq v_{m+1}^*(x) \leq v_m^*(x) \leq KW(x) \quad \text{para toda } x \in X. \tag{4.4.21}$$

Por lo tanto,

$$v_0^*(x) \leq l(x) := \lim_{m\to\infty} v_m^*(x) = \inf_m v_m^*(x) \quad \text{para toda } x \in X. \tag{4.4.22}$$

Ahora consideremos un punto $x \in X$ fijado de antemano. Luego, de la hipótesis 2.2.1(a), el conjunto $A(x)$ es compacto, y por la semicontinuidad superior de las funciones $v_m(x,\cdot)$ en $A(x)$, se tiene que para cada $m = 0, 1, \cdots$, existe un punto $a_m \in A(x)$ tal que

$$v_m^*(x) = \sup_{a\in A(x)} v_m(x,a) = v_m(x, a_m). \tag{4.4.23}$$

Nuevamente, de la compacidad de $A(x)$ existe una subsucesión $\{a_{\varphi(n)}\}$ y un punto $a^* \in A(x)$ tal que

$$a^* = \lim_{n\to\infty} a_{\varphi(n)}. \tag{4.4.24}$$

Sea $m = 1, 2, \cdots$ fijo. Luego, para cada $n \geq m$, por (4.4.23) y de la propiedad de monotonía de la sucesión $\{v_k(x, a_{\varphi(n)})\}_k$, tenemos

$$l(x) \leq v_{\varphi(n)}^*(x) = v_{\varphi(n)}(x, a_{\varphi(n)}) \leq v_m(x, a_{\varphi(n)}) \quad \text{para toda } n \geq m. \tag{4.4.25}$$

Así, tomando el límite superior en (4.4.25) cuando $n \to \infty$, y teniendo en mente la s.c.s. de $v_m(x,\cdot)$ en $A(x)$, del límite en (4.4.24) se sigue

$$l(x) \leq \limsup_{n\to\infty} v_m(x, a_{\varphi(n)}) \leq v_m(x, a^*) \quad \text{para toda } m = 1, 2, \cdots . \tag{4.4.26}$$

Tomando el límite en (4.4.26) cuando $m \to \infty$, obtenemos

$$l(x) \leq v_0(x, a^*) \leq \sup_{a\in A(x)} v_0(x, a) = v_0^*(x) \quad \text{para toda } x \in X. \tag{4.4.27}$$

Finalmente, comparando (4.4.22) y (4.4.27), obtenemos el resultado deseado. ∎

Nuestro siguiente resultado constituye la versión promedio de la α-EGDO (3.3.1).

Teorema 4.4.1 *Suponga que las hipótesis 2.2.1, 2.2.2, y 4.3.1 se satisfacen. Suponga además que existe una sucesión de factores de descuento $\{\alpha_m\}_m$ tal que el límite $\alpha_m \uparrow 1$ cuando $m \to \infty$, así también asumimos la existencia de una sucesión acotada $\{\theta_{\alpha_m}\}_m$ en $B_W(X)$ de funciones de restricción (esto es, $\sup_m \|\theta_{\alpha_m}\|_W < \infty$), $\theta(\cdot) \in B_W(X)$ tales que $(1-\alpha_m)\theta_{\alpha_m}(x) \to \theta(x)$ cuando $m \to \infty$ para toda $x \in X$, y una sucesión $\{\lambda_m\}_m$ de números no positivos de tal forma que $\lambda_m \to \lambda_0 \in (-\infty, 0]$ cuando $m \to \infty$. Luego:*

(a) *Para cada $z \in X$ la sucesión $\{(1-\alpha_m)V_{\alpha_m}^*(z, r_{\alpha_m}^{\lambda_m})\}_m$ converge a una constante denotada $\rho_\theta^{\lambda_0}$, la cual es independiente de z, esto es,*

$$\rho_\theta^{\lambda_0} = \lim_{m\to\infty} (1-\alpha_m)V_{\alpha_m}^*(z, r_{\alpha_m}^{\lambda_m}) \quad \textit{para toda } z \in X. \tag{4.4.28}$$

Además, existen funciones $\overline{h}, \underline{h} \in B_W(X)$ que satisfacen las desigualdades de optimalidad

$$\overline{h}(x) + \rho_\theta^{\lambda_0} \leq \max_{a\in A(x)} \left\{ r(x,a) + \lambda_0[c(x,a) - \theta(x)] + \int_X \overline{h}(y) Q(dy|x,a) \right\}, \tag{4.4.29}$$

y

$$\underline{h}(x) + \rho_\theta^{\lambda_0} \geq \max_{a\in A(x)} \left\{ r(x,a) + \lambda_0[c(x,a) - \theta(x)] + \int_X \underline{h}(y) Q(dy|x,a) \right\}, \tag{4.4.30}$$

para todo $x \in X$.

(b) *La constante $\rho_\theta^{\lambda_0}$ coincide con el valor óptimo promedio sin restricciones*

$$\rho_\theta^{\lambda_0} = \max_{\pi\in\Pi} J(x, \pi, r^{\lambda_0}) \quad \textit{para toda } x \in X, \tag{4.4.31}$$

donde $r^{\lambda_0}(x,a) = r(x,a) + \lambda_0[c(x,a) - \theta(x)]$ *y*

$$J(x,\pi,r^{\lambda_0}) = \liminf_{n\to\infty} \frac{1}{n} E_x^\pi \sum_{k=0}^{n-1} r^{\lambda_0}(x_k,a_k).$$

Más aún, si $\mathcal{F}_\theta$ *es no vacío, luego*

$$\rho_\theta^{\lambda_0} \geq \sup_{\pi\in\mathcal{F}_\theta} J(x,\pi,r) \quad \textit{para toda } x \in X. \tag{4.4.32}$$

(c) *Para cada* $m = 1, \cdots$, *sea* $f_{\alpha_m}^{\lambda_m} \in \mathbb{F} \cap \Pi_{\alpha_m}^{\lambda_m}$ *la política óptima determinista* α_m*-descontada para el PDSR definido en la observación 3.3.1. Supongamos que existe* $f \in \mathbb{F}$ *tal que* $f_{\alpha_m}^{\lambda_m}(x) \to f(x)$ *para toda* $x \in X$. *Luego*

$$\rho_\theta^{\lambda_0} = \mu_f(r_f^{\lambda_0}) = g(r,f) + \lambda_0[g(c,f) - \mu_f(\theta)]. \tag{4.4.33}$$

Demostración

Prueba de (a). Considere la función de ganancia óptima α_m-descontada para el PDSR definida en la definición 3.2.1, donde $V_{\alpha_m}^*(\cdot, r_{\alpha_m}^{\lambda_m}) \in B_W(X)$. Por la proposición 3.3.1(i)-(ii), esta función satisface la α_m-EGDO.

$$V_{\alpha_m}^*(x, r_{\alpha_m}^{\lambda_m}) = \max_{a\in A(x)} \left\{ r_{\alpha_m}^{\lambda_m}(x,a) + \alpha_m \int_X V_{\alpha_m}^*(y, r_{\alpha_m}^{\lambda_m}) Q(dy|x,a) \right\} \tag{4.4.34}$$

$$= r_{\alpha_m f_{\alpha_m}^{\lambda_m}}^{\lambda_m}(x) + \alpha_m \int_X V_{\alpha_m}^*(y, r_{\alpha_m}^{\lambda_m}) Q_{f_{\alpha_m}^{\lambda_m}}(dy|x), \tag{4.4.35}$$

para toda $x \in X$, con $r_\alpha^\lambda(x,a) = r(x,a) + \lambda[c(x,a) - (1-\alpha)\theta_\alpha(x)]$, y donde $f_{\alpha_m}^{\lambda_m} \in \mathbb{F} \cap \Pi_{\alpha_m}^{\lambda_m}$ se obtiene de la proposición 3.3.1(ii). Usando las funciones $h_{\alpha_m}^{\lambda_m}(\cdot) \in B_W(X)$ definidas en (4.4.18), y sustituyéndolas en (4.4.34), obtenemos la ecuación de optimalidad

$$h_{\alpha_m}^{\lambda_m}(x) + (1-\alpha_m) V_{\alpha_m}^*(z, r_{\alpha_m}^{\lambda_m}) = \max_{a\in A(x)} \left\{ r_{\alpha_m}^{\lambda_m}(x,a) + \alpha_m \int_X h_{\alpha_m}^{\lambda_m}(y) Q(dy|x,a) \right\} \tag{4.4.36}$$

$$= r_{\alpha_m f_{\alpha_m}^{\lambda_m}}^{\lambda_m}(x) + \alpha_m \int_X h_{\alpha_m}^{\lambda_m}(y) Q_{f_{\alpha_m}^{\lambda_m}}(dy|x), \tag{4.4.37}$$

para toda $x \in X$. Por otro lado, por el hecho de que la sucesión $\{\theta_{\alpha_m}(\cdot)\}_m$ está acotada en $B_W(X)$, y $\lambda_m \to \lambda_0$ cuando $m \to \infty$, la sucesión $\{(1-\alpha_m)V^*_{\alpha_m}(z, r^{\lambda_m}_{\alpha_m})\}_m$ está contenida en un conjunto compacto, a saber, por (3.2.3) y (3.2.4), obtenemos

$$\left|(1-\alpha_m)V^*_{\alpha_m}(z, r^{\lambda_m}_{\alpha_m})\right| \leq M\left(\frac{1+b}{1-\beta}\right)\left[1+\sup_m|\lambda_m|\left(1+\sup_m\|\theta_{\alpha_m}\|_W/M\right)\right]W(z) < \infty. \tag{4.4.38}$$

Luego, existe una constante $\rho^{\lambda_0}_\theta$ y una subsucesión $\{\alpha_{m_k}\}_k$ (la cual denotaremos de nuevo por $\{\alpha_m\}_m$) tal que

$$\rho^{\lambda_0}_\theta := \lim_{m\to\infty}(1-\alpha_m)V^*_{\alpha_m}(z, r^{\lambda_m}_{\alpha_m}). \tag{4.4.39}$$

Usando el lema 4.4.1, podemos probar que la sucesión $\{h^{\lambda_m}_{\alpha_m}(\cdot)\}_m$ está acotada en $B_W(X)$. Definamos

$$\underline{h}(x) := \liminf_{m\to\infty} h^{\lambda_m}_{\alpha_m}(x), \quad \text{y} \quad \overline{h}(x) := \limsup_{m\to\infty} h^{\lambda_m}_{\alpha_m}(x).$$

Denotemos, para cada $m = 1, 2, \cdots$, y $(x,a) \in \mathbb{K}$,

$$v_m(x,a) := \sup_{n\geq m} r^{\lambda_n}_{\alpha_n}(x,a) + \int_X \sup_{n\geq m} h^{\lambda_n}_{\alpha_n}(y)Q(dy|x,a), \tag{4.4.40}$$

y

$$v_0(x,a) := r(x,a) + \lambda_0[c(x,a) - \theta(x)] + \int_X \overline{h}(y)Q(dy|x,a). \tag{4.4.41}$$

Bajo nuestras hipótesis, $\{v_m\}_m$ es una sucesión acotada en $B_W(\mathbb{K})$ tal que $v_m(x,a) \downarrow v_0(x,a)$ sobre $\mathbb{K}$. Más aún, del lema 4.3.2(a) y la hipótesis 2.2.1(b), las funciones $a \mapsto v_m(x,a)$ son continuas en $A(x)$ para cada $m = 0, 1, \cdots$, y para cada $x \in X$. Así, la sucesión $\{v_m\}_m$ satisface las hipótesis del lema 4.4.2, por lo que

$$\begin{aligned}\lim_{m\to\infty}\max_{a\in A(x)}\left\{\sup_{n\geq m} r^{\lambda_n}_{\alpha_n}(x,a) + \int_X \sup_{n\geq m} h^{\lambda_n}_{\alpha_n}(y)Q(dy|x,a)\right\} = \\ = \max_{a\in A(x)}\left\{r(x,a) + \lambda_0[c(x,a)-\theta(x)] + \int_X \overline{h}(y)Q(dy|x,a)\right\},\end{aligned} \tag{4.4.42}$$

Note que para cada $m \geq 1$, $\sup_{a\in A(x)} v_m(x,a)$ acota el lado derecho de (4.4.36). De aquí, tomando $\limsup_{m\to\infty}$ en ambos lados de (4.4.36), por las relaciones (4.4.39) y (4.4.42), obtenemos la desigualdad (4.4.29). Por otro lado, tomando $\liminf_{m\to\infty}$ en ambos lados de (4.4.36), de (4.3.7) en el lema 4.3.2(b), obtenemos la desigualdad (4.4.30).

Prueba de (b). Sea $f \in \mathbb{F}$ tal que alcanza el máximo en el lado derecho de (4.4.29). Luego,

$$\overline{h}(x) + \rho_\theta^{\lambda_0} \leq r_f(x) + \lambda_0[c_f(x) - \theta(x)] + \int_X \overline{h}(y) Q_f(dy|x) \quad \text{para toda } x \in X. \tag{4.4.43}$$

Integrando ambos lados de (4.4.43) por la medida de probabilidad invariante correspondiente μ_f, obtenemos

$$\rho_\theta^{\lambda_0} \leq J(x, f, r^{\lambda_0}) = g(f, r) + \lambda_0[g(f, c) - \mu_f(\theta)]. \tag{4.4.44}$$

Por otro lado, de (4.4.30) tenemos

$$\underline{h}(x) + \rho_\theta^{\lambda_0} \geq r(x, a) + \lambda_0[c(x, a) - \theta(x)] + \int_X \underline{h}(y) Q(dy|x, a)$$

para cada par de estado y acción admisible $(x, a) \in \mathbb{K}$. Sea $\pi \in \Pi$ una política arbitraria, $x \in X$ un estado inicial, y $\{(x_n, a_n)\}_n$ el proceso estocástico correspondiente. Luego, para cada $n = 1, 2, \cdots$,

$$E_x^\pi \underline{h}(x_{n-1}) + \rho_\theta^{\lambda_0} \geq E_x^\pi r^{\lambda_0}(x_{n-1}, a_{n-1}) + E_x^\pi h(x_n).$$

Realizando iteraciones de esta desigualdad obtenemos

$$E_x^\pi \sum_{k=0}^{n-1} \underline{h}(x_k) + n\rho_\theta^{\lambda_0} \geq E_x^\pi \sum_{k=0}^{n-1} r^{\lambda_0}(x_k, a_k) + E_x^\pi \sum_{k=1}^{n} \underline{h}(x_k),$$

o, equivalentemente,

$$\frac{1}{n}\underline{h}(x) - \frac{1}{n} E_x^\pi \underline{h}(x_n) + \rho_\theta^{\lambda_0} \geq \frac{1}{n} E_x^\pi \sum_{k=0}^{n-1} r^{\lambda_0}(x_k, a_k). \tag{4.4.45}$$

De (2.2.2) y el hecho que $\underline{h} \in B_W(X)$, obtenemos

$$\frac{1}{n} E_x^\pi |\underline{h}(x_n)| \leq \frac{\|\underline{h}\|_W}{n} E_x^\pi W(x_n) \leq \frac{\|\underline{h}\|_W}{n} \left[\beta^n W(x) + \frac{b}{1-\beta}\right] \to 0 \quad \text{cuando } n \to \infty.$$

Por lo tanto, tomando el límite en (4.4.45) cuando $n \to \infty$, obtenemos

$$\rho_\theta^{\lambda_0} \geq \liminf_{n\to\infty} \frac{1}{n} E_x^\pi \sum_{k=0}^{n-1} r^{\lambda_0}(x_k, a_k) = J(x, \pi, r^{\lambda_0}). \tag{4.4.46}$$

Ya que $\pi \in \Pi$ es una política arbitraria, (4.4.44) y (4.4.46) prueban (4.4.31), esto es

$$\rho_\theta^{\lambda_0} = \max_{\pi \in \Pi} J(x, \pi, r^{\lambda_0}) \quad \text{para toda } x \in X.$$

probando (4.4.31). Además, de (4.4.39) y (4.4.31), obtenemos (4.4.28), es decir el límite considerado se cumple para toda z. Por otro lado, puesto que $\lambda_0 \leq 0$, notemos que

$$\begin{aligned} J(x,\pi,r^{\lambda_0}) &= \liminf_{n\to\infty} \frac{1}{n} E_x^\pi \sum_{k=0}^{n-1} \left(r(x_k,a_k) + \lambda_0[c(x_k,a_k) - \theta(x_k)]\right) \\ &\geq \liminf_{n\to\infty} \frac{1}{n} E_x^\pi \sum_{k=0}^{n-1} r(x_k,a_k) + \\ &+ \lambda_0 \left[\limsup_{n\to\infty} \frac{1}{n} E_x^\pi \sum_{k=0}^{n-1} c(x_k,a_k) - \liminf_{n\to\infty} \frac{1}{n} E_x^\pi \sum_{k=0}^{n-1} \theta(x_k) \right] = \\ &= J(x,\pi,r) + \lambda_0[J(x,\pi,c) - \overline{\theta}(x,\pi)] \quad \text{para toda} \quad x \in X. \end{aligned} \tag{4.4.47}$$

Considerando que $\mathcal{F}_\theta$ es no vacío y $\lambda_0 \leq 0$, luego $J(x,\pi,c) \leq \overline{\theta}(x,\pi)$, y $\lambda_0[J(x,\pi,c) - \overline{\theta}(x,\pi)] \geq 0$ para toda $x \in X$ y para cada $\pi \in \mathcal{F}_\theta$. Así, de (4.4.47) y (4.4.31) obtenemos (4.4.32).

Prueba de (c). Tomando $\limsup_{m\to\infty}$ en (4.4.37), de las hipótesis 2.2.1(b), y de los límites $f_{\alpha_m}^{\lambda_m}(x) \to f(x)$ y $(1-\alpha_m)\theta_{\alpha_m}(x) \to \theta(x)$ cuando $m \to \infty$, para toda $x \in X$, por (4.3.8) en el lema 4.3.2, obtenemos

$$\overline{h}(x) + \rho_\theta^{\lambda_0} \leq r_f^{\lambda_0} + \int_X \overline{h}(y) Q_f(dy|x) \quad \text{para toda } x \in X.$$

Integrando ambos lados de la última desigualdad por la medida invariante de probabilidad μ_f y considerando (4.4.31), se tiene finalmente

$$\rho_\theta^{\lambda_0} = \mu_f(r_f^{\lambda_0}) = g(f,r) + \lambda_0[g(f,c) - \mu_f(\theta)].$$

■

El siguiente corolario es una consecuencia directa del teorema 4.4.1.

Corolario 4.4.1 *Bajo las hipótesis 2.2.1, 2.2.2, y 4.3.1 del teorema 4.4.1, dada cualquier función arbitraria $\theta(\cdot) \in B_W(X)$, entonces para cada $\lambda \leq 0$, tenemos:*

(a) *Existe una constante ρ_θ^λ, y funciones $\overline{h}_\lambda, \underline{h}_\lambda \in B_W(X)$ que satisfacen las desigualdades de optimalidad*

$$\overline{h}_\lambda(x) + \rho_\theta^\lambda \leq \max_{a\in A(x)} \left\{ r(x,a) + \lambda[c(x,a) - \theta(x)] + \int_X \overline{h}_\lambda(y) Q(dy|x,a) \right\}, \tag{4.4.48}$$

y

$$\underline{h}_\lambda(x) + \rho_\theta^\lambda \geq \max_{a\in A(x)} \left\{ r(x,a) + \lambda[c(x,a) - \theta(x)] + \int_X \underline{h}_\lambda(y) Q(dy|x,a) \right\}, \tag{4.4.49}$$

para todo $x \in X$.

(b) *La constante* ρ_θ^λ *coincide con el valor óptimo del problema ergódico sin restricciones*

$$\rho_\theta^\lambda = \max_{\pi\in\Pi} J(x,\pi,r^\lambda) \quad \textit{para toda } x \in X, \tag{4.4.50}$$

donde $r^\lambda(x,a) = r(x,a) + \lambda[c(x,a) - \theta(x)]$, *y*

$$J(x,\pi,r^\lambda) = \liminf_{n\to\infty} \frac{1}{n} E_x^\pi \sum_{k=0}^{n-1} r^\lambda(x_k,a_k).$$

Además, si $\varphi \in \Phi$ *es una política aleatorizada que alcanza el máximo en el lado izquierdo de (4.4.48), esto es,*

$$\overline{h}_\lambda(x) + \rho_\theta^\lambda \leq r_\varphi(x) + \lambda[c_\varphi(x) - \theta(x)] + \int_X \overline{h}_\lambda(y) Q_\varphi(dy|x) \quad \textit{para toda } x \in X,$$

entonces φ *es una política óptima para este problema sin restricciones, vale decir,*

$$\rho_\theta^\lambda = \max_{\pi\in\Pi} J(x,\pi,r^\lambda) = J(x,\varphi,r^\lambda) = \mu_\varphi(r_\varphi^\lambda) \quad \textit{para toda} \quad x \in X.$$

Demostración En la demostración de los incisos (a), (b) del teorema 4.4.1, basta considerar las funciones de restricción $\theta_\alpha(\cdot) := \frac{\theta(\cdot)}{1-\alpha} \in B_W(X)$ para cada $0 < \alpha < 1$. Además, basta tomar cualquier sucesión $\{\alpha_m\}_m$ en $(0,1)$ tal que $\alpha_m \uparrow 1$. También definimos $\lambda_m := \lambda$ para toda $m = 1, 2\cdots$, para cualquier $\lambda \leq 0$. Posteriormente, basta imitar los pasos de la prueba dada para los incisos (a) y (b) del teorema 4.4.1.

■

Hemos llegado a nuestro principal resultado de esta sección concerniente a la existencia de políticas de control óptimas para el problema esperado con restricciones junto con una caracterización deseable de la ganancia esperada óptima promedio.

Teorema 4.4.2 *Supongamos válidas las hipótesis 2.2.1, 2.2.2, y 4.2.1. Sea* $\{\alpha_m\}_m$ *una sucesión que satisface* $\alpha_m \uparrow 1$, *una sucesión acotada* $\{\theta_{\alpha_m}(\cdot)\}_m$ *en* $B_W(X)$ *y* $\theta(\cdot) \in B_W(X)$ *tal que* $(1-\alpha_m)\theta_{\alpha_m}(x) \to \theta(x)$ *para toda* $x \in X$, *y sea* $z \in X$ *un estado fijo, y una sucesión* $\{\lambda_m^*\}_m$ *en* $(-\infty, 0)$ *tal que para cada* m, λ_m^* *es un punto crítico del mapeo* $\lambda \mapsto V_{\alpha_m}^*(z, r_{\alpha_m}^\lambda)$, *y* $f_{\alpha_m}^{\lambda_m^*} \in \mathbb{F} \cap \Pi_{\alpha_m}^{\lambda_m^*}$ *una política* α_m*-óptima para el PDSR (obtenido por la proposición 3.3.1(ii) y el teorema 3.4.1(b)). Supongamos también que* $\lambda_m^* \to \lambda_0^* \in (-\infty, 0]$, *y* $f_{\alpha_m}^{\lambda_m^*}(x) \to f^*(x)$ *para toda* $x \in X$, *con* $f^* \in \mathbb{F}$. *Entonces:*

(i) *La política determinista* f^* *satisface* $J(x, f^*, c) = g(f^*, c) = \underline{\theta}(x, f^*) = \mu_{f^*}(\theta)$ *para toda* $x \in X$, *de donde* $f^* \in \mathcal{F}_\theta$. *Además,* f^* *resulta ser una política óptima para el PERG, y la constante* $\rho_\theta^{\lambda_0^*}$ *en (4.4.29)-(4.4.30) coincide con el valor óptimo para el PERG, esto es*

$$V_\theta^*(x) = \rho_\theta^{\lambda_0^*} = J(x, f^*, r) = g(r, f^*) = \sup_{\pi \in \mathcal{F}_\theta} J(x, \pi, r) \quad \textit{para cada} \quad x \in X.$$

(ii) *El valor óptimo* $\rho_\theta^{\lambda_0^*}$ *satisface:*

$$\rho_\theta^{\lambda_0^*} = \min_{\lambda \le 0} \rho_\theta^\lambda. \tag{4.4.51}$$

Demostración Prueba de (i). Para cada $m \ge 1$, consideremos el multiplicador de Lagrange $\lambda_m^* < 0$, que bajo nuestras hipótesis es un punto crítico de la función $\lambda \mapsto V_{\alpha_m}^*(z, r_{\alpha_m}^\lambda)$, donde $V_{\alpha_m}^*(z, r_{\alpha_m}^\lambda)$ es la ganancia óptima α_m-descontada para el PDSR. Del Teorema 3.4.1(b), y la definición de $f_{\alpha_m}^{\lambda_m^*}$, para cada $m \ge 1$, tenemos

$$V_{\alpha_m}^*(z, r_{\alpha_m}^{\lambda_m^*}) = V_{\alpha_m}(z, f_{\alpha_m}^{\lambda_m^*}, r_{\alpha_m}^{\lambda_m^*}), \quad \text{y} \quad V_{\alpha_m}(z, f_{\alpha_m}^{\lambda_m^*}, c) = \overline{\theta}_{\alpha_m}(z, f_{\alpha_m}^{\lambda_m^*}), \tag{4.4.52}$$

De donde $V_{\alpha_m}^*(z, r_{\alpha_m}^{\lambda_m^*}) = V_{\alpha_m}(z, f_{\alpha_m}^{\lambda_m^*}, r)$, por lo que se sigue

$$(1-\alpha_m)V_{\alpha_m}^*(z, r_{\alpha_m}^{\lambda_m^*}) = (1-\alpha_m)V_{\alpha_m}(z, f_{\alpha_m}^{\lambda_m^*}, r), \tag{4.4.53}$$

y

$$(1-\alpha_m)V_{\alpha_m}(z, f_{\alpha_m}^{\lambda_m^*}, c) = (1-\alpha_m)\overline{\theta}_{\alpha_m}(z, f_{\alpha_m}^{\lambda_m^*}). \tag{4.4.54}$$

Tomando el límite cuando $m \to \infty$ en (4.4.53) y (4.4.54), de las proposiciones 4.3.1 y 4.3.2, y por (4.4.28) en el teorema 4.4.1, obtenemos

$$\rho_\theta^{\lambda_0^*} = g(f^*, r), \quad \text{y} \quad g(f^*, c) = \mu_{f^*}(\theta), \tag{4.4.55}$$

equivalentemente,

$$\rho_\theta^{\lambda_0^*} = J(x, f^*, r), \quad \text{y} \quad J(x, f^*, c) = \underline{\theta}(x, f^*), \quad \text{para toda } x \in X. \tag{4.4.56}$$

lo cual implica que $f^* \in \mathcal{F}_\theta$, por lo que $\mathcal{F}_\theta$ es un conjunto no vacío. Así, de (4.4.32) en el teorema 4.4.1 y la primera igualdad en (4.4.56), obtenemos que f^* es una política óptima para la PERG, y además la función de valor óptimo para la PERG, $V_\theta^*(\cdot)$, es constante y coincide con $\rho_\theta^{\lambda_0^*}$.

Prueba de (ii). De (4.4.49) en el corolario 4.4.1, para cada $\lambda \leq 0$ tenemos

$$\underline{h}_\lambda(x) + \rho_\theta^\lambda \geq r_{f^*}(x) + \lambda[c_{f^*}(x) - \theta(x)] + \int_X \underline{h}_\lambda(y) Q_{f^*}(dy|x) \quad \text{para toda } x \in X.$$

Integrando ambos lados de esta desigualdad con respecto a la medida invariante μ_{f^*}, obtenemos

$$\rho_\theta^\lambda \geq g(f^*, r) + \lambda[g(f^*, c) - \mu_{f^*}(\theta)].$$

Luego, por (4.4.55), tenemos que $\rho_\theta^\lambda \geq \rho_\theta^{\lambda_0^*}$ para cada $\lambda \leq 0$, de donde $\rho_\theta^{\lambda_0^*} = \text{mín}_{\lambda \leq 0}\, \rho_\theta^\lambda$.

■

Capítulo 5

Ejemplo: Sistema cuadrático-lineal LQ

En este capítulo presentaremos un sistema cuadrático-lineal (el cual nos referiremos como LQ) que satisface todas las hipótesis de los teoremas 3.4.1, 4.4.1, y 4.4.2.

5.1. Hipótesis y resultados importantes del sistema LQ con restricciones

Consideremos el sistema lineal

$$x_{t+1} = k_1 x_t + k_2 a_t + z_t, \quad t = 0, 1, \cdots, \tag{5.1.1}$$

con espacio de estados $X := \mathbb{R}$ y coeficientes positivos k_1, k_2. El conjunto de control es $A := \mathbb{R}$, y el conjunto de acciones admisibles en cada estado x es el intervalo

$$A(x) := [-k_1|x|/k_2, k_1|x|/k_2]. \tag{5.1.2}$$

El ruido z_t en (5.1.1) se modelará a través de variables aleatorias i.i.d. con valores en $Z := \mathbb{R}$, con media cero y varianza finita, esto es,

$$E(z_t) = 0 \quad \text{y} \quad \sigma^2 := E(z_t^2) < \infty. \tag{5.1.3}$$

Para completar la descripción de nuestro modelo de control con restricciones introducimos la función de ganacia cuadrática

$$r(x, a) := e - (r_1 x^2 + r_2 a^2) \quad \forall (x, a) \in \mathbb{K}, \tag{5.1.4}$$

con coeficientes positivos e, r_1, y r_2, y la función de costo

$$c(x,a) := c_1 x^2 + c_2 a^2 \quad \forall (x,a) \in \mathbb{K}, \tag{5.1.5}$$

con coeficientes positivos c_1, c_2. También definimos

$$W(x) := \exp[\zeta |x|] \quad \text{para toda} \quad x \in X, \tag{5.1.6}$$

con $\zeta \geq 2$. Más aún, sea $\hat{s} > 0$ tal que

$$\zeta \hat{s} < \log(\zeta/2 + 1)$$

lo cual implica

$$\beta := \frac{2}{\zeta}(\exp[\zeta \hat{s}] - 1) < 1.$$

Con este valor de β, tenemos que es válida la hipótesis 2.2.2-(c). Por otro lado, observemos que r, c son funciones en $B_W(\mathbb{K})$, y $W \geq 1$. De esto se sigue que valen las hipótesis 2.2.1 y 2.2.2.

Al igual que en [21, Sección 5], supondremos lo siguiente.

Hipótesis 5.1.1 $0 < k_1 < 1/2$.

Hipótesis 5.1.2 *El ruido z_t i.i.d. tienen densidad común $d(\cdot)$, la cual es una función acotada continua con soporte compacto $S := [-\hat{s}, \hat{s}]$. Más aún, existe un número positivo ε tal que $d(s) \geq \varepsilon$ para toda $s \in S$.*

Sea $S_0 := [0, \hat{s}]$, y sea Υ la medida de Lebesgue en $X = \mathbb{R}$. Definimos

$$l(x,a) := 1_{S_0}(x) \ \forall (x,a) \in \mathbb{K}, \text{ y } \nu(B) := \varepsilon \Upsilon(B \cap S_0) \ \forall B \in \mathcal{B}(\mathbb{R}).$$

Luego, tenemos que el sistema LQ (5.1.1)-(5.1.5) satisface los lemas 4.4-4.9 en [24]. En particular, es válida la hipótesis 4.2.1. Por esta razón se tiene la siguiente proposición.

Proposición 5.1.1 *Bajo las hipótesis 5.1.1 y 5.1.2, el sistema LQ (5.1.1)-(5.1.5) satisface las hipótesis 2.2.1, 2.2.2, y 4.2.1.*

Hipótesis 5.1.3 *Sea $\alpha \in (0,1)$. Consideremos la de restricción cuadrática $\theta_\alpha : \mathbb{R} \to \mathbb{R}$ definida por*

$$\theta_\alpha(x) = \theta_{2,\alpha}x^2 + \theta_{0,\alpha}, \quad \text{para toda } x \in \mathbb{R}, \quad \text{con } \theta_{2,\alpha} < \frac{c_1}{1-\alpha}, \tag{5.1.7}$$

donde $\theta_{2,\alpha}$ y $\theta_{0,\alpha}$ constantes dadas.

Notemos que $\theta_\alpha(\cdot)$ pertenece al espacio normado $B_W(\mathbb{R})$, por lo que se cumple la hipótesis 2.3.1. Además, la ganancia generalizada r_α^λ en (3.1.1) puede ser escrita como

$$r_\alpha^\lambda(x,a) = r_{1,\alpha}(\lambda)x^2 + r_{2,\alpha}(\lambda)a^2 + b_\alpha(\lambda), \tag{5.1.8}$$

donde

$$\begin{aligned} &r_{1,\alpha}(\lambda) := \lambda[c_1 - (1-\alpha)\theta_{2,\alpha}] - r_1 < 0, \quad r_{2,\alpha}(\lambda) := \lambda c_2 - r_2 < 0, \\ &\text{y} \quad b_\alpha(\lambda) = e - \lambda(1-\alpha)\theta_{0,\alpha}. \end{aligned} \tag{5.1.9}$$

La siguiente proposición nos dará la forma explícita de los valores óptimos $\alpha-$descontados así como las políticas óptimas corresondientes a los multiplicadores λ.

Proposición 5.1.2 *Supongamos válidas las hipótesis 5.1.1, 5.1.2, y 5.1.3. Entonces:*

(i) *Sea $\alpha \in (0,1)$, $\lambda \leq 0$, θ_α como en (5.1.7), y $x \in \mathbb{R}$ un estado inicial. La ganancia óptima α-descontada $V_\alpha^*(x, r_\alpha^\lambda)$ para el sistema LQ (5.1.1)-(5.1.5), la cual satisface la α-EGDO (3.3.1) en la proposición 3.3.1, tiene la forma*

$$V_\alpha^*(x, r_\alpha^\lambda) = v_\alpha(\lambda)\left[x^2 + \frac{\alpha\sigma^2}{1-\alpha}\right] + \frac{b_\alpha(\lambda)}{1-\alpha}, \tag{5.1.10}$$

con σ como en (5.1.3), y $v_\alpha(\lambda)$ es la única solución negativa a la ecuación cuadrática (también llamada ecuación de Ricatti)

$$\alpha k_2^2 v_\alpha(\lambda)^2 + \left[r_{2,\alpha}(\lambda) - \alpha k_2^2 r_{1,\alpha}(\lambda) - \alpha k_1^2 r_{2,\alpha}(\lambda)\right] v_\alpha(\lambda) - r_{1,\alpha}(\lambda) r_{2,\alpha}(\lambda) = 0, \tag{5.1.11}$$

con $r_{1,\alpha}(\lambda)$, $r_{2,\alpha}(\lambda)$, $b_\alpha(\lambda)$ dados como en (5.1.9). Además, $\lambda \mapsto V_\alpha^*(x, r_\alpha^\lambda)$ es diferenciable en $(-\infty, 0)$ y continua en $(-\infty, 0]$, y la derivada cuando $\Lambda < 0$ satisface (3.3.6).

(ii) *Existe una única política determinista $f_\alpha^\lambda \in \mathbb{F}$ que maximiza el lado derecho de (3.3.1), y la cual es una política óptima α-descontada para el PDSR, esto es, $f_\alpha^\lambda \in \mathbb{F} \cap \Pi_\alpha^\lambda$, dada por*

$$f_\alpha^\lambda(x) = -\widehat{f}_\alpha^\lambda \; x \in A(x) \quad \text{para toda } x \in \mathbb{R}, \text{ donde } \widehat{f}_\alpha^\lambda = \frac{\alpha k_1 k_2 v_\alpha(\lambda)}{r_{2,\alpha}(\lambda) + \alpha k_2^2 v_\alpha(\lambda)} > 0, \tag{5.1.12}$$

con $A(x)$ dada por (5.1.2).

(iii) *Para cada estado inicial* x

$$V_\alpha(x, f_\alpha^\lambda, c) = \left(\frac{c_1 + c_2(\widehat{f}_\alpha^\lambda)^2}{1 - \alpha\widehat{k}_\alpha(\lambda)^2}\right)\left[x^2 + \frac{\alpha\sigma^2}{1-\alpha}\right], \tag{5.1.13}$$

y

$$\overline{\theta}_\alpha(x, f_\alpha^\lambda) = \left(\frac{(1-\alpha)\theta_{2,\alpha}}{1 - \alpha\widehat{k}_\alpha(\lambda)^2}\right)\left[x^2 + \frac{\alpha\sigma^2}{1-\alpha}\right] + \theta_{0,\alpha}, \tag{5.1.14}$$

donde $\widehat{k}_\alpha(\lambda) := k_1 - k_2\widehat{f}_\alpha^\lambda \in (0,1)$. *Además, los mapeos* $\lambda \mapsto V_\alpha(x, f_\alpha^\lambda, c)$, $\lambda \mapsto \overline{\theta}_\alpha(x, f_\alpha^\lambda)$ *son continuos en* $(-\infty, 0]$.

Demostración Prueba de (i) y (ii). Estas se tienen resolviendo la α-EGDO (3.3.1) para el sistema LQ (5.1.1)-(5.1.5).

Prueba de (iii). La parte (iii) de este lema es una consecuencia directa del lema 5.1.1 dado a continuación, y el hecho que $\lambda \mapsto v_\alpha(\lambda)$ es continua. ∎

Lema 5.1.1 *Sea* $\widehat{f}$ *una constante, y sea* $f \in \mathbb{F}$ *una política determinista dada por* $f(x) := -\widehat{f}x \in A(x)$ *para toda* $x \in \mathbb{R}$. *Además, sea* $\widehat{k} := k_1 - k_2\widehat{f}$, *donde* k_1, k_2 *son los coeficientes dados en (5.1.1). Supongamos que* $|\widehat{k}| < 1$. *Entonces, para toda* $x \in \mathbb{R}$,

$$E_x^f(x_t^2) = \widehat{k}^{2t}x^2 + \frac{(1-\widehat{k}^{2t})\sigma^2}{1-\widehat{k}^2}, \quad \textit{para toda } t = 0, 1, \cdots, \tag{5.1.15}$$

y

$$E_x^f\left[\sum_{t=0}^{\infty}\alpha^t x_t^2\right] = \left(\frac{1}{1-\alpha\widehat{k}^2}\right)\left[x^2 + \frac{\alpha\sigma^2}{1-\alpha}\right]. \tag{5.1.16}$$

Demostración Remplazando a_t en (5.1.1) con $a_t := f(x_t) = -\widehat{f}x_t$, obtenemos

$$x_t = (k_1 - k_2\widehat{f})x_{t-1} + z_{t-1} = \widehat{k}x_{t-1} + z_{t-1} \quad \forall t = 1, 2, \cdots.$$

Por un proceso inductivo, obtenemos que para toda $t = 1, 2, \cdots$,

$$x_t = \widehat{k}^t x_0 + \sum_{j=0}^{t-1}\widehat{k}^j z_{t-1-j}.$$

De esta relación se siguen (5.1.15) y (5.1.16). ■

Para cada $\alpha \in (0,1)$, $\lambda \leq 0$, y un estado inicial $x \in \mathbb{R}$, definimos

$$h_\alpha(x,\lambda) := V_\alpha(x, f_\alpha^\lambda, c) - \overline{\theta}_\alpha(x, f_\alpha^\lambda) = \tag{5.1.17}$$

$$= \left(\frac{c_1 + c_2(\widehat{f}_\alpha^\lambda)^2 - (1-\alpha)\theta_{2,\alpha}}{1 - \alpha \widehat{k}_\alpha(\lambda)^2} \right) \left[x^2 + \frac{\alpha\sigma^2}{1-\alpha} \right] - \theta_{0,\alpha}.$$

Luego, de la igualdad (3.3.5), el mapeo $\lambda \mapsto h_\alpha(x,\lambda)$ es creciente en $(-\infty, 0]$. Así,

$$-\infty < h_\alpha(x,-\infty) := \inf_{\Lambda \leq 0} h(x,\Lambda) \leq h_\alpha(x,\lambda) \leq h_\alpha(x,0) < \infty \quad \text{para toda } \lambda \leq 0. \tag{5.1.18}$$

Por un cálculo elemental

$$h_\alpha(x,-\infty) = V_\alpha(x, f_\alpha^{-\infty}, c) - \overline{\theta}_\alpha(x, f_\alpha^{-\infty}),$$

con

$$f_\alpha^{-\infty}(x) := -\widehat{f}_\alpha^{-\infty}\ x \in A(x) \quad \text{para toda } x \in \mathbb{R}, \quad \widehat{f}_\alpha^{-\infty} := \frac{\alpha k_1 k_2 v_\alpha(-\infty)}{c_2 + \alpha k_2^2 v_\alpha(-\infty)} > 0,$$

donde $v_\alpha(-\infty) := \lim_{\lambda \to -\infty} v_\alpha(\lambda)/\lambda$.

Teorema 5.1.1 *Suponga válidas las hipótesis 5.1.1, 5.1.2, y 5.1.3. Sea $\alpha \in (0,1)$ un factor de descuento y x un estado inicial fijo. Entonces:*

(a) *Un número $\lambda^* < 0$ es un punto crítico de la función $\lambda \mapsto V_\alpha^*(x, r_\alpha^\lambda)$ si y sólo si λ^* es una raíz de la función $\lambda \mapsto h_\alpha(x,\lambda)$. Si este es el caso, $f_\alpha^{\lambda^*}$ es α-óptimo para el PDR. También tenemos que $V_\alpha(x, f_\alpha^{\lambda^*}, c) = \overline{\theta}_\alpha(x, f_\alpha^{\lambda^*})$, y $V_\alpha^*(x, r_\alpha^{\lambda^*})$ es el valor α-óptimo para el PDR, el cual coincide con $V_\alpha(x, f_\alpha^{\lambda^*}, r)$. Además,*

$$V_\alpha^*(x, r_\alpha^{\lambda^*}) = \inf_{\lambda \leq 0} V_\alpha^*(x, r_\alpha^\lambda).$$

(b) *Si $h_\alpha(x,-\infty) < 0 < h_\alpha(x,0)$, entonces existe un punto crítico $\lambda^* < 0$ del mapeo $\lambda \mapsto V_\alpha^*(x, r_\alpha^\lambda)$.*

(c) *Si $h_\alpha(x,0) \leq 0$, la política f_α^0 está en $\mathcal{F}_{\theta_\alpha}^x$, y es una política α-óptima para el α-PDR. Además, $V_\alpha^*(x, r_\alpha^0) = V_\alpha^*(x,r)$ resulta ser el valor α-óptimo par el PDR y coincide con $V_\alpha(x, f_\alpha^0, r)$, y satisface*

$$V_\alpha^*(x, r_\alpha^0) = \inf_{\lambda \leq 0} V_\alpha^*(x, r_\alpha^\lambda).$$

Demostración Prueba de (a). De la proposición 5.1.2-(i), el mapeo $\lambda \mapsto V_\alpha^*(x, r_\alpha^\lambda)$ es diferenciable en $(-\infty, 0)$, y de (3.3.6) del lema 3.3.1, y la definición de $h_\alpha(x, \lambda)$ en (5.2.1),

$$\frac{\partial V_\alpha^*(x, r_\alpha^\lambda)}{\partial \lambda} = h_\alpha(x, \lambda) \quad \text{para toda } \lambda < 0.$$

De lo cual, $\lambda^* < 0$ es un punto crítico de $\lambda \mapsto V_\alpha^*(x, r_\alpha^\lambda)$ si y sólo si λ^* es una raíz del mapeo $\lambda \mapsto h_\alpha(x, \lambda)$. El resto de la prueba de sigue del teorema 3.4.1-(b).

Prueba de (b). Esta parte se sigue del inciso (a) anterior, la definición $h_\alpha(x, -\infty)$ y la continuidad de $\lambda \mapsto h_\alpha(x, \lambda)$.

Prueba de (c). Ya que $h_\alpha(x, 0) \leq 0$ es equivalente a $V_\alpha(x, f_\alpha^0, c) \leq \overline{\theta}_\alpha(x, f_\alpha^0)$, donde $f_\alpha^0 \in \mathbb{F} \cap \Pi_\alpha^0$. Así, esta parte se sigue del teorema 3.4.1-(c). ∎

A continuación se presentará un ejemplo numérico.

5.2. Ejemplo numérico

Considere el sistema LQ (5.1.1)-(5.1.5) de tal forma que la función de ganancia (5.1.4) y la función de costo (5.1.5) satisfacen $r_1 = 1, r_2 = 2$, $e = 10$, y $c_1 = 2$, $c_2 = 1$, respectivamente. Más aún, supongamos que $k_1 = 1/3$, $k_2 = 1$ en (5.1.1), y $\sigma^2 = 1$ en (5.1.3). También, para cada $\alpha \in (0, 1)$, definimos la función de restricción $\theta_\alpha(\cdot) = \theta_{2,\alpha} x^2 + \theta_{0,\alpha}$ de tal forma que se cumpla (b) del teorema 5.1.1. Entonces la condición

$$h_\alpha(x, -\infty) < 0 < h_\alpha(x, 0)$$

equivale a que se cumpla

$$L_\alpha(-\infty) := (1-\alpha)\left(\frac{c_1 + c_2(\widehat{f}_\alpha^{-\infty})^2 - (1-\alpha)\theta_{2,\alpha}}{1 - \alpha\widehat{k}_\alpha(-\infty)^2}\right)\left[x^2 + \frac{\alpha\sigma^2}{1-\alpha}\right] < (1-\alpha)\theta_{0,\alpha} := \theta_0$$

$$< (1-\alpha)\left(\frac{c_1 + c_2(\widehat{f}_\alpha^{0})^2 - (1-\alpha)\theta_{2,\alpha}}{1 - \alpha\widehat{k}_\alpha(0)^2}\right)\left[x^2 + \frac{\alpha\sigma^2}{1-\alpha}\right] := L_\alpha(0).$$

Por ejemplo, si $\alpha = 0{,}95$, con $x = 1$ y $(1-\alpha)\theta_{2,\alpha} = 1$, tenemos que un cálculo directo nos muestra que $L_\alpha(-\infty) = 1{,}05563469...$ y $L_\alpha(0) = 1{,}06180858....$ Por esta razón, basta elegir $\theta_0 = 191/180 = 1{,}06111111...$ para que la aplicación $\lambda \mapsto V_{0,95}^*(1, r_{0,95}^\lambda)$ tenga un punto crítico de

acuerdo al teorema 5.1.1(b). Más aún, elijamos para toda $0 < \alpha < 1$ las funciones de restricción θ_α como

$$(1-\alpha)\theta_\alpha(x) = x^2 + \frac{191}{180}, \quad \text{para toda } x \in \mathbb{R}.$$

Usando esta elección para las funciones de restricción, no es más que una rutina verificar que $h_\alpha(1,-\infty) < 0 < h_\alpha(1,0)$ para todo $0{,}95 \leq \alpha < 1$. Por el teorema 5.1.1(a)-(b), podemos obtener una sucesión $\{\alpha_m\}_m$ en $(0,1)$ tal que $\alpha_m \uparrow 1$, y una sucesión $\{\lambda_m^*\}_m$ en $(-\infty, 0)$ tal que para cada m, λ_m^* es un punto crítico del mapeo $\lambda \mapsto V_{\alpha_m}(1, r^\lambda_{\alpha_m})$ (equivalentemente, $h_{\alpha_m}(1,\lambda_m^*) = 0$). Ahora, consideremos la sucesión $\{f^{\lambda_m^*}_{\alpha_m}\}_m$, tal que para cada m, $f^{\lambda_m^*}_{\alpha_m}(\cdot) \in \mathbb{F} \cap \Pi^{\lambda_m^*}_{\alpha_m}$ es una política α_m-óptima para el PDR, donde por (5.1.12) en la proposición 5.1.2, toma la forma $f^{\lambda_m^*}_{\alpha_m}(x) = -\widehat{f}^{\lambda_m^*}_{\alpha_m} x$ para toda $x \in \mathbb{R}$. Luego, obtenemos,

$$\lambda_m^* \to \lambda_0 := -0{,}38767819\cdots, \quad \widehat{f}^{\lambda_m^*}_{\alpha_m} \to \widehat{f}^* := 0{,}12806245\cdots, \tag{5.2.1}$$

y

$$(1-\alpha_m)V^*_{\alpha_m}(1, r^{\lambda_m^*}_{\alpha_m}) \to \rho_\theta^{\lambda_0} = 8{,}92176746\cdots, \tag{5.2.2}$$

cuando $m \to \infty$. Así, $f^{\lambda_m^*}_{\alpha_m}(x) \to f^*(x) \in A(x)$ cuando $m \to \infty$, para cada $x \in \mathbb{R}$, con $f^*(x) = -\widehat{f}^* x$. Del teorema 4.4.2, la política determinista f^* está en $\mathcal{F}_\theta$, y resulta ser una política óptima para el PERG. Además, del teorema 4.4.1 y 4.4.2, la constante $\rho_\theta^{\lambda_0} = 8{,}92176746\cdots$ es el valor óptimo para el PERG, esto es,

$$\rho_\theta^{\lambda_0} = 8{,}92176746\cdots = \sup_{\pi \in \mathcal{F}_\theta} J(x,\pi,r) = J(x,f^*,r) \quad \text{para toda } x \in \mathbb{R}.$$

Finalmente, de las proposiciones 4.3.1 y 4.3.2, obtenemos

$$(1-\alpha_m)V_{\alpha_m}(1, f^{\lambda_m^*}_{\alpha_m}, c) \to J(x,f^*,c) = g(f^*,c) = 2{,}10510079\cdots \quad \text{para toda } x \in \mathbb{R},$$

$$(1-\alpha_m)\overline{\theta}_{\alpha_m}(1, f^{\lambda_m^*}_{\alpha_m}) \to \underline{\theta}(x,f^*) = \mu_{f^*}(\theta) = 2{,}10510079\cdots, \quad \text{para toda } x \in \mathbb{R},$$

comprobando de paso que $J(x,f^*,c) = \underline{\theta}(x,f^*) = 2{,}10510079\cdots$. En la siguiente tabla consignamos algunos cálculos que nos dan idea de los límites anteriores cuando el factor de descuento $\alpha \uparrow 1$.

α	$h_\alpha(1,\lambda^*)=0$	$\widehat{f}_\alpha^{\lambda^*}$	$(1-\alpha)V_\alpha^*(1,r_\alpha^{\lambda^*})$	$(1-\alpha)V_\alpha(1,f_\alpha^{\lambda^*},c)$
0,950000	$-0{,}09102632\cdots$	$0{,}11580229\cdots$	$8{,}92484750\cdots$	$2{,}10818083\cdots$
0,960000	$-0{,}14200398\cdots$	$0{,}11823179\cdots$	$8{,}92426036\cdots$	$2{,}10759369\cdots$
0,970000	$-0{,}19657718\cdots$	$0{,}12066968\cdots$	$8{,}92365962\cdots$	$2{,}10699295\cdots$
0,980000	$-0{,}25526817\cdots$	$0{,}12311866\cdots$	$8{,}92304452\cdots$	$2{,}10637785\cdots$
0,990000	$-0{,}31870988\cdots$	$0{,}12558175\cdots$	$8{,}92241416\cdots$	$2{,}10574749\cdots$
0,999000	$-0{,}38050692\cdots$	$0{,}12781348\cdots$	$8{,}92183290\cdots$	$2{,}10516623\cdots$
0,999900	$-0{,}38695817\cdots$	$0{,}12803755\cdots$	$8{,}92177401\cdots$	$2{,}10510735\cdots$
0,999990	$-0{,}38760615\cdots$	$0{,}12805996\cdots$	$8{,}92176812\cdots$	$2{,}10510145\cdots$
0,999999	$-0{,}38767098\cdots$	$0{,}12806220\cdots$	$8{,}92176753\cdots$	$2{,}10510086\cdots$

Capítulo 6

Conclusiones

En este trabajo de tesis estudiamos ganancias descontadas y ganancias esperadas promedio de procesos de control de Markov a tiempo discreto sobre espacios Borel, con restricciones en costos descontados y costos promedio, respectivamente. Nuestros principales resultados obtenidos incluyen:

(a) existencia de políticas óptimas para la ganancia α-descontada con restricciones en el costo α-descontado,

(b) método para calcular políticas óptimas y el valor óptimo para el caso descontado,

(c) existencia de políticas óptimas para la ganancia promedio esperada con restricciones en el costo promedio esperado,

(d) método para calcular las políticas óptimas y el valor óptimo para el caso promedio.

Para analizar nuestros problemas procedimos esencialmente a través de dos pasos fundamentales:

En la primer paso se estudió el problema de control α-descontado con en el costo descontado, y se proporcionó un método basado en la técnica de los multiplicadores de Lagrange para la optimización de una familia paramétrica de problemas de control α-descontado sin restricciones. Para ello se utilizaron argumentos basados en la programación dinámica a través de ecuaciones de optimalidad α-descontadas de los problemas sin restricciones, y que conjuntamente con el cálculo diferencial de una variable, permitió encontrar puntos críticos a las funciones de valor óptimo asociados a esta familia paramétrica. Dichos puntos críticos corresponden al multiplicador adecuado, y al mismo tiempo, proporcionan un conjunto de políticas óptimas asociadas al problema sin restricciones parametrizada por dicho multiplicador adecuado. Más aún, estas políticas

también resultan ser óptimas para nuestro problema original α-descontado con restricciones, y el valor óptimo correspondiente coincide con el valor óptimo del problema sin restricciones asociado a dicho parámetro.

En el segundo paso, utilizamos la técnica de aproximación desvanescente de los problemas α-descontados sin restricciones cuando $\alpha \uparrow 1$, parametrizados con multiplicadores de Lagrange que son los puntos críticos encontrados a través de la técnica empleada en el primer paso. Esto conduce a que los valores óptimos α-descontados converjan al valor óptimo del problema promedio esperado con restricciones, y que, bajo nuestras hipótesis, las políticas óptimas α-descontadas converjan a una política óptima para el caso promedio con restricciones.

Entre las ventajas que se obtienen usando estos métodos (multiplicadores de Lagrange conjuntamente con aproximación desvanescente), es que bajo hipótesis relativamente más débiles que las empleadas de manera estándar en la literatura conocida, probamos la existencia de políticas óptimas de los problemas con restricciones, así también se proporciona un método para computarlas. La desventaja estriba en que debemos tener hipótesis de regularidad en nuestras funciones de valor óptimo descontado como funciones del parámetro asociado al multiplicador de Lagrange, en otras palabras, debemos asumir que dichas funciones son diferenciables en tales parámetros, lo que no necesariamente ocurre de manera general.

Finalmente, consideramos que el trabajo tiene cierta relevancia para trabajos posteriores, que permitan extender el análisis para un número mayor de restricciones, así como tratar de extender al caso de cadenas de Markov controladas a tiempo continuo. También consideramos que se podría extender nuestro análisis a sistemas más generales donde no se cumplan necesariamente hipótesis tan restrictivas como lo es la W-ergodicidad geométrica.

Bibliografía

[1] Altman, E., *Constrained Markov Decision Processes*, Chapman & Hall/CRC, Boca Raton, FL., 1999.

[2] Billingsley, P., *Convergence of probability measures*, Wiley, New York, 1968.

[3] Borkar V.S., and Ghosh M. K., *Controlled diffusions with constraints*, J. Math. Anal. Appl. 152 (1990), 88-108.

[4] Borkar V.S., *Ergodic control of Markov chains with constraints- the general case*, SIAM J. Control Optim. 32 (1994), 176-186.

[5] Chang H.S., *A policy improvement method in constrained stochastic dynamic programming*, IEEE Trans. Automat. Control 51 (2006), No. 9, 1523-1526.

[6] Chen R.C. and Feinberg E.A., *Nonradomized policies for constrained Markov decision processes*, Maths. Methods Oper. Res. 66 (2007), No. 1, 165-179.

[7] Costa O.L.V., Dufour F., *Average control of Markov decision processes with Feller transition probabilities and general action spaces*, J. Math. Anal. Appl. 396 (2012) 5869.

[8] Dufour F. and Stockbridge R.H., *Existence of strict optimal controls for discounted stochastic control problems, in Modern Trends in Controlled Stochastic Processes: Theory and Applications*, edited by A.B. Piunovskiy, Luniver Press, Frome, U.K., 2010, 12-22.

[9] Dutta P.K., *What do discounted optima converge to? A theory of discount rate asymptotics in economic models*, J. Econom. Theory 55 (1991) 6494.

[10] Feinberg E. A., Kasyanov P.O., Zadoianchuk N. V., *Average cost Markov decision processes with weakly continuous transition probabilities*, Math. Oper. Res 37 (4) (2012) 591607.

[11] Feinberg E. A. and Shwartz A., *Constrained discounted dynamic programming*, Math. Oper. Res. 21 (1996), 922-945

[12] Gordienko E. and Hernández-Lerma O., *Average cost Markov control processes with weigthed norms: existence of canonical policies*, Appl. Math. (Warsaw) 23 (1995), 199-218.

[13] Guo X. P., Hernández-Lerma O., *Continuous-Time Markov Decision Processes*, Springer-Verlag, Berlin, Heidelberg, 2009.

[14] Guo X. P., Huang Y., and Song X., *Linear programming and constrained average optimality for general continuous-time Markov decision processes in history-dependent policies,* SIAM J. Control Optim. 50 (2012), 23-47.

[15] Guo X. P., Quanxin Z., *Average optimality for Markov decision processes in Borel spaces: a new condition and approach*, J. Appl. Probab 43 (2006) 318334

[16] Haviv M., *On constrained Markov decision processes*, Oper. Res. Lett. 19 (1996), 25-28.

[17] Hernández-Lerma O., González-Hernández J., *Constrained Markov control processes in Borel space: The discounted case,* Math. Methods Oper. Res. 52 (2000), 271-285.

[18] Hernández-Lerma O., González-Hernández J. and López-Martínez, R. R., *Constrained average cost Markov control processes in Borel spaces*, SIAM J. Control Optim. 42 (2003), 442-468.

[19] Hernández-Lerma O. and Laserre, J. B., *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, Springer-Verlag, New York, 1999.

[20] Hernández-Lerma O. and Lasserre J.B., *Further Topics on Discrete-time Markov Control Processes*, Springer-Verlag, New York, 1999.

[21] Hilgert N. and Hernández-Lerma O., *Bias optimality versus strong 0-discount optimality in Markov control processes with unbounded costs*, Acta Appl. Math. 77 (2003), 215-235.

[22] Jaso Fuentes, H., Escobedo-Trujillo B. A., y Mendoza-Pérez A.F., *The Lagrange and the vanishing discount techniques to controlled difussions with cost constrants.* Sometido.

[23] Lyer K. and Hamachandra N., *Sensitivity analysis and optimal ultimately stationary deterministic policies in some constrained discounted cost models*, Maths. Methods Oper. Res 71 (2010), No. 3, pp. 404-425.

[24] Mendoza-Pérez, A.F. and Hernández-Lerma, O., *Markov control processes with pathwise constraints*, Math. Methods Oper. Res. 71 (2010), 477-502.

[25] Mendoza-Pérez A. F. and Hernández-Lerma O., *Deterministic optimal policies for Markov control processes with pathwise constraints*, Applications Mathematicae 39.2 (2012), 185-209.

[26] Piunovskiy, A. B., *The problem of convex programming with linear constraints,* Comput. Math. Phys., 34, (1994).

[27] Piunovskiy, A. B., *Optimal Control of Radom Sequences in Problems with Constraints,* Kluwer, Boston, 1997.

[28] Prieto-Rumeau T.and Hernández-Lerma O., *Ergodic control of continuous-time Markov chains with pathwise constraints*, SIAM J. Control Optim. 47 (2008), 1888-1908.

[29] Prieto-Rumeau T., Hernández-Lerma O., *The vanishing discount approach to constrained continuous-time controlled Markov chains*, Systems Control Lett. 59 (2010) 504-509.

[30] Puterman M.L., *Markov Decision Process*, Wiley, New York, 1994.

[31] Vega-Amaya O. *The average cost optimality equation: a fixed point approach.* Bol Soc Mat Mexicana 9:185-195. (2003).

[32] Vega-Amaya O. *On the vanishing discount factor approach for Markov decision processes with weakly continuous transition probabilities.* J. Math. Anal. Appl. 426 (2015), 978-985.

[33] Zhang L. L., Guo X. P., *Cosntrained continuous-time Markov decision processes with average criteria,* Math. Methods Oper. Res., 67 (2008), 323-340.

Índice

W-ergodicidad geométrica, 28

costo esperado, 27
costo esperado promedio, 27
costo por trayectorias, 27
criterio α-descontado, 11

función de peso, 8
función de restricción, 16
función selectora, 5

ganancia esperada, 27
ganancia esperada promedio, 27
ganancia por trayectorias, 27

historias, 4

Irreducibilidad, 28

Lyapunov, condicion tipo, 11

modelo de control de markov, 4
multiplicadores de Lagrange, 19

norma del supremo, 8

política de control aleatorizada, 5
política determinista estacionaria, 5
política estacionaria, 5
política Markov aleatorizada, 5
problema descontado con restricciones, 18
problema descontado sin restricciones, 20
Problema esperado con restricciones, 28

Propiedad Markoviana, 7
Propiedad Markoviana Estacionaria, 8

Teorema de Ionescu-Tulcea, 6

W-norma, 8