



BUAP

Benemérita Universidad Autónoma de Puebla

Nº 319661

DIRECCIÓN DE ADMINISTRACIÓN ESCOLAR

Folio BUAP: 0191316

ACTA DE EXAMEN DE GRADO

En la Ciudad de Puebla, Puebla siendo las 12:00 horas del día once del
mes de julio de 2025, reunido en

FACULTAD DE CIENCIAS FISICO MATEMATICAS

Los integrantes del Honorable Jurado de Examen de Grado:

Presidente **DRA. HORTENSIA JOSEFINA REYES CERVANTES**

Secretario **DR. FERNANDO VELASCO LUNA**

Vocal **DR. VICTOR HUGO VAZQUEZ GUEVARA**

Vocal **DR. CARLOS CAMILO GARAY**

Proceden a examinar a: **PORTILLO RAMIREZ GUSTAVO**

Quienes al terminar el examen resolvieron: **APROBARLO POR CUM LAUDE**

para obtener: **EL GRADO DE DOCTOR EN CIENCIAS MATEMÁTICAS**

Acto continuo el presidente del Jurado le hizo saber el resultado de su examen y le tomó
la protesta de Ley correspondiente.

Se expide la presente Acta de Examen en la ciudad de Puebla, Puebla a
11 de agosto del 2025.

**Con las facultades conferidas por la Ley de la Benemérita Universidad Autónoma de
Puebla y el Estatuto Orgánico de la misma**

CERTIFICAN

La presente Acta de Examen de Grado

EL SECRETARIO GENERAL DE LA BUAP

MTRO. JOSÉ MANUEL ALONSO OROZCO

EL DIRECTOR DE ADMINISTRACIÓN ESCOLAR

MTRO. JUAN MANUEL ROSAS TAPIA

NOTAS IMPORTANTES AL REVERSO



Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias Físico Matemáticas

Posgrado en Ciencias Matemáticas

**Convergencia de sistemas estocásticos Markovianos
controlados**

Tesis

Presentada para obtener el grado de
Doctorado en Ciencias Matemáticas

Presenta
M.C. Gustavo Portillo Ramírez

Director de Tesis
Dr. Hugo Adán Cruz Suárez

Puebla, Puebla. Julio 2025



BUAP

DR. SEVERINO MUÑOZ AGUIRRE
SECRETARIO DE INVESTIGACIÓN Y
ESTUDIOS DE POSGRADO, FCFM-BUAP
P R E S E N T E:

Por este medio le informo que el C:

GUSTAVO PORTILLO RAMÍREZ

estudiante del Doctorado en Ciencias (Matemáticas), ha cumplido con las indicaciones que el Jurado le señaló en el Coloquio que se realizó el día 19 de junio de 2025, con la tesis titulada:

“Convergencia de sistemas estocásticos Markovianos controlados”

Por lo que se le autoriza a proceder con los trámites y realizar el examen de grado en la fecha que se le asigne.

A T E N T A M E N T E.
H. Puebla de Z, a 19 de junio de 2025

DR. RAÚL ESCOBEDO CONDE
COORDINADOR DEL POSGRADO
EN MATEMÁTICAS.



D*REC/mtrv

Dedicado a mi familia...

Agradecimientos

Al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT) por la beca otorgada durante el periodo de doctorado para la obtención del grado.

A toda mi familia que me ha apoyado y motivado durante todo este tiempo. A mis papás Teresa y Jacinto, a mis hermanas y hermanos, Amada, Bertha, Reyna, Edmundo y Josué, a mis sobrinas y a mi sobrino Jeansli, Mariana y Caleb. A mi prometida Sinai y a su familia.

A mi director de tesis el doctor Hugo Adán Cruz Suárez y a su familia por sus consejos, enseñanzas y por el tiempo dedicado a la producción de todo el trabajo desarrollado durante esta etapa.

A mis sinodales la doctora Hortensia Josefina Reyes Cervantes, a la doctora Rosa María Flores Hernández, al doctor Carlos Camilo Garay, al doctor Fernando Velasco Luna y al doctor Víctor Hugo Vázquez Guevara, por el tiempo dedicado a la revisión de este escrito y por las sugerencias para mejorarlo.

A todos mis amigos que preservo desde la licenciatura y los nuevos que se agregaron, por su compañía y sus pláticas.

Muchas gracias a todos.

Introducción

Los modelos de decisión secuencial son abstracciones matemáticas de situaciones en las que se toman decisiones en varias etapas. En cada una de estas etapas se incurre en un cierto costo. Cada decisión puede influir en las circunstancias bajo las que se tomarán decisiones futuras, de modo que, si se quiere minimizar un costo total, se debe equilibrar el deseo de minimizar el costo de la decisión presente con el deseo de evitar situaciones futuras donde el alto costo es inevitable. Entre la variedad de problemas de decisión secuenciales se tienen Problemas de Control Óptimo Determinista (PCOD) y Problemas de Control Óptimo Estocástico (PCOE), problemas de decisión de Markov y semi-Markoviano, problemas de control minimax y juegos secuenciales [7].

Como uno de los principales métodos para el análisis de problemas de decisiones secuenciales se tiene la Programación Dinámica (PD) [6, 24]. Aunque la naturaleza de estos problemas varía de manera considerable, sus estructuras subyacentes resultan ser muy similares. En todos los casos, el costo correspondiente a una política y la iteración básica del algoritmo PD pueden describirse mediante un determinado mapeo que difiere de un problema a otro en detalles que en gran medida no son esenciales. Normalmente, tal mapeo resume los datos del problema y determina las cantidades de interés para el analista. Si se toma este mapeo como punto de partida se pueden proporcionar resultados analíticos poderosos que son aplicables a una gran colección de problemas de decisión secuencial.

El trabajo está enfocado a Procesos de Decisión de Markov (PDMs) a tiempo discreto [7, 37, 38, 51]. En los tres capítulos que

componen este escrito, se proponen algunos resultados que involucren convergencia en PDMs. En el Capítulo 1, la atención se centra en PDMs donde la ley de transición es inducida por un sistema de dos ecuaciones en diferencias acopladas que son perturbadas por un ruido pequeño que depende de parámetros numéricos ϵ y δ , ver Sección 1.1. La primera ecuación representa la evolución de estados x , mientras que la segunda representa la evolución de estados α , que se relacionan directamente con parámetros del modelo. En este contexto, estamos interesados en los siguientes problemas:

- Estudiar aproximaciones de PDMs por procesos de control deterministas. En particular, estamos interesados en asegurar que la política de un sistema determinista es asintóticamente óptima para el sistema aleatorio, ver Teorema 1.3.7.
- Analizar la convergencia de la función de valor óptimo y la política óptima del sistema estocástico cuando $\epsilon \rightarrow 0$ y $\delta \rightarrow 0$, ver Teorema 1.3.10.

La metodología para resolver estos problemas es imponer restricciones de continuidad de Lipschitz [39, 46] sobre las componentes del modelo de control y aplicar técnicas de PD. Nuestro enfoque asegura los siguientes tres aspectos importantes:

- La existencia de una cota superior para el índice de estabilidad [29–31] cuando aplicamos la política óptima del sistema determinista. En consecuencia, resulta que la política óptima del sistema determinista es asintóticamente óptima para el sistema estocástico.
- Una tasa de convergencia de la función de costo óptimo para el sistema aleatorio con respecto al sistema determinista.
- La convergencia uniforme de las políticas estocásticas óptimas para la política determinista cuando $\epsilon \rightarrow 0$ y $\delta \rightarrow 0$, sobre subconjuntos compactos del espacio de estados.

En el Capítulo 2, se aborda un problema de crecimiento económico que se ha estudiado ampliamente en [58]. Tal problema se conforma de producción, acumulación de capital y crecimiento estocástico de la productividad. La variante presentada en este trabajo es a través la perturbación por un ruido pequeño de la tasa

con la que evoluciona la depreciación del capital. Esta idea, permite identificar el problema perturbado con un Proceso de Decisión de Markov, que se desarrolla por dos ecuaciones en diferencias acopladas (Capítulo 1), cada una perturbada por un ruido pequeño. En este contexto, se presenta un teorema del límite funcional para una transformación de los logaritmos del capital y de los logaritmos de la tasa con la que evoluciona la depreciación del capital, tanto del sistema estocástico como el sistema determinista. Esta conclusión se establece en el Teorema 2.3.1.

Adicionalmente, presentamos experimentos numéricos desarrollados con técnicas de Aprendizaje por Refuerzo, específicamente utilizando Q -learning, para presentar la solución del problema de crecimiento económico perturbado y para mostrar la convergencia al estado estable determinista. Finalmente, se muestra evidencia de la normalidad dada en el teorema principal a través de histogramas y de pruebas estadísticas. Esto se logró gracias a la implementación de los Algoritmos 1-3, que fueron implementados en el software R [52].

En el Capítulo 3, se aborda un tema independiente a los que se presentaron en los capítulos 1 y 2. De manera precisa se trabaja con cadenas de decisión de Markov sobre espacio de estados numerable, donde la función de costo en un paso es acotada y el rendimiento se mide por el criterio promedio asociado con un controlador de decisiones propenso al riesgo (risk-seeking). Las condiciones estructurales sobre la ley de transición aseguran que el costo promedio es constante, pero no garantizan que la ecuación de optimalidad admite solución. En este contexto, se aborda el siguiente problema: Obtener aproximaciones convergentes al costo promedio óptimo, y para determinar aproximadamente las políticas estacionarias óptimas usando los puntos fijos de una familia de operadores contractivos. Las conclusiones principales sobre este problema se establecen en el Teorema 3.2.1 y representan una extensión del clásico “enfoque descontado” en el caso neutral al riesgo [1, 35]. En este contexto, se extienden los resultados establecidos en [53], donde una versión del caso averso al riesgo (risk-averse) de este problema fue analizado.

El estudio de cadenas de decisión de Markov dotado con un

criterio sensible al riesgo se remonta, al menos a los artículos proporcionados por Howard y Matheson [40], donde fueron analizadas las cadenas de decisión de Markov con espacio de estados finito y el costo promedio óptimo fue caracterizado vía una ecuación de optimalidad. El interés sobre este tema ha sido motivado por las aplicaciones, por ejemplo, en finanzas [4, 5, 47, 57], gestión de ingresos [2], y teoría de las grandes desviaciones [8]. Modelos con espacio de estados finito o infinito son considerados, por ejemplo, en [11, 12], [55, 56] mientras que cadenas de decisión de Markov sobre espacios de estado de Borel fueron analizados en [20–22], [41, 42] y [54]. Juegos estocásticos con criterio sensible al riesgo son estudiados en [3].

Índice general

Agradecimientos	III
Introducción	VIII
1. Aproximación determinista de PDMs con dinámica acoplada	1
1.1. Modelo de control de Markov	4
1.2. Planteamiento del problema	8
1.3. Suposiciones y resultados	8
1.4. Ejemplos	19
1.4.1. Problema de consumo-inversión	19
1.4.2. Problema de control con ruido aditivo pequeño	22
1.4.3. Importancia de las suposiciones	25
2. Comportamiento normal en un modelo de crecimiento económico	29
2.1. Q -learning	31
2.1.1. Q -factor	32
2.1.2. Q -factores y Robbins-Monro	33
2.2. El modelo de crecimiento identificado con un PDM	36
2.3. Teorema del límite central	38
2.3.1. Comparación de los modelos estocásticos y deterministas	40
2.4. Experimentos numéricos	43
2.4.1. Algoritmos	48
3. Cadenas de decisión de Markov promedio: caso propenso al riesgo	51
3.1. Modelo de decisión	52

ÍNDICE GENERAL

3.2. Aproximaciones contractivas	56
3.3. Herramientas auxiliares	59
3.4. Demostración del resultado principal	65
Bibliografía	68

Capítulo 1

Aproximación determinista de PDMs con dinámica acoplada

En este capítulo se abordan los llamados Procesos de Decisión de Markov discretos con horizonte infinito y costo total descontado [7, 24, 37, 38, 51]. La importancia de trabajar con PDMs radica en el amplio rango de aplicación en distintas disciplinas, por ejemplo: ingeniería, informática, comunicaciones y economía [9, 36]. El problema principal en PDMs es determinar una política óptima y la función de valor óptimo. Para caracterizar y determinar las soluciones de los PDMs se tiene disponible el enfoque de Programación Dinámica [6, 24].

En este capítulo, los PDMs de interés son aquellos que evolucionan mediante una dinámica de estados compuesta por dos ecuaciones en diferencias acopladas como se muestra en (1.1) y (1.2). La ecuación (1.1) modela la transición de estados del sistema, donde el conjunto de todos los estados se denota por X , en este capítulo, tales elementos serán llamados x -estados. Similarmente, la ecuación (1.2) modela el cambio en los parámetros del sistema; el conjunto de todos los parámetros es denotado por Γ , estos estados serán llamados α -estados. Sean ϵ_0 y δ_0 números positivos y sean $\epsilon \in [0, \epsilon_0]$ y $\delta \in [0, \delta_0]$, entonces considere las perturbaciones $\{\xi_t(\epsilon)\}$ y $\{\eta_t(\delta)\}$, que son sucesiones de elementos aleatorios independientes e idénticamente distribuidos en algunos

CAPÍTULO 1. APROXIMACIÓN DETERMINISTA DE PDMS CON DINÁMICA ACOPLADA

espacios de Borel (S_1, r_1) y (S_2, r_2) (espacios métricos de perturbación [18] o espacios de ruido [30]), respectivamente.

Suponga que existen $s_1 \in S_1$ y $s_2 \in S_2$ tales que $s_1 = \xi_t(0)$ y $s_2 = \eta_t(0)$, cada elemento de las sucesiones anteriores depende de parámetros ϵ y δ tales que $\mathbb{E}r_1(\xi(\epsilon), s_1) \rightarrow 0$ cuando $\epsilon \rightarrow 0$ y $\mathbb{E}r_2(\eta(\delta), s_2) \rightarrow 0$ cuando $\delta \rightarrow 0$, donde ξ y η son elementos genéricos de $\{\xi_t(\epsilon)\}$ y $\{\eta_t(\delta)\}$, respectivamente. Bajo el contexto anterior, estamos interesados en los siguientes problemas:

- Estudiar aproximaciones de PDMs por procesos de control deterministas. En particular, estamos interesados en asegurar que la política de un sistema determinista es asintóticamente óptima para el sistema aleatorio, ver Teorema 1.3.7.
- Analizar la convergencia de la función de valor óptimo y la política óptima del sistema estocástico cuando $\epsilon \rightarrow 0$ y $\delta \rightarrow 0$, ver Teorema 1.3.10.

En lo que sigue, describimos brevemente el trabajo relacionado con los problemas discutidos en este capítulo.

El estudio en [45], consideró el problema de aproximación de un proceso de control estocástico por un proceso determinista en el caso continuo. En este artículo, los autores demostraron que el problema estocástico puede ser aproximado por un determinista cuando el ruido es pequeño y las fluctuaciones se vuelven rápidas. En este contexto, se demuestra que el control óptimo para el problema determinista es asintóticamente óptimo para problemas estocásticos. En el caso continuo, [23] aborda un problema similar, es decir, cuando los efectos del ruido en un sistema físico son pequeños, estos autores realizaron un análisis asintótico de aproximación de difusión y usaron esto para estimaciones deseadas del sistema original. Para PDMs a tiempo discreto, esta clase de problemas fueron estudiados por [17, 18], donde la dinámica del sistema está descrita por una sola ecuación en diferencias. La convergencia entre estos modelos también fue estudiada en [44]. Sin embargo, la convergencia fue estudiada usando sucesiones que pertenecen al conjunto de parejas de estado-acción admisibles, que se supone es un subconjunto de un espacio euclidiano. Además, este estudio se lleva a cabo bajo el supuesto de que el espacio de

acciones es un conjunto compacto y que la función de costos es acotada. Ahora, cuando consideramos PDMs que se desarrollan con respecto a (1.1) y (1.2), los resultados que aparecen en [17] se generalizan. El enfoque de ecuaciones acopladas puede aplicarse, por ejemplo, cuando se considera un factor de descuento aleatorio [25–28], donde la segunda ecuación en diferencias se refiere a la evolución del factor de descuento aleatorio.

La metodología para resolver los problemas descritos arriba es imponer restricciones de continuidad de Lipschitz [39, 46] sobre las componentes del modelo de control y aplicar técnicas de PD. Específicamente, suponemos condiciones Lipschitz para las funciones c , F y G involucradas en la dinámica del sistema compuesta de dos ecuaciones en diferencias acopladas (ver ecuaciones (1.1) y (1.2)). Una consecuencia de esta suposición es la continuidad de Lipschitz de la función de costo óptimo. Este enfoque asegura los siguientes tres aspectos importantes:

- La existencia de una cota superior para el índice de estabilidad [29–31] cuando aplicamos la política óptima del sistema determinista. En consecuencia, resulta que la política óptima del sistema determinista es asintóticamente óptima para el sistema estocástico.
- Una tasa de convergencia de la función de costo óptimo para el sistema aleatorio con respecto al sistema determinista.
- La convergencia uniforme de las políticas estocásticas óptimas para la política determinista cuando $\epsilon \rightarrow 0$ y $\delta \rightarrow 0$, sobre subconjuntos compactos del espacio de estados.

Este capítulo está organizado como sigue. En la Sección 1.1, se presenta la teoría básica de los PDMs con estados que evolucionan con la dinámica compuesta por dos ecuaciones en diferencias acopladas. En la Sección 1.2, se establece el planteamiento del problema de aproximación de la función de valor y la política óptima. En la Sección 1.3, se presenta el resultado que proporciona la cota para el índice de estabilidad $\hat{\delta}_{\epsilon, \delta}$, la tasa de convergencia del costo óptimo y la convergencia de la política óptima sobre subconjuntos compactos. Finalmente, en la Sección 1.4, se ilustra la teoría desarrollada con dos ejemplos. El primero es referente al problema

CAPÍTULO 1. APROXIMACIÓN DETERMINISTA DE PDMS CON DINÁMICA ACOPLADA

de consumo-inversión. Este ejemplo tiene como antecedentes los ejemplos desarrollados en [19, 25]. El segundo ejemplo es un problema de control con ruido aditivo pequeño. En ambos problemas se proporcionan explícitamente la cota para el índice de estabilidad y la tasa de convergencia de la función de valor óptimo.

1.1. Modelo de control de Markov

Considere el modelo de Markov que sigue:

$$\mathcal{M} := (X \times \Gamma, A, \{A(x, \alpha) \mid (x, \alpha) \in X \times \Gamma\}, Q, c),$$

donde $X \times \Gamma$ y A son espacios de Borel, denominados espacio de estados y espacio de acciones, respectivamente; $\{A(x, \alpha) \mid (x, \alpha) \in X \times \Gamma\}$ es una familia de subconjuntos medibles no vacíos $A(x, \alpha)$ de A , donde $A(x, \alpha)$ denota el conjunto de acciones admisibles cuando el sistema se encuentra en el estado $(x, \alpha) \in X \times \Gamma$. El conjunto de estado-acción admisibles está definido como sigue:

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

el cual es un subconjunto medible de $X \times \Gamma \times A$; la siguiente componente es un kernel estocástico Q sobre $X \times \Gamma$ dado \mathbb{K} , i.e., $Q(\cdot \mid x, \alpha, a)$ es una medida de probabilidad sobre $X \times \Gamma$ para cada $(x, \alpha, a) \in \mathbb{K}$ y $Q(B \mid \cdot)$ es una función medible sobre \mathbb{K} para cada $B \in \mathcal{B}(X \times \Gamma)$, donde $\mathcal{B}(X \times \Gamma)$ denota la σ -álgebra de Borel de $X \times \Gamma$; c una función medible sobre \mathbb{K} que toma valores reales, llamada función de costo por etapa.

Observación 1.1.1. *En el desarrollo posterior, las métricas de los espacios métricos X , Γ , y A serán denotadas por d_x , d_α y d_2 , respectivamente. En consecuencia, sobre $X \times \Gamma$ se define la métrica d_1 como*

$$d_1((x, \alpha), (x', \alpha')) = \max\{d_x(x, x'), d_\alpha(\alpha, \alpha')\},$$

para todo $(x, \alpha), (x', \alpha') \in X \times \Gamma$. Además, sobre \mathbb{K} es considerada la métrica d definida por

$$d((x, \alpha, a), (x', \alpha', a')) = \max\{d_1((x, \alpha), (x', \alpha')), d_2(a, a')\},$$

para todo $(x, \alpha, a), (x', \alpha', a') \in \mathbb{K}$.

La dinámica del sistema se describe a continuación. Suponga que al tiempo t , $t = 0, 1, \dots$, el sistema se encuentra en el estado $(x_t, \alpha_t) = (x, \alpha) \in X \times \Gamma$. Entonces, el controlador elige un control $a_t = a \in A(x, \alpha)$. Como consecuencia de esto, ocurren dos cosas:

- a) Se incurre en un costo $c(x_t, \alpha_t, a_t)$, y
- b) el sistema se mueve a un nuevo estado $(x_{t+1}, \alpha_{t+1}) = (x', \alpha')$ de acuerdo a la ley de transición $Q(\cdot|x, \alpha, a)$, i.e.,

$$Q(B|x, \alpha, a) = \Pr((x_{t+1}, \alpha_{t+1}) \in B | x_t = x, \alpha_t = \alpha, a_t = a),$$

$$B \in \mathcal{B}(X \times \Gamma) \text{ y } (x, \alpha, a) \in \mathbb{K}.$$

Posteriormente, el sistema pasa a un estado (x_{t+1}, α_{t+1}) y se repite el proceso.

En este capítulo, se supone que la ley de transición Q es inducida por un sistema de ecuaciones en diferencias acopladas, como sigue

$$x_{t+1} = F(x_t, \alpha_t, a_t, \xi_t(\epsilon)), \quad (1.1)$$

$$\alpha_{t+1} = G(\alpha_t, \eta_t(\delta)), \quad (1.2)$$

donde $t = 0, 1, \dots$, con $(x_0, \alpha_0) \in X \times \Gamma$ dado, donde $F : \mathbb{K} \times A \times S_1 \rightarrow X$ y $G : \Gamma \times S_2 \rightarrow \Gamma$ son funciones medibles. Sean ϵ_0, δ_0 números fijos positivos y sean $\epsilon \in [0, \epsilon_0]$, $\delta \in [0, \delta_0]$, las perturbaciones $\{\xi_t(\epsilon)\}$ y $\{\eta_t(\delta)\}$ son sucesiones de elementos aleatorios independientes e idénticamente distribuidos (i.i.d.) con valores en algunos espacios de Borel (S_1, r_1) y (S_2, r_2) , respectivamente.

Observación 1.1.2. *Se asume que las variables aleatorias $\xi : \Omega_1 \rightarrow S_1$ y $\eta : \Omega_2 \rightarrow S_2$ están definidas sobre espacios de probabilidad $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ y $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$, donde ξ y η son elementos genéricos de $\{\xi_t(\epsilon)\}$ y $\{\eta_t(\delta)\}$, respectivamente. Además, la terna $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mathbb{P})$ denota el espacio de probabilidad producto, donde $\mathcal{F}_1 \otimes \mathcal{F}_2$ es la σ -álgebra producto y \mathbb{P} es la medida producto inducida por el teorema de Ionescu-Tulcea [37]. El valor esperado con respecto a la medida de probabilidad \mathbb{P} será denotado por \mathbb{E} .*

CAPÍTULO 1. APROXIMACIÓN DETERMINISTA DE PDMS CON DINÁMICA ACOPLADA

En el desarrollo posterior, será considerado el espacio $S := S_1 \times S_2$ con la métrica r definida por

$$\begin{aligned} & r\left((\xi(\omega_1), \eta(\omega_2)), (\hat{\xi}(\omega'_1), \hat{\eta}(\omega'_2))\right) \\ &= \max\left\{r_1(\xi(\omega_1), \hat{\xi}(\omega'_1)), r_2(\eta(\omega_2), \hat{\eta}(\omega'_2))\right\}, \end{aligned}$$

para todo $(\xi(\omega_1), \eta(\omega_2)), (\hat{\xi}(\omega'_1), \hat{\eta}(\omega'_2)) \in S$, donde $\omega_1, \omega'_1 \in \Omega_1$ y $\omega_2, \omega'_2 \in \Omega_2$.

Ahora, considere $\chi(\epsilon, \delta) = (\xi(\epsilon), \eta(\delta))$, entonces las ecuaciones en diferencias (1.1) y (1.2) pueden ser expresadas como

$$\begin{aligned} (x_{t+1}, \alpha_{t+1}) &= H(x_t, \alpha_t, a_t, \chi_t(\epsilon, \delta)) \\ &:= (F(x_t, \alpha_t, a_t, \xi_t(\epsilon)), G(\alpha_t, \eta_t(\delta))). \end{aligned}$$

Supongamos que existen $s_1 \in S_1$ y $s_2 \in S_2$ tales que $s_1 = \xi(0)$ y $s_2 = \eta(0)$. Cada elemento de las sucesiones anteriores depende de un parámetro numérico ϵ y δ , de tal modo que $\mathbb{E}r_1(\xi(\epsilon), s_1) \rightarrow 0$ cuando $\epsilon \rightarrow 0$ y $\mathbb{E}r_2(\eta(\delta), s_2) \rightarrow 0$ cuando $\delta \rightarrow 0$. Por otro lado, se considera un PDM determinista cuya dinámica evoluciona de acuerdo con las ecuaciones en diferencias que se muestran en (1.3) y (1.4):

$$x_{t+1} = F(x_t, \alpha_t, a_t, s_1), \tag{1.3}$$

$$\alpha_{t+1} = G(\alpha_t, s_2), \tag{1.4}$$

para todo $t = 0, 1, \dots$. Observe que $\chi(0, 0) = (\xi(0), \eta(0)) = (s_1, s_2)$, entonces la dinámica conjunta dada por (1.3) y (1.4) se denota de la siguiente manera

$$\begin{aligned} (x_{t+1}, \alpha_{t+1}) &= H(x_t, \alpha_t, a_t, \chi_t(0, 0)) \\ &:= (F(x_t, \alpha_t, a_t, \xi_t(0)), G(\alpha_t, \eta_t(0))). \end{aligned}$$

En este contexto, estamos interesados en la aproximación de Procesos de Decisión de Markov que evolucionan a través de (1.1) y (1.2) mediante el proceso de control determinista dado por las ecuaciones (1.3) y (1.4).

Cuando los procesos x -estados y α -estados estén especificados por el modelo dinámico dado por las ecuaciones (1.1) y (1.2), la ley de transición toma la forma

$$\begin{aligned} Q(B|x, \alpha, a) &:= Pr[(x_{t+1}, \alpha_{t+1}) \in B | x_t = x, \alpha_t = \alpha, a_t = a] \\ &= \int_{S_1 \times S_2} 1_B(H(x, \alpha, a, s_{\epsilon, \delta})) \mu(ds) \\ &= \mu(\{s \in S_1 \times S_2 : H(x, \alpha, a, s_{\epsilon, \delta}) \in B\}), \end{aligned} \quad (1.5)$$

donde $B \in \mathcal{B}(X \times \Gamma)$, $1_B(\cdot)$ denota la función indicadora sobre B y μ es la distribución común del vector aleatorio $\chi_t(\epsilon, \delta)$.

Por otro lado, cuando los procesos x -estados y α -estados están especificados por el modelo dinámico de las ecuaciones (1.3) y (1.4), la ley de transición toma la forma

$$Q_H(B|x, \alpha, a) := 1_B\left(H(x, \alpha, a, \chi(0, 0))\right), \quad (1.6)$$

donde $B \in \mathcal{B}(X \times \Gamma)$, $(x, \alpha, a) \in \mathbb{K}$. Así, el modelo de control de Markov está dado por $(X \times \Gamma, A, \{A(x, \alpha) : (x, \alpha) \in X \times \Gamma\}, Q_H, c)$.

Una política de control π es una secuencia $\{\pi_t : t = 0, 1, \dots\}$, donde para cada $t = 0, 1, \dots$, $\pi_t(\cdot | h_t)$ es una probabilidad condicional sobre la σ -álgebra de Borel $\mathcal{B}(A)$, dada la historia $h_t := (x_0, \alpha_0, a_0, \dots, x_{t-1}, \alpha_{t-1}, a_{t-1}, x_t, \alpha_t)$, tal que $\pi_t(A(x_t, \alpha_t) | h_t) = 1$. El conjunto de todas las políticas es denotado por Π .

Sea $\mathbb{F} := \{\phi : X \times \Gamma \rightarrow A \mid \phi \text{ es medible y } \phi(x, \alpha) \in A(x, \alpha), (x, \alpha) \in X \times \Gamma\}$. Una sucesión $\pi = \{\phi_t \mid t = 0, 1, \dots\}$ de funciones $\phi_t \in \mathbb{F}$ es llamada política de Markov. Una política de Markov $\pi = \{\phi_t \mid t = 0, 1, \dots\}$ es llamada política estacionaria si $\phi_t = \phi \in \mathbb{F}$, para todo $t = 0, 1, \dots$.

Dados los estados iniciales $(x_0 = x, \alpha_0 = \alpha) \in X \times \Gamma$ y cualquier política $\pi \in \Pi$, existe una medida de probabilidad $P_{(x, \alpha)}^\pi$ inducida por la tripleta (x, α, π) sobre el espacio $\Omega = (X \times \Gamma \times A)^\infty$, con \mathcal{F} la σ -álgebra producto. La existencia de esta medida de probabilidad se verifica de manera análoga a la realizada en [28]. El correspondiente operador esperanza será denotado por $\mathbb{E}_{(x, \alpha)}^\pi$. La tripleta (x, α, π) determina un proceso estocástico $(\Omega, \mathcal{F}, P_{(x, \alpha)}^\pi, \{(x_t, \alpha_t)\})$ llamado Proceso de Decisión de Markov. En lo sucesivo se denota $y = (x, \alpha)$ y $Y = X \times \Gamma$.

CAPÍTULO 1. APROXIMACIÓN DETERMINISTA DE PDMS CON DINÁMICA ACOPLADA

1.2. Planteamiento del problema

Considere un modelo de control de Markov determinista $(Y, A, \{A(y) : y \in Y\}, Q_H, c)$ como se presentó en la Sección 1.1. Además, considere un sistema de control estocástico con el mismo espacio de estados Y , espacio de controles A , conjuntos admisibles $A(y)$, $y \in Y$ y función de costo c , pero con el sistema dinámico descrito por

$$y_{t+1} = H(y_t, a_t, \chi_t(\epsilon, \delta)), \quad t = 0, 1, \dots$$

Observe que cuando el sistema es manejado por una política determinista, en la ley de transición estocástica (1.5), el sistema estocástico se convierte en un sistema determinista, con ley de transición (1.6), cuando $\epsilon \rightarrow 0$ y $\delta \rightarrow 0$.

Para cada política $\pi \in \Pi$ y estado inicial $(x, \alpha) \in Y$, considere el costo total esperado descontado, denotado por $\hat{V}_{\epsilon, \delta}(\pi, x, \alpha)$, y definido como

$$\hat{V}_{\epsilon, \delta}(\pi, x, \alpha) = \mathbb{E}_{(x, \alpha)}^{\pi} \left[\sum_{t=0}^{\infty} \beta^t c(x_t, \alpha_t, a_t) \right],$$

donde $\beta \in (0, 1)$ es el factor de descuento.

Por lo tanto, el problema de control óptimo es encontrar una política $\pi^* \in \Pi$ tal que

$$\hat{V}_{\epsilon, \delta}(\pi^*, x, \alpha) = \inf_{\pi \in \Pi} \{ \hat{V}_{\epsilon, \delta}(\pi, x, \alpha) \} := V_{\epsilon, \delta}(x, \alpha),$$

$(x, \alpha) \in X \times \Gamma$. A π^* se le conoce como política óptima, mientras que $V_{\epsilon, \delta}(x, \alpha)$, $(x, \alpha) \in Y$ es llamada función de valor óptimo. En el caso determinista cuando $\epsilon = 0$ y $\delta = 0$, se denotará a $V_{\epsilon, \delta}$ por V .

En el resto del capítulo, se establecen condiciones para realizar un análisis asintótico de la solución óptima para el sistema estocástico.

1.3. Suposiciones y resultados

En esta sección, introducimos tres bloques de condiciones para estudiar la convergencia del sistema estocástico definido por las ecuaciones (1.1) y (1.2). Además, se proporciona una cota para el índice de estabilidad que depende de un parámetro de perturbación de ruido pequeño $\hat{\delta}_{\epsilon, \delta}$. En lo que sigue $\chi(\epsilon, \delta)$ denota un elemento genérico de $\{\chi_t(\epsilon, \delta)\}$.

Suposición 1.3.1. a) El conjunto $A(x, \alpha)$ es compacto para cada $(x, \alpha) \in Y$ y el mapeo de valores-conjunto $(x, \alpha) \rightarrow A(x, \alpha)$ es semicontinuo superior con respecto a la métrica de Hausdorff.

b) La función de costo $c(y, \cdot)$ es semicontinua inferior sobre $A(y)$ para cada $y \in Y$.

c) Para cada función continua acotada $U : Y \rightarrow \mathbb{R}$ la función

$$U'(y, a) := \mathbb{E}U [H(y, a, \chi(\epsilon, \delta))],$$

$(y, a) \in \mathbb{K}$, es una función continua sobre \mathbb{K} y \mathbb{E} es introducida en la Observación 1.1.2.

La Suposición 1.3.1 es necesaria para asegurar la existencia de minimizadores en la correspondiente ecuación de optimalidad. La Suposición 1.3.1 a) es similar a la Suposición 1 presentada en [30].

Sea $Z : X \times \Gamma \rightarrow [1, \infty)$ una función medible. Si U es una función con valores reales sobre $X \times \Gamma$, entonces su norma ponderada está definida como

$$\|U\|_Z := \sup_{(x, \alpha) \in X \times \Gamma} \frac{|U(x, \alpha)|}{Z(x, \alpha)},$$

donde Z denota la función de peso. Sea \mathbb{B}_Z el espacio de Banach de funciones medibles $U : Y \rightarrow \mathbb{R}$ tales que $\|U\|_Z < \infty$.

Suposición 1.3.2. Existen constantes γ tal que $\gamma \in (\beta, 1)$ y una función de peso W sobre Y tal que para cualesquiera $\epsilon \in [0, \epsilon_0]$, $\delta \in [0, \delta_0]$ se tiene que

a) $|c(y, a)| \leq W(y)$, $(y, a) \in \mathbb{K}$.

b) $\mathbb{E}W [H(y, a, \chi(\epsilon, \delta))] \leq \frac{\gamma}{\beta} W(y)$, $(y, a) \in \mathbb{K}$.

c) Para cada estado $y \in Y$, la función

$$W'(y, a) := \mathbb{E}W [H(y, a, \chi(\epsilon, \delta))],$$

es continua en $a \in A(y)$.

La Suposición 1.3.2 se usa para garantizar la existencia de soluciones a la ecuación de optimalidad [30]. Además, bajo las Suposiciones 1.3.1 y 1.3.2, el enfoque de PD es válido. Por lo tanto, para cada $(x, \alpha) \in X \times \Gamma$, la siguiente relación es válida:

$$V_{\epsilon, \delta}(x, \alpha) = \inf_{a \in A(x, \alpha)} \left\{ c(x, \alpha, a) + \beta \int_{S_1 \times S_2} V_{\epsilon, \delta}(y) Q(y|x, \alpha, a) \right\}.$$

CAPÍTULO 1. APROXIMACIÓN DETERMINISTA DE PDMS CON DINÁMICA ACOPLADA

Un procedimiento para aproximar la función de valor es por medio de las funciones de iteración valor que se definen como:

$$V_{\epsilon, \delta}^n(x, \alpha) = \inf_{a \in A(x, \alpha)} \left\{ c(x, \alpha, a) + \beta \int_{S_1 \times S_2} V_{\epsilon, \delta}^{n-1}(y) Q(y|x, \alpha, a) \right\},$$

donde $(x, \alpha) \in X \times \Gamma$ y $n = 1, 2, \dots$, con $V_{\epsilon, \delta}^0(\cdot) = 0$.

Suposición 1.3.3. *Existen constantes $L_0, L_1, L_{2,x}$ y $L_{2,\alpha}$ tales que*

- a) $|c(y, a) - c(y', a)| \leq L_0 d_1(y, y')$, para cada $(y, a), (y', a) \in \mathbb{K}$.
- b) $d_1\left(H(y, a, (s_1, s_2)), H(y', a, (s_1, s_2))\right) \leq L_1 d_1(y, y')$, para $(y, a), (y', a) \in \mathbb{K}, (s_1, s_2) \in S_1 \times S_2$ con $L_1 \leq 1$.
- c) Las funciones F y G satisfacen:
 - (i) $d_x(F(x, \alpha, a, s_1), F(x, \alpha, a, s'_1)) \leq L_{2,x} r_1(s_1, s'_1)$,
para todo $(x, \alpha, a) \in \mathbb{K}$ y para todo $s_1, s'_1 \in S_1$.
 - (ii) $d_\alpha(G(\alpha, s_2), G(\alpha, s'_2)) \leq L_{2,\alpha} r_2(s_2, s'_2)$,
para cualquier $\alpha \in \Gamma$ y $s_2, s'_2 \in S_2$.

Observación 1.3.4. *Bajo la Suposición 1.3.3, la función de costo y la función H involucrada en la dinámica de los estados son funciones Lipschitz respecto a la variable $y \in Y$. Además, las funciones F y G son funciones Lipschitz respecto a ξ y η , respectivamente.*

Si se cumplen las Suposiciones 1.3.1 y 1.3.2 con argumentos similares a los expuestos en [38] (considerando los respectivos cambios), se garantiza la existencia de una política óptima $\pi_{\epsilon, \delta} = \{f_{\epsilon, \delta}, f_{\epsilon, \delta}, \dots\}$ donde $f_{\epsilon, \delta} : Y \rightarrow A$, $a_{t+1} = f_{\epsilon, \delta}(x_t, \alpha_t)$, $f_{\epsilon, \delta} \in \mathbb{F}$, $t = 0, 1, \dots$ con correspondiente función de valor $\hat{V}_{\epsilon, \delta}(x, \alpha, \pi_{\epsilon, \delta}) = V_{\epsilon, \delta}(x, \alpha) \in \mathbb{B}_W$ y la esperanza $\mathbb{E}V_{\epsilon, \delta}[H(y, a, \chi(\epsilon, \delta))]$ existe para cada $\epsilon \in [0, \epsilon_0]$ y $\delta \in [0, \delta_0]$, $(y, a) \in \mathbb{K}$. Además, la política óptima para el problema de control determinista es denotado por $\pi_0^* = \{f^*, f^*, \dots\}$ con $f^* \in \mathbb{F}$.

Sea \mathbf{L} la distancia de Kantorovich definida en (S, \mathbb{B}_s) :

$$\mathbf{L}(\chi, \chi') = \sup\{|\mathbb{E}\varphi(\chi) - \mathbb{E}\varphi(\chi')| \mid \varphi \text{ tal que } |\varphi(s) - \varphi(s')| \leq r(s, s'), s, s' \in S\}. \quad (1.7)$$

Por otra parte, el índice de estabilidad $\Delta_{\epsilon, \delta}$ se define como

$$\Delta_{\epsilon, \delta}(y, \pi) := \hat{V}_{\epsilon, \delta}(y, \pi) - V_{\epsilon, \delta}(y), \quad y \in Y, \quad \pi \in \Pi.$$

El índice $\Delta_{\epsilon, \delta}(y, \pi)$ expresa el exceso del costo descontado cuando se aplica la política π para el proceso de control estocástico relacionado con

1.3. SUPOSICIONES Y RESULTADOS

(1.1) y (1.2) para $\epsilon > 0$ y $\delta > 0$, $y \in Y$. La calidad de la aproximación para el sistema estocástico por la política π_0^* será medida por el índice de estabilidad $\Delta_{\epsilon,\delta}(y, \pi_0^*)$ (ver [17, 30]), i.e.,

$$\Delta_{\epsilon,\delta}(y, \pi_0^*) := \hat{V}_{\epsilon,\delta}(y, \pi_0^*) - V_{\epsilon,\delta}(y), \quad y \in Y.$$

Por otra parte, defina un parámetro perturbación de ruido pequeño $\hat{\delta}_{\epsilon,\delta}$ como sigue:

$$\hat{\delta}_{\epsilon,\delta} := \mathbb{E} \max \left\{ r_1(\xi(\epsilon), \xi(0)), r_2(\eta(\delta), \eta(0)) \right\},$$

para $\epsilon \in [0, \epsilon_0]$ y $\delta \in [0, \delta_0]$. El Teorema 1.3.7 proporciona una cota superior para $\Delta_{\epsilon,\delta}(y, \pi_0^*)$ que involucra el parámetro de ruido $\hat{\delta}_{\epsilon,\delta}$. Los siguientes lemas serán utilizados para demostrar los Teoremas 1.3.7 y 1.3.10.

Lema 1.3.5. *Bajo las Suposiciones 1.3.1, 1.3.2 y 1.3.3 a) y b), para cada $\epsilon \in [0, \epsilon_0]$ y $\delta \in [0, \delta_0]$ fijos se tiene que $V_{\epsilon,\delta}^n$ es una función Lipschitz, para todo $n = 1, 2, \dots$. En consecuencia, $V_{\epsilon,\delta}$ es una función Lipschitz con constante Lipschitz $\frac{L_0}{1-\beta L_1}$.*

Demostración. Sean $\epsilon \in [0, \epsilon_0]$ and $\delta \in [0, \delta_0]$. La demostración es por inducción sobre $n \geq 1$. Para $n = 1$, observe que si $(x, \alpha), (x', \alpha') \in Y$ se tiene que $(x, \alpha), (x', \alpha') \in Y$

$$\begin{aligned} |V_{\epsilon,\delta}^1(x, \alpha) - V_{\epsilon,\delta}^1(x', \alpha')| &= \left| \inf_{a \in A} \{c(x, \alpha, a)\} - \inf_{a \in A} \{c(x', \alpha', a)\} \right| \\ &\leq \sup_{a \in A} |c(x, \alpha, a) - c(x', \alpha', a)| \\ &\leq L_0 d_1((x, \alpha), (x', \alpha')). \end{aligned}$$

Para $n > 1$, suponga que $V_{\epsilon,\delta}^{n-1}$ es una función Lipschitz con constante

CAPÍTULO 1. APROXIMACIÓN DETERMINISTA DE PDMS CON DINÁMICA ACOPLADA

$L_0 \sum_{i=0}^{n-2} (\beta L_1)^i$. Entonces,

$$\begin{aligned}
& |V_{\epsilon, \delta}^n(x, \alpha) - V_{\epsilon, \delta}^n(x', \alpha')| \\
&= \left| \inf_{a \in A} \left\{ c(x, \alpha, a) + \beta \mathbb{E} V_{\epsilon, \delta}^{n-1} [H(x, \alpha, a, \chi(\epsilon, \delta))] \right\} \right. \\
&\quad \left. - \inf_{a \in A} \left\{ c(x', \alpha', a) + \beta \mathbb{E} V_{\epsilon, \delta}^{n-1} [H(x', \alpha', a, \chi(\epsilon, \delta))] \right\} \right| \\
&\leq \sup_{a \in A} \left\{ |c(x, \alpha, a) - c(x', \alpha', a)| \right. \\
&\quad \left. + \left| \beta \mathbb{E} V_{\epsilon, \delta}^{n-1} [H(x, \alpha, a, \chi(\epsilon, \delta))] - \beta \mathbb{E} V_{\epsilon, \delta}^{n-1} [H(x', \alpha', a, \chi(\epsilon, \delta))] \right| \right\} \\
&\leq L_0 d_1((x, \alpha), (x', \alpha')) \\
&\quad + \beta \sup_{a \in A} \mathbb{E} L_0 \sum_{i=0}^{n-2} (\beta L_1)^i d_1(H(x, \alpha, a, \chi(\epsilon, \delta)), H(x', \alpha', a, \chi(\epsilon, \delta))) \\
&\leq L_0 d_1((x, \alpha), (x', \alpha')) + \beta L_0 \sum_{i=0}^{n-2} (\beta L_1)^i L_1 d_1((x, \alpha), (x', \alpha')) \\
&= \left(L_0 + L_0 \sum_{i=1}^{n-1} (\beta L_1)^i \right) d_1((x, \alpha), (x', \alpha')) \\
&= L_0 \sum_{i=0}^{n-1} (\beta L_1)^i d_1((x, \alpha), (x', \alpha')).
\end{aligned}$$

Por lo tanto, $V_{\epsilon, \delta}^n$ es una función Lipschitz con constante de Lipschitz $L_0 \sum_{i=0}^{n-1} (\beta L_1)^i$, para $n \in \mathbb{N}$.

Para verificar la segunda parte, observe que $\beta L_1 < 1$, entonces se cumple que $\sum_{i=0}^{\infty} (\beta L_1)^i = \frac{1}{1-\beta L_1}$. Además, como $V_{\epsilon, \delta}^n \rightarrow V_{\epsilon, \delta}$, cuando $n \rightarrow \infty$, se consigue que $V_{\epsilon, \delta}$ es una función Lipschitz con constante de Lipschitz $\frac{L_0}{1-\beta L_1}$. \square

Lema 1.3.6. *Bajo las Suposiciones 1.3.1 y 1.3.2 b), para cada $\epsilon \in [0, \epsilon_0]$ y $\delta \in [0, \delta_0]$ se cumple que*

$$\mathbb{E}_y^{\pi_0^*} \sup_{a \in A(y_{t-1})} \left\{ \mathbb{E} W [H(y_{t-1}, a, \chi_{t-1}(\epsilon, \delta))] \right\} \leq \left(\frac{\gamma}{\beta} \right)^{t-1} W(y). \quad (1.8)$$

Demostración. Sean $\epsilon \in [0, \epsilon_0]$, $\delta \in [0, \delta_0]$ y $y \in Y$. De la Suposición 1.3.2 b), se tiene que

$$\mathbb{E} W [H(y_{t-1}, a, \chi_{t-1}(\epsilon, \delta))] \leq \frac{\gamma}{\beta} W(y_{t-1}),$$

para cualquier $t \geq 1$ fijo. Entonces se consigue que

$$\mathbb{E}_y^{\pi_0^*} \sup_{a \in A(y_{t-1})} \left\{ \mathbb{E} W [H(y_{t-1}, a, \chi_{t-1}(\epsilon, \delta))] \right\} \leq \frac{\gamma}{\beta} \mathbb{E}_y^{\pi_0^*} W(y_{t-1}),$$

para cualquier $t \geq 1$ fijo. Ahora, considere $\hat{h}_t = \{y, a_1, y_1, a_2, \dots, y_{t-1}, a_t\}$, la historia del proceso conjunto descrito por las ecuaciones (1.1) y (1.2) bajo la política $\pi_0^* = \{f^*, f^*, \dots\}$, entonces

$$\begin{aligned} \mathbb{E}_y^{\pi_0^*} W(y_{t-1}) &= \mathbb{E}_y^{\pi_0^*} W(H(y_{t-2}, a_{t-2}, \chi_{t-2}(\epsilon, \delta))) \\ &= \mathbb{E}_y^{\pi_0^*} \left[\mathbb{E} W(H(y_{t-2}, a_{t-2}, \chi_{t-2}(\epsilon, \delta))) | \hat{h}_{t-2} \right] \\ &\leq \frac{\gamma}{\beta} \mathbb{E}_y^{\pi_0^*} [W(y_{t-2}) | \hat{h}_{t-2}] \\ &= \frac{\gamma}{\beta} \mathbb{E}_y^{\pi_0^*} [W(H(y_{t-3}, a_{t-3}, \chi_{t-3}(\epsilon, \delta))) | \hat{h}_{t-2}] \\ &= \frac{\gamma}{\beta} \mathbb{E}_y^{\pi_0^*} W(H(y_{t-3}, a_{t-3}, \chi_{t-3}(\epsilon, \delta))). \end{aligned}$$

Por lo tanto,

$$\begin{aligned} \mathbb{E}_y^{\pi_0^*} \sup_{a \in A(y_{t-1})} \{ \mathbb{E} W(H(y_{t-1}, a, \chi_{t-1}(\epsilon, \delta))) \} \\ \leq \left(\frac{\gamma}{\beta} \right)^2 \mathbb{E}_y^{\pi_0^*} W(H(y_{t-3}, a_{t-3}, \chi_{t-3}(\epsilon, \delta))). \end{aligned}$$

Continuando con este procedimiento iterativo se tiene que

$$\mathbb{E}_y^{\pi_0^*} \sup_{a \in A(y_{t-1})} \{ \mathbb{E} W(H(y_{t-1}, a, \chi_{t-1}(\epsilon, \delta))) \} \leq \left(\frac{\gamma}{\beta} \right)^{t-1} W(y).$$

□

La demostración del siguiente teorema está basada en el Teorema 1 de [30].

Teorema 1.3.7. *Bajo las Suposiciones 1.3.1 y 1.3.3 se cumple que*

$$\Delta_{\epsilon, \delta}(y, \pi_0^*) \leq \hat{C}(y) \hat{\delta}_{\epsilon, \delta}, \quad y \in Y,$$

donde

$$\hat{C}(y) = \frac{2\beta L_0 \max\{L_{2,x}, L_{2,\alpha}\}}{1 - \beta L_1} \left[\frac{1}{1 - \beta} + \frac{\beta}{(1 - \gamma)^2} W(y) \right],$$

para cada $\epsilon \in [0, \epsilon_0]$ y $\delta \in [0, \delta_0]$.

Demostración. Observe que para $\epsilon \in [0, \epsilon_0]$ y $\delta \in [0, \delta_0]$ se cumple que $V_{\epsilon, \delta}$ y $f_{\epsilon, \delta}$ satisfacen la siguiente ecuación de optimalidad

$$\begin{aligned} V_{\epsilon, \delta}(y) &= \inf_{a \in A(y)} \{ c(y, a) + \beta \mathbb{E} V_{\epsilon, \delta} [H(y, a, \chi(\epsilon, \delta))] \} \\ &= c(y, f_{\epsilon, \delta}(y)) + \beta \mathbb{E} V_{\epsilon, \delta} [H(y, f_{\epsilon, \delta}(y), \chi(\epsilon, \delta))] . \end{aligned} \tag{1.9}$$

CAPÍTULO 1. APROXIMACIÓN DETERMINISTA DE PDMS CON DINÁMICA ACOPLADA

Denote

$$R_{\epsilon,\delta}(y, a) := c(y, a) + \beta \mathbb{E} V_{\epsilon,\delta} [H(y, a, \chi(\epsilon, \delta))], \quad (y, a) \in \mathbb{K}, \quad (1.10)$$

y considere $\hat{h}_t = \{y, a_1, y_1, a_2, \dots, y_{t-1}, a_t\}$, como en la demostración del Lema 1.3.6. Por la propiedad de Markov se verifica que

$$\begin{aligned} \mathbb{E}^{\pi_0^*} [\beta V_{\epsilon,\delta}(y_t) | \hat{h}_t] &= R_{\epsilon,\delta}(y_{t-1}, a_t) - c(y_t, a_t) \\ &\quad - \inf_{a \in A(y_{t-1})} R_{\epsilon,\delta}(y_{t-1}, a) + \inf_{a \in A(y_{t-1})} R_{\epsilon,\delta}(y_{t-1}, a) \end{aligned} \quad (1.11)$$

Denotando $\Lambda_t^{\epsilon,\delta} := R_{\epsilon,\delta}(y_{t-1}, a_t) - \inf_{a \in A(y_{t-1})} R_{\epsilon,\delta}(y_{t-1}, a)$, por (1.11) se obtiene que

$$\mathbb{E}^{\pi_0^*} [\beta V_{\epsilon,\delta}(y_t) | \hat{h}_t] = \Lambda_t^{\epsilon,\delta} - c(y_t, a_t) + V_{\epsilon,\delta}(y_{t-1}). \quad (1.12)$$

Si tomamos valor esperado en (1.12), obtenemos que

$$\mathbb{E}_y^{\pi_0^*} [\beta V_{\epsilon,\delta}(y_t)] = \mathbb{E}_y^{\pi_0^*} [V_{\epsilon,\delta}(y_{t-1})] - \mathbb{E}_y^{\pi_0^*} [c(y_{t-1}, a_t)] + \mathbb{E}_y^{\pi_0^*} [\Lambda_t^{\epsilon,\delta}]. \quad (1.13)$$

Sumando (1.13) sobre $t = 1, 2, \dots, n$ con pesos β^{t-1} conseguimos que

$$\begin{aligned} \sum_{t=1}^n \beta^{t-1} \mathbb{E}_y^{\pi_0^*} [c(y_{t-1}, a_t)] &= \sum_{t=1}^n \beta^{t-1} [\mathbb{E}_y^{\pi_0^*} V_{\epsilon,\delta}(y_{t-1}) - \beta \mathbb{E}_y^{\pi_0^*} V_{\epsilon,\delta}(y_t)] \\ &\quad + \sum_{t=1}^n \beta^{t-1} \mathbb{E}_y^{\pi_0^*} [\Lambda_t^{\epsilon,\delta}] \\ &= V_{\epsilon,\delta}(y) - \beta^n \mathbb{E}_y^{\pi_0^*} V_{\epsilon,\delta}(y_n) + \sum_{t=1}^n \beta^{t-1} \mathbb{E}_y^{\pi_0^*} \Lambda_t^{\epsilon,\delta}. \end{aligned} \quad (1.14)$$

Como $V_{\epsilon,\delta} \in \mathbb{B}_W$, $\lim_{n \rightarrow \infty} \beta^n \mathbb{E}_y^{\pi_0^*} V_{\epsilon,\delta}(y_n) = 0$. Por lo tanto, cuando $n \rightarrow \infty$, se sigue de (1.14) que

$$\Delta_{\epsilon,\delta_0}(y, \pi_0^*) = \sum_{t=1}^{\infty} \beta^{t-1} \mathbb{E}_y^{\pi_0^*} \Lambda_t^{\epsilon,\delta} = \sum_{t=1}^{\infty} \beta^{t-1} \mathbb{E}_y^{\pi_0^*} c(y_{t-1}, a_t) - V_{\epsilon,\delta}(y). \quad (1.15)$$

Ahora, por (1.9) y (1.10), se sigue que

$$R_{0,0}(y, f^*(y)) = \inf_{a \in A(y)} R_{0,0}(y, a),$$

entonces

$$\begin{aligned} \Lambda_t^{\epsilon,\delta} &= R_{\epsilon,\delta}(y_{t-1}, a) - R_{0,0}(y_{t-1}, f^*(y_{t-1})) \\ &\quad + \inf_{a \in A(y_{t-1})} \{R_{0,0}(y_{t-1}, a)\} - \inf_{a \in A(y_{t-1})} \{R_{\epsilon,\delta}(y_{t-1}, a)\}, \end{aligned}$$

esto implica que

$$\begin{aligned} \Lambda_t^{\epsilon, \delta} &\leq R_{\epsilon, \delta}(y_{t-1}, a) - R_{0,0}(y_{t-1}, f^*(y_{t-1})) \\ &\quad + \sup_{a \in A(y_{t-1})} \{R_{0,0}(y_{t-1}, a) - R_{\epsilon, \delta}(y_{t-1}, a)\}. \end{aligned}$$

Por lo tanto,

$$\begin{aligned} |\Lambda_t^{\epsilon, \delta}| &\leq 2 \sup_{a \in A(y_{t-1})} |R_{\epsilon, \delta}(y_{t-1}, a) - R_{0,0}(y_{t-1}, a)| \\ &\leq 2\beta \sup_{a \in A(y_{t-1})} |\mathbb{E}V_{\epsilon, \delta}(H(y_{t-1}, a, \chi(\epsilon, \delta))) - \mathbb{E}V(H(y_{t-1}, a, \chi(0, 0)))|, \end{aligned}$$

donde el valor esperado en el último término es tomado con respecto al vector aleatorio $\chi(\epsilon, \delta)$ para t fijo. De la última desigualdad se sigue que

$$\begin{aligned} |\Lambda_t^{\epsilon, \delta}| &\leq 2\beta \sup_{a \in A(y_{t-1})} |\mathbb{E}V_{\epsilon, \delta}(H(y_{t-1}, a, \chi(\epsilon, \delta))) - \mathbb{E}V_{\epsilon, \delta}(H(y_{t-1}, a, \chi(0, 0)))| \\ &\quad + 2\beta \sup_{a \in A(y_{t-1})} |\mathbb{E}V_{\epsilon, \delta}(H(y_{t-1}, a, \chi(0, 0))) - \mathbb{E}V(H(y_{t-1}, a, \chi(0, 0)))| \\ &\leq 2\beta \mu_1(\chi(\epsilon, \delta), \chi(0, 0)) \\ &\quad + 2\beta \|V_{\epsilon, \delta} - V\|_W \sup_{a \in A(y_{t-1})} \mathbb{E}W(H(y_{t-1}, a, \chi(0, 0))), \end{aligned} \tag{1.16}$$

donde

$$\mu_1(\chi(\epsilon, \delta), \chi(0, 0)) = \sup_{(y, a) \in \mathbb{K}} |\mathbb{E}V_{\epsilon, \delta}(H(y, a, \chi(\epsilon, \delta))) - \mathbb{E}V_{\epsilon, \delta}(H(y, a, \chi(0, 0)))|.$$

De la Proposición 8.3.9 parte (a) de [38], se puede demostrar que

$$T_{\epsilon, \delta} u(y) := \inf_{a \in A(y)} \left\{ c(y, a) + \beta \mathbb{E}u(H(y, a, \chi(\epsilon, \delta))) \right\},$$

es un operador contractivo en \mathbb{B}_W con módulo γ , para cada $\epsilon \in [0, \epsilon_0]$ y $\delta \in [0, \delta_0]$. Como $V_{\epsilon, \delta}$ y V son puntos fijos para el operador $T_{\epsilon, \delta}$ se consigue que

$$\|V_{\epsilon, \delta} - V\|_W \leq \|T_{\epsilon, \delta} V_{\epsilon, \delta} - T_{0,0} V_{\epsilon, \delta}\|_W + \|T_{0,0} V_{\epsilon, \delta} - T_{0,0} V\|_W.$$

Esta última relación implica que

$$\begin{aligned} &\|V_{\epsilon, \delta} - V\|_W \\ &\leq (1 - \gamma)^{-1} \|T_{\epsilon, \delta} V_{\epsilon, \delta} - T_{0,0} V_{\epsilon, \delta}\|_W \\ &\leq \frac{\beta}{1 - \gamma} \sup_{y \in \mathbb{Y}} \left\{ \frac{\sup_{a \in A(y)} |\mathbb{E}V_{\epsilon, \delta}[H(y, a, \chi(\epsilon, \delta))] - \mathbb{E}V_{\epsilon, \delta}[H(y, a, \chi(0, 0))]|}{W(y)} \right\}. \end{aligned}$$

Combinando la desigualdad (1.8) del Lema 1.3.6 y las expresiones (1.16) y (1.17), se consigue que

$$\mathbb{E}_{y_0}^{\pi_0^*} |\Lambda_t^{\epsilon, \delta}| \leq 2\beta \left[1 + \frac{\beta}{1 - \gamma} \left(\frac{\gamma}{\beta} \right)^{t-1} W(y) \right] \mu_1(\chi(\epsilon, \delta), \chi(0, 0)).$$

CAPÍTULO 1. APROXIMACIÓN DETERMINISTA DE PDMS CON DINÁMICA ACOPLADA

Finalmente, por (1.15) se obtiene que

$$\Delta_{\epsilon, \delta}(y, \pi_0^*) \leq 2\beta \left[\frac{1}{1-\beta} + \frac{\beta}{(1-\gamma)^2} W(y) \right] \mu_1(\chi(\epsilon, \delta), \chi(0, 0)).$$

Por el Lema 1.3.5, se tiene que

$$\Delta_{\epsilon, \delta}(y, \pi_0^*) \leq 2\beta \frac{L_0}{1-\beta L_1} \left[\frac{1}{1-\beta} + \frac{\beta}{(1-\gamma)^2} W(y) \right] \mathbf{L}(\chi(\epsilon, \delta), \chi(0, 0)). \quad (1.18)$$

Considere el caso particular $\chi' = \chi(0, 0)$ on (1.7), entonces se sigue que

$$\mathbf{L}(\chi(\epsilon, \delta), \chi(0, 0)) = \mathbb{E}r(\chi(\epsilon, \delta), \chi(0, 0)) = \hat{\delta}_{\epsilon, \delta}.$$

Por lo tanto, sustituyendo la igualdad anterior en la ecuación (1.18) el resultado se sigue. \square

Observación 1.3.8. *Observe que el Teorema 1.3.7 garantiza que la política óptima del sistema determinista (ver (1.3) y (1.4)) $\pi_0^* \in \mathbb{F}$ es asintóticamente óptima para el sistema estocástico (ver (1.1) y (1.2)), i.e.,*

$$\lim_{\epsilon, \delta \rightarrow 0} \left| \hat{V}_{\epsilon, \delta}(y, \pi_0^*) - V_{\epsilon, \delta}(y) \right| = 0.$$

En el siguiente lema se verifica la continuidad de la función f^* , bajo el supuesto de que existe una única política óptima π_0^* . La unicidad de la política óptima es una condición restrictiva, pero en [16] se pueden encontrar tres bloques de condiciones para las componentes del modelo de decisión para garantizar esta suposición. En particular, en [16] se proporcionan condiciones para la unicidad cuando el espacio de estados es un subconjunto de \mathbb{R}^n .

Lema 1.3.9. *Bajo las Suposiciones 1.3.1, 1.3.2 y si además, la política óptima estacionaria para el problema determinista $\pi_0^* = \{f^*, f^*, \dots\}$ es única, entonces f^* es una función continua.*

Demostración. Por contradicción se demostrará que para $\epsilon = 0$ y $\delta = 0$, la política óptima $f^* : Y \rightarrow A$ es una función continua. Bajo las Suposiciones 1.3.1 y 1.3.2 se tiene que

$$\begin{aligned} V(x, \alpha) &= \inf_{a \in A} \left\{ c(x, \alpha, a) + \beta V \left(F(x, \alpha, a, \xi(0)), G(\alpha, \eta(0)) \right) \right\} \\ &= c(x, \alpha, f^*(x, \alpha)) + \beta \left(F(x, \alpha, f^*(x, \alpha), s_1), G(\alpha, s_2) \right), \end{aligned} \quad (1.19)$$

$(x, \alpha) \in Y$. Suponga que existe $(\hat{x}, \hat{\alpha}) \in Y$ donde f^* no es continua. Entonces, existe una sucesión $\{(x_n, \alpha_n)\}$ tal que $(x_n, \alpha_n) \rightarrow (\hat{x}, \hat{\alpha})$ pero

$d_2(f^*(x_n, \alpha_n), f^*(\hat{x}, \hat{\alpha})) \not\rightarrow 0$, cuando $n \rightarrow \infty$. Después de tomar una subsucesión -si es necesario- sin pérdida de generalidad, existe $\tau > 0$ tal que $d_2(f^*(x_n, \alpha_n), f^*(\hat{x}, \hat{\alpha})) \geq \tau$, para todo $n \in \mathbb{N}$. Como A es compacto, existe una subsucesión $\{z_{n_k}\}$ de $\{z_n = f^*(x_n, \alpha_n)\}$ que converge a $z \in A$, donde $z \neq f^*(\hat{x}, \hat{\alpha})$. Considere (x_{n_k}, α_{n_k}) y $f^*(x_{n_k}, \alpha_{n_k})$ en vez de (x, α) y $f^*(x, \alpha)$ en (1.19), entonces se consigue que

$$V(x_{n_k}, \alpha_{n_k}) = c(x_{n_k}, \alpha_{n_k}, f^*(x_{n_k}, \alpha_{n_k})) + \beta V\left(F(x_{n_k}, \alpha_{n_k}, f^*(x_{n_k}, \alpha_{n_k}), \xi(0)), G(\alpha_{n_k}, \eta(0))\right).$$

Por la continuidad de las funciones c , F , G y V , cuando $k \rightarrow \infty$, se obtiene que

$$V(\hat{x}, \hat{\alpha}) = c(\hat{x}, \hat{\alpha}, z) + \beta V\left(F(\hat{x}, \hat{\alpha}, z, \xi(0)), G(\hat{\alpha}, \eta(0))\right),$$

$(\hat{x}, \hat{\alpha}) \in Y$. Por las Suposiciones 1.3.1 y 1.3.2, existe una política óptima \bar{f} con $z = \bar{f}(\hat{x}, \hat{\alpha})$. Pero $\bar{f}(\hat{x}, \hat{\alpha}) \neq f^*(\hat{x}, \hat{\alpha})$, lo cual contradice la unicidad de la política óptima. Por lo tanto, f^* es continua. \square

A continuación, se enuncia y demuestra el teorema principal.

Teorema 1.3.10. *Bajo las Suposiciones 1.3.1, 1.3.2 y 1.3.3, para cada $\epsilon \in [0, \epsilon_0]$ y $\delta \in [0, \delta_0]$ se cumplen las siguientes afirmaciones*

- (a) $\|V_{\epsilon, \delta} - V\|_W \leq \frac{\beta}{1-\gamma} \frac{L_0 \max\{L_{2,x}, L_{2,\alpha}\}}{1-\beta L_1} \hat{\delta}_{\epsilon, \delta}$.
- (b) Sea K un subconjunto compacto de Y . Si la política estacionaria óptima para el problema determinista $\pi_0^* = \{f^*, f^*, \dots\}$ es única, entonces $f_{\epsilon, \delta} \rightarrow f^*$ uniformemente sobre K cuando $\epsilon \rightarrow 0$ y $\delta \rightarrow 0$.

Demostración. (a) La expresión (1.17) implica que

$$\begin{aligned} \|V_{\epsilon, \delta} - V\|_W &\leq \beta(1-\gamma)^{-1} \sup_{y \in Y} \{W^{-1}(y) \\ &\sup_{a \in A(y)} |\mathbb{E}V_{\epsilon, \delta}(H(y, a, \chi(\epsilon, \delta))) - \mathbb{E}V_{\epsilon, \delta}(H(y, a, \chi(0, 0)))|\}. \end{aligned}$$

Ahora, por el Lema 1.3.5 y como $W(y) \geq 1$ para $y \in Y$, se tiene que

$$\begin{aligned} \|V_{\epsilon, \delta} - V\|_W &\leq \beta(1-\gamma)^{-1} \frac{L_0}{1-\beta L_1} \\ &\sup_{y \in Y} \sup_{a \in A(y)} \left| \mathbb{E}d_1\left(H(y, a, \chi(\epsilon, \delta)), H(y, a, \chi(0, 0))\right) \right|. \end{aligned}$$

Por otro lado, las siguientes expresiones son válidas:

$$\begin{aligned}
& \left| \mathbb{E} d_1 \left(H(y, a, \chi(\epsilon, \delta)), H(y, a, \chi(0, 0)) \right) \right| \\
&= \left| \mathbb{E} \max \left\{ d_x \left(F(x, \alpha, a, \xi(\epsilon)), F(x, \alpha, a, \xi(0)) \right), \right. \right. \\
&\quad \left. \left. d_\alpha \left(G(\alpha, \eta(\delta)), G(\alpha, \eta(0)) \right) \right\} \right| \\
&\leq \left| \mathbb{E} L_{2,x} r_1(\xi(\epsilon), \xi(0)) \right| 1_{\{d_\alpha \leq d_x\}} \left(\left(F(x, \alpha, a, \xi(\epsilon)), G(\alpha, \eta(\delta)) \right), \right. \\
&\quad \left. \left(F(x, \alpha, a, \xi(0)), G(\alpha, \eta(0)) \right) \right) + \\
&\quad \left| \mathbb{E} L_{2,\alpha} r_2(\eta(\delta), \eta(0)) \right| 1_{\{d_x < d_\alpha\}} \left(\left(F(x, \alpha, a, \xi(\epsilon)), G(\alpha, \eta(\delta)) \right), \right. \\
&\quad \left. \left(F(x, \alpha, a, \xi(0)), G(\alpha, \eta(0)) \right) \right),
\end{aligned}$$

donde $1_{\{d_\alpha \leq d_x\}}(\cdot)$ denota la función indicadora sobre $\{d_\alpha \leq d_x\}$, mientras que $1_{\{d_x < d_\alpha\}}(\cdot)$ denota la función indicadora sobre $\{d_x < d_\alpha\}$. Entonces, se concluye que

$$\|V_{\epsilon, \delta} - V\|_W \leq \beta(1 - \gamma)^{-1} \frac{L_0 \max\{L_{2,x}, L_{2,\alpha}\}}{1 - \beta L_1} \hat{\delta}_{\epsilon, \delta}.$$

(b) Suponga que existen $K \subset Y$ compacto, un número real $\tau > 0$ y sucesiones $\{\epsilon_n\}$, $\{\delta_n\}$ convergentes a 0 tales que

$$d_2(f_{\epsilon_n, \delta_n}(x_n, \alpha_n), f^*(x_n, \alpha_n)) \geq \frac{\tau}{2}, \quad n = 1, 2, \dots, \quad (1.20)$$

para alguna sucesión convergente $\{(x_n, \alpha_n)\} \subset K$, tal que $(x_n, \alpha_n) \rightarrow (x, \alpha) \in K$, cuando $n \rightarrow \infty$. Como A es compacto, elija una subsucesión $\{(x_m, \alpha_m)\}$ de $\{(x_n, \alpha_n)\}$ tal que $f_{\epsilon_n, \delta_n}(x_m, \alpha_m) \rightarrow a \in A$. Ahora, por la continuidad de f^* que provee el Lema 1.3.5 y por (1.20) se tiene que $d_2(a, f^*(x, \alpha)) \geq \frac{\tau}{2}$. Como f_{ϵ_n, δ_n} es una política óptima, análogamente a (1.19) se obtiene que para $m = 1, 2, \dots$,

$$\begin{aligned}
V_{\epsilon_m, \delta_m}(x_m, \alpha_m) &= c(x_m, \alpha_m, f_{\epsilon_m, \delta_m}(x_m, \alpha_m)) \\
&+ \beta \mathbb{E} V_{\epsilon_m, \delta_m} \left(F(x_m, \alpha_m, f^*(x_m, \alpha_m), \xi(\epsilon_m)), G(\alpha_m, \eta(\delta_m)) \right). \quad (1.21)
\end{aligned}$$

Por otro lado, note que

$$\begin{aligned}
 & \left| \mathbb{E} V_{\epsilon_m, \delta_m} \left(F(x_m, \alpha_m, f^*(x_m, \alpha_m), \xi(\epsilon_m)), G(\alpha_m, \eta(\delta_m)) \right) - \right. \\
 & \quad \left. V \left(F(x, \alpha, a, \xi(0)), G(\alpha, \eta(0)) \right) \right| \\
 & \leq \left| \mathbb{E} V_{\epsilon_m, \delta_m} \left(F(x_m, \alpha_m, f^*(x_m, \alpha_m), \xi(\epsilon_m)), G(\alpha_m, \eta(\delta_m)) \right) - \right. \\
 & \quad \left. V_{\epsilon_m, \delta_m} \left(F(x_m, \alpha_m, f^*(x_m, \alpha_m), \xi(0)), G(\alpha_m, \eta(0)) \right) \right| \\
 & + \left| V_{\epsilon_m, \delta_m} \left(F(x_m, \alpha_m, f^*(x_m, \alpha_m), \xi(0)), G(\alpha_m, \eta(0)) \right) - \right. \\
 & \quad \left. V \left(F(x_m, \alpha_m, f_{\epsilon_m, \delta_m}(x_m, \alpha_m), \xi(0)), G(\alpha_m, \eta(0)) \right) \right| \\
 & + \left| V \left(F(x_m, \alpha_m, f_{\epsilon_m, \delta_m}(x_m, \alpha_m), \xi(0)), G(\alpha_m, \eta(0)) \right) - \right. \\
 & \quad \left. V \left(F(x, \alpha, a, \xi(0)), G(\alpha, \eta(0)) \right) \right|. \tag{1.22}
 \end{aligned}$$

Por el Lema 1.3.5, el primer término del lado derecho de (1.22) es menor o igual que $\frac{L_0}{1-\beta L_1} \hat{\delta}_{\epsilon_m, \delta_m}$. Los términos restantes convergen a 0 cuando $m \rightarrow \infty$, por la continuidad de las funciones F , G y V . Por lo tanto, cuando $m \rightarrow \infty$, (1.21) se convierte en

$$V(x, \alpha) = c(x, \alpha, a) + \beta V \left(F(x, \alpha, a, \xi(0)), G(\alpha, \eta(0)) \right).$$

Por argumentos similares a los proporcionados en la demostración del Lema 1.3.9, se sabe que existe una política óptima \bar{f} con $a = \bar{f}(x, \alpha)$, pero $\bar{f}(x, \alpha) \neq f^*(x, \alpha)$. Lo cual contradice la unicidad de la política óptima. Por lo tanto, f_{ϵ_n, δ_n} converge uniformemente a f^* . \square

1.4. Ejemplos

En esta sección, se presentan dos ejemplos que ilustran la teoría desarrollada y dos ejemplos que no verifican alguna de las suposiciones del Teorema 1.3.7 y que, por tanto, proporcionan conclusiones muy distintas a las que provee tal resultado. En esta sección, se consideran d_x, d_α, d_2, r_1 y r_2 como la métrica usual en \mathbb{R} .

1.4.1. Problema de consumo-inversión

Será considerado un problema de consumo-inversión en el que un inversor debe destinar su riqueza actual, digamos x_t , entre inversión a_t y consumo $x_t - a_t$, en cada etapa $t = 0, 1, 2, \dots$. Además, en cada etapa t se impone un factor de descuento $\exp(-\alpha_t)$, que depende de la tasa de interés bancario actual α_t . Los espacios de estados y acciones serán

CAPÍTULO 1. APROXIMACIÓN DETERMINISTA DE PDMS CON DINÁMICA ACOPLADA

$X = A = [0, \infty)$. Suponiendo que no se permite el endeudamiento, el conjunto de controles admisibles toma la forma: $A(x, \alpha) = [0, x]$. Además se supone que el banco recibe al menos una tasa de interés de $\exp(\alpha^*) - 1$ para $\alpha^* > 0$. De este modo, el espacio de la tasa de descuento es $\Gamma = [\alpha^*, \infty)$. El proceso de estados $\{x_t\}$ y el proceso de descuentos $\{\alpha_t\}$ evolucionan de acuerdo con las ecuaciones en diferencias:

$$\begin{aligned} x_{t+1} &= \xi_t(\epsilon) (x_t - a_t), \\ \alpha_{t+1} &= \alpha_t + \eta_t(\delta), \end{aligned} \tag{1.23}$$

$t = 0, 1, 2, \dots$, con (x_0, α_0) fijo dado, $\{\xi_t\}$ y $\{\eta_t\}$ son sucesiones de variables aleatorias independientes e idénticamente distribuidas (v.a.i.i.d.) independientes de (x_0, α_0) que tienen distribución discreta con valores en $S_1 = S_2 = [0, 1]$.

Observación 1.4.1. *En particular, si $\eta_t(0) = s_2 = 0$, $t = 0, 1, \dots$, en la ecuación (1.23) el correspondiente PDM determinista tiene factor de descuento constante.*

El objetivo es maximizar la utilidad de consumo del inversor sobre todo $\pi \in \Pi$,

$$\hat{V}_{\epsilon, \delta}(\pi, x, \alpha) = \mathbb{E}_{(x, \alpha)}^{\pi} \left[\sum_{t=0}^{\infty} e^{-S_t} u(x_t, \alpha_t, a_t) \right],$$

donde $S_t = \alpha_0 + \alpha_1 + \dots + \alpha_{t-1}$ y u es la función de utilidad. Considere la función de utilidad u definida por

$$u(x, \alpha, a) = \frac{b}{\gamma_1} a^{\gamma_1},$$

$(x, \alpha, a) \in \mathbb{K}$, donde $b > 0$, $\gamma_1 \in (0, 1)$. Además, suponga que $\mu_{\gamma_1} := \mathbb{E}[\xi^{\gamma_1}] < \infty$ con $0 < \beta \mu_{\gamma_1} < 1$, donde $\beta = e^{-\alpha_0}$. Por la definición de la función de utilidad, de manera inmediata se satisfacen las Suposiciones 1.3.1 b) y 1.3.3 a) con $L_0 := 1$.

Note que $A(x, \alpha)$ es compacto para todo $(x, \alpha) \in X \times \Gamma$. Ahora, considere H_a la métrica de Hausdorff, entonces para $(x, \alpha), (x', \alpha') \in X \times \Gamma$ se tiene que

$$\begin{aligned} H_a(A(x, \alpha), A(x', \alpha')) &= H_a([0, x], [0, x']) \\ &= |x - x'| \\ &\leq \max \left\{ |x - x'|, |\alpha - \alpha'| \right\} \\ &= d_1 \left((x, \alpha), (x', \alpha') \right). \end{aligned}$$

Luego el mapeo de valores-conjunto $(x, \alpha) \rightarrow A(x, \alpha)$ es continuo respecto a la métrica de Hausdorff, por lo que la Suposición 1.3.1 a) es válida. Además, por la continuidad de H la función $U'(x, \alpha, a)$ también es continua, por lo que se cumple 1.3.1 c).

Considere $W : X \times \Gamma \rightarrow [1, \infty)$ definida por

$$W(x, \alpha) = \frac{b\mu_{\gamma_1}}{\gamma_1 (1 - \beta\mu_{\gamma_1})} x^{\gamma_1} + 1, \quad (x, \alpha) \in X \times \Gamma.$$

En [19] se verifica que la función W satisface las Suposiciones 1.3.2 a), b). Además, note que

$$W'(x, \alpha, a) = \frac{b\mu_{\gamma_1}}{\gamma_1 (1 - \beta\mu_{\gamma_1})} (x - a) \mathbb{E}\xi(\epsilon) + 1,$$

es continua sobre \mathbb{K} , por lo que la Suposición 1.3.2 se sigue.

Por otra parte, para $(x, \alpha, a), (x', \alpha', a) \in \mathbb{K}$ y $(s_1, s_2) \in S_1 \times S_2$ se tiene que

$$\begin{aligned} d_1\left(H(x, \alpha, a, (s_1, s_2)), H(x', \alpha', a, (s_1, s_2))\right) &= \max\{x - x', |\alpha - \alpha'|\} \\ &\leq \max\{|x - x'|, |\alpha - \alpha'|\} \\ &= d_1((x, \alpha), (x', \alpha')), \end{aligned}$$

con lo que 1.3.3 b) se cumple para $L_1 := 1$.

Finalmente, se mostrará la condición 1.3.3 c):

$$\begin{aligned} &d_x(F(x, \alpha, a, \xi(\omega_1)), F(x, \alpha, a, \xi'(\omega'_1))) \\ &= |F(x, \alpha, a, \xi(\omega_1)) - F(x, \alpha, a, \xi'(\omega'_1))| \\ &= \left| \left(\xi(\omega_1) - \xi'(\omega'_1) \right) (x - a) \right| \\ &\leq x |\xi(\omega_1) - \xi'(\omega'_1)| \\ &= L_{2,x} r_1(\xi(\omega_1), \xi'(\omega'_1)), \end{aligned}$$

para todo $(x, \alpha, a) \in \mathbb{K}$ y $\xi(\omega_1), \xi'(\omega'_1) \in S_1$, donde $L_{2,x} := x$.

CAPÍTULO 1. APROXIMACIÓN DETERMINISTA DE PDMS CON DINÁMICA ACOPLADA

También, se tiene que

$$\begin{aligned}
 d_\alpha(G(\alpha, \eta(\omega_2)), G(\alpha, \eta'(\omega'_2))) &= |G(\alpha, \eta(\omega_2)) - G(\alpha, \eta'(\omega'_2))| \\
 &= |\alpha + \eta(\omega_2) - (\alpha + \eta'(\omega'_2))| \\
 &= \delta |\eta(\omega_2) - \eta'(\omega'_2)| \\
 &\leq \delta_0 |\eta(\omega_2) - \eta'(\omega'_1)| \\
 &= L_{2,\alpha} r_2(\eta(\omega_2), \eta'(\omega'_2)),
 \end{aligned}$$

para todo $\alpha \in \Gamma$ y $\eta(\omega_2), \eta'(\omega'_2) \in S_2$, donde $L_{2,\alpha} := \delta_0$.

Por el Teorema 1.3.7, se obtiene la desigualdad

$$\begin{aligned}
 &\Delta_{\epsilon,\delta}((x, \alpha), \pi_0^*) \\
 &\leq \frac{2\beta \max\{x, \delta_0\}}{1 - \beta} \left[\frac{1}{1 - \beta} + \frac{\beta}{(1 - \gamma)^2} \left(\frac{b\mu_{\gamma_1}}{\gamma_1(1 - \beta\mu_{\gamma_1})} x^{\gamma_1} + 1 \right) \right] \hat{R},
 \end{aligned}$$

donde $\hat{R} = \mathbb{E} \max\{|\xi(\epsilon) - \xi(0)|, |\eta(\delta) - \eta(0)|\}$. Por el Teorema 1.3.10 inciso a) la tasa de convergencia de la función de valor óptimo es:

$$\frac{\beta \max\{x, \delta_0\}}{(1 - \gamma)(1 - \beta)}.$$

Finalmente, por el Teorema 1.3.10 inciso b) para cualquier subconjunto K del espacio de estados $X \times \Gamma$ se tiene que

$$\sup_{(x,\alpha) \in K} |f_{\epsilon,\delta}(x, \alpha) - f^*(x, \alpha)| \rightarrow 0,$$

cuando $\epsilon \rightarrow 0$ y $\delta \rightarrow 0$.

1.4.2. Problema de control con ruido aditivo pequeño

Suponga que la dinámica del sistema está dada por las ecuaciones en diferencias siguientes:

$$\begin{aligned}
 x_{t+1} &= \frac{1}{2} (\alpha_t x_t + a_t + \xi_t(\epsilon)), \\
 \alpha_{t+1} &= h\alpha_t + \eta_t(\delta),
 \end{aligned} \tag{1.24}$$

$t = 0, 1, 2, \dots$, donde $0 < h \leq 1$ y $\{\xi_t(\epsilon)\}, \{\eta_t(\delta)\}$ son sucesiones de v.a.i.i.d. que toman valores en $S_1 = [0, \frac{B}{3}]$ y $S_2 = [0, \frac{1}{2}]$, respectivamente.

El espacio de x -estados es $X = [0, B]$, donde $0 < B < 6 \left(\frac{1}{\beta} - 1 \right)$, con β el factor de descuento y el espacio de α -estados es $\Gamma = [0, 1]$, es decir, $0 \leq \alpha \leq 1$. El espacio de control es $A = [0, \frac{B}{3}]$. El conjunto de controles admisibles en los estados (x, α) es $A(x, \alpha) = [0, x\alpha]$ y la función de costo es $c(x, \alpha, a) = x\alpha - a$, $(x, \alpha, a) \in \mathbb{K}$.

Observación 1.4.2. *En particular, si $\eta_t(0) = s_2 = 0$, $t = 0, 1, \dots$, y $h = 1$, el correspondiente PDM determinista tiene parámetro constante α .*

En este ejemplo, la Suposición 1.3.1 se verifica de manera inmediata. A continuación, serán verificadas las Suposiciones 1.3.2 y 1.3.3.

Considere $W : X \times \Gamma \rightarrow [1, \infty)$ definida por $W(x, \alpha) = x + 1$ para todo $(x, \alpha) \in X \times \Gamma$. Entonces se cumple que

$$\begin{aligned} |c(x, \alpha, a)| &= (x\alpha - a) \\ &\leq x - a \\ &\leq x \\ &< x + 1 \\ &:= W(x, \alpha), \end{aligned}$$

para todo $(x, \alpha, a) \in \mathbb{K}$.

También, se tiene que

$$\begin{aligned} \mathbb{E}W[H(x, \alpha, a, (\xi(\epsilon), \eta(\delta)))] &= \mathbb{E}\left[\frac{1}{2}(\alpha x + a + \xi(\epsilon)) + 1\right] \\ &= \frac{1}{2}(\alpha x + a + \mathbb{E}\xi(\epsilon)) + 1 \\ &\leq \frac{1}{2}\left(2\alpha x + \frac{B}{3}\right) + 1 \\ &\leq \frac{1}{2}\left(2x + 2 + \frac{B}{3}\right) \tag{1.25} \\ &\leq x + 1 + \frac{B}{6} \\ &\leq \left(1 + \frac{B}{6}\right)(x + 1) \\ &= \frac{\gamma}{\beta}W(x, \alpha), \end{aligned}$$

para todo $(x, \alpha, a) \in \mathbb{K}$ donde $\gamma = \beta\left(1 + \frac{B}{6}\right)$. Es claro que $\gamma > \beta$. Además, note que como $B < 6\left(\frac{1}{\beta} - 1\right)$ se tiene que $\gamma < 1$ y por lo tanto, $\gamma \in (\beta, 1)$.

De la segunda igualdad en (1.25) se observa que la función

$$\mathbb{E}W[H(x, \alpha, a, \chi(\epsilon, \delta))],$$

es continua sobre \mathbb{K} . Por lo tanto, la Suposición 1.3.2 se cumple.

CAPÍTULO 1. APROXIMACIÓN DETERMINISTA DE PDMS CON DINÁMICA ACOPLADA

Finalmente, se verifica la Suposición 1.3.3. Note que

$$\begin{aligned}
 |c(x, \alpha, a) - c(x', \alpha', a)| &= |x\alpha - a - (x'\alpha' - a)| \\
 &= |x\alpha - (x'\alpha')| \\
 &\leq |x - x'| \\
 &\leq \max\{|x - x'|, |\alpha - \alpha'|\} \\
 &= L_0 d_1((x, \alpha), (x', \alpha')),
 \end{aligned}$$

para todo $(x, \alpha, a), (x', \alpha', a) \in \mathbb{K}$, donde $L_0 := 1$.

Las siguientes desigualdades se cumplen para la dinámica conjunta de los estados

$$\begin{aligned}
 &d_1\left(H(x, \alpha, a, (s_1, s_2)), H(x', \alpha', a, (s_1, s_2))\right) \\
 &= \max\left\{\left|\frac{1}{2}\alpha x - \frac{1}{2}\alpha' x'\right|, \left|h(\alpha - \alpha')\right|\right\} \\
 &\leq \max\left\{\frac{1}{2}|x - x'|, h|\alpha - \alpha'|\right\} \\
 &\leq L_1 d_1((x, \alpha), (x', \alpha')),
 \end{aligned}$$

para todo $(x, \alpha, a) \in \mathbb{K}$ y $(s_1, s_2) \in S_1 \times S_2$, donde $L_1 := \max\{\frac{1}{2}, h\} \leq 1$.

Finalmente, se verifican las condiciones Lipschitz para las funciones F y G respecto a las variables de perturbación:

$$\begin{aligned}
 &d_x\left(F(x, \alpha, a, \xi(\omega_1)), F(x, \alpha, a, \xi'(\omega'_1))\right) \\
 &= \left|\frac{1}{2}(\alpha x + a + \xi(\omega_1)) - \frac{1}{2}(\alpha x + a + \xi'(\omega'_1))\right| \\
 &= \left|\frac{1}{2}\xi(\omega_1) - \frac{1}{2}\xi'(\omega'_1)\right| \\
 &\leq \frac{1}{2}|\xi(\omega_1) - \xi'(\omega'_1)| \\
 &\leq L_{2,x} r_1\left(\xi(\omega_1), \xi'(\omega'_1)\right),
 \end{aligned}$$

para todo $(x, \alpha, a) \in \mathbb{K}$, $\xi(\omega_1), \xi'(\omega_1') \in S_1$, donde $L_{2,x} := \frac{1}{2}$ y

$$\begin{aligned} d_\alpha \left(G(\alpha, \eta(\omega_2)), G(\alpha, \eta'(\omega_2')) \right) &= \left| (h\alpha + \eta(\omega_2)) - (h\alpha + \eta'(\omega_2')) \right| \\ &= \left| \eta(\omega_2) - \eta'(\omega_2') \right| \\ &\leq L_{2,\alpha} r_2 \left(\eta(\omega_2), \eta'(\omega_2') \right), \end{aligned}$$

para cada $\alpha \in \Gamma$ y $\eta(\omega_2), \eta'(\omega_2') \in S_2$, donde $L_{2,\alpha} := 1$. Por lo tanto, la Suposición 1.3.3 se cumple.

Por el Teorema 1.3.7 se tiene que

$$\Delta_{\epsilon,\delta}((x, \alpha), \pi_0^*) \leq \frac{2\beta}{1 - \beta \left(\max \left\{ \frac{1}{2}, h \right\} \right)} \left[\frac{1}{1 - \beta} + \frac{\beta}{\left(1 - \left(\beta \left(1 + \frac{\beta}{6} \right) \right) \right)^2} (x + 1) \right] \hat{R},$$

donde $\hat{R} := \mathbb{E} \max \{ |\xi(\epsilon) - \xi(0)|, |\eta(\delta) - \eta(0)| \}$. Mientras que por el Teorema 1.3.10 la tasa de convergencia de la función de valor es:

$$\frac{\beta}{\left[1 - \left(\beta \left(1 + \frac{\beta}{6} \right) \right) \right] \left[1 - \beta \left(\max \left\{ \frac{1}{2}, h \right\} \right) \right]}.$$

Además, por el inciso b) del Teorema 1.3.10 se tiene que $f_{\epsilon,\delta}(x, \alpha) \rightarrow f^*(x, \alpha)$ sobre cualquier subconjunto K del espacio de estados $X \times \Gamma$, cuando $\epsilon \rightarrow 0$ y $\delta \rightarrow 0$.

1.4.3. Importancia de las suposiciones

Finalmente, presentamos dos ejemplos donde no se satisfacen las Suposiciones 1.3.2 y 1.3.3 y por lo tanto, no se garantizan las conclusiones del Teorema 1.3.7.

Ejemplo 1.4.3. Sean $X = [0, \infty)$, $\Gamma = [0, 1]$, $A = [1, \frac{1}{\beta}]$, $\epsilon, \delta \in [0, 1]$ y la función de costo en una etapa dada por

$$\begin{aligned} c(x, \alpha, 0) &= 1, \quad (x, \alpha) \in X \times \Gamma, \\ c(x, \alpha, a) &= \begin{cases} a, & x \in [0, 1], \alpha \in \Gamma, \\ a + x - 1, & x > 1, a \in (0, \frac{1}{\beta}), \alpha \in \Gamma, \end{cases} \\ c(x, \alpha, \frac{1}{\beta}) &= \begin{cases} 0, & x \in [0, 1], \alpha \in \Gamma, \\ x - 1, & x > 1, \alpha \in \Gamma. \end{cases} \end{aligned}$$

Considere las ecuaciones en diferencias:

$$\begin{aligned} x_{t+1} &= a_t x_t + \epsilon \alpha_t \xi_t, \\ \alpha_{t+1} &= k \alpha_t + \delta \eta_t, \end{aligned} \tag{1.26}$$

CAPÍTULO 1. APROXIMACIÓN DETERMINISTA DE PDMS CON DINÁMICA ACOPLADA

$t = 0, 1, \dots$, donde $\{\xi_t\}$ es una sucesión de v.a.i.i.d. con distribución uniforme sobre $(0, 1)$, $\eta_t = 0$, $t = 1, 2, \dots$ y $k < 1$. La aproximación determinista al proceso (1.26) está dada por las ecuaciones:

$$\begin{aligned} x_{t+1} &= a_t x_t, \\ \alpha_{t+1} &= k \alpha_t, \end{aligned} \tag{1.27}$$

$t = 1, 2, \dots$ y $k < 1$. Considere $x_0 = 0$ y $\alpha_0 = 1$, entonces para cualquier política de control en (1.27), se tiene que $(x_t, \alpha_t) = (0, h^t)$, $t = 1, 2, \dots$. Por lo tanto, la política $\pi_0^* = \{\frac{1}{\beta}, \frac{1}{\beta}, \dots\}$ proporciona el valor mínimo de $\hat{V}((0, 1), \pi_0^*) = V((0, 1)) = 0$. Ahora, si se aplica la política π_0^* en la ecuación (1.26) con estado inicial $(0, 1)$ se obtiene que

$$x_t = \frac{1}{\beta^t} + \frac{\epsilon k}{\beta^{t-1}} \sum_{i=0}^{t-1} (\beta k)^i \xi_{i+1}, \tag{1.28}$$

$t = 1, 2, \dots$. Observe que el primer término del lado derecho de (1.28) es mayor que 1, por tanto, $x_t > 1$ para $t = 1, 2, \dots$. Como $c(x, \alpha, \frac{1}{\beta}) = x - 1$ para $x > 1$ y usando (1.28) se consigue que

$$\begin{aligned} \beta^t \mathbb{E}_{(0,1)}^{\pi_0^*} c(x_t, \alpha_t, \frac{1}{\beta}) &= \beta^t \mathbb{E}_{(0,1)}^{\pi_0^*} \left[\frac{1}{\beta^t} + \frac{\epsilon k}{\beta^{t-1}} \sum_{i=0}^{t-1} (\beta k)^i \xi_{i+1} - 1 \right] \\ &= 1 + \frac{\epsilon k \beta}{2} \sum_{i=0}^{t-1} (\beta k)^i - \beta^t \\ &\geq \epsilon \frac{k \beta}{2} \sum_{i=0}^{t-1} (\beta k)^i \\ &= \epsilon \frac{k \beta}{2} \frac{1 - (k \beta)^t}{1 - k \beta}, \end{aligned}$$

$t = 1, 2, \dots$. Ahora, para cada $\epsilon_1 \in (0, \frac{1}{2})$ elija la política estacionaria $\pi_1 = \{\epsilon_1, \epsilon_1, \dots\}$, por (1.26) se obtiene que $x_t \in [0, 1]$, $t = 0, 1, 2, \dots$. Observe que

$$\begin{aligned} \hat{V}_{\epsilon, \delta}((0, 1), \pi_1) &= \mathbb{E}_{(0,1)}^{\pi_1} \left[\sum_{t=0}^{\infty} \beta^t c(y_t, \alpha_t, a_t) \right] \\ &= \mathbb{E}_{(0,1)}^{\pi_1} \left[\sum_{t=0}^{\infty} \beta^t \epsilon_1 \right] \\ &= \frac{\epsilon_1}{1 - \beta}. \end{aligned}$$

De esta manera,

$$0 \leq V_{\epsilon, \delta}((0, 1)) \leq \hat{V}_{\epsilon, \delta}((0, 1), \pi_1) \rightarrow 0,$$

cuando $\epsilon_1 \rightarrow 0$. Por otro lado, observe que $(k\beta)^t \leq k\beta$, para $t \geq 1$, entonces

$$\begin{aligned}
 \Delta_{\epsilon,\delta}((0,1),\pi_1) &= \hat{V}_{\epsilon,\delta}((0,1),\pi_1) - V_{\epsilon,\delta}((0,1)) \\
 &= \mathbb{E}_{(0,1)}^{\pi_1} \left[\sum_{t=0}^{\infty} \beta^t c(y_t, \alpha_t, a_t) \right] \\
 &\geq \epsilon \frac{k\beta}{2} \sum_{t=0}^{\infty} \frac{1 - (k\beta)^t}{1 - k\beta} \\
 &= \epsilon \frac{k\beta}{2} \sum_{t=1}^{\infty} \frac{1 - (k\beta)^t}{1 - k\beta} \\
 &\geq \frac{k\beta}{2} \sum_{t=1}^{\infty} \frac{1 - k\beta}{1 - k\beta} \\
 &= \infty.
 \end{aligned}$$

Por lo tanto, $\Delta_{\epsilon,\delta}((0,1),\pi_1) = \infty$. En este ejemplo, no se cumplen las condiciones de la Suposición 1.3.2, en particular, no existe una función $W : Y \rightarrow [1, \infty)$ continua tal que $|c(y, a)| \leq W(y)$, para $(y, a) \in \mathbb{K}$. En este caso, ocurre que $\Delta_{\epsilon,\delta}((0,1),\pi_1) = \infty$, para cualesquiera $\epsilon, \delta > 0$.

Ejemplo 1.4.4. Sean $X = \mathbb{R}$, $\Gamma = [0, \infty)$, $A = \{0, 1\}$, $\epsilon, \delta \in [0, 1]$ y para $i \in \{0, 1\}$ se define la función de costo en una etapa por

$$c(x, \alpha, i) = \begin{cases} 1, & x \leq 0, \alpha \in \Gamma, \\ 3, & \text{en otro caso.} \end{cases}$$

Además, considere las ecuaciones en diferencias:

$$\begin{aligned}
 x_{t+1} &= x_t \alpha_t (a_t - \epsilon \xi_t), \\
 \alpha_{t+1} &= h \alpha_t - \delta \eta_t,
 \end{aligned} \tag{1.29}$$

$t = 0, 1, \dots$, donde $\{\xi_t\}$ es una sucesión de variables aleatorias con distribución normal estándar y $\{\eta_t\}$ es una sucesión de variables aleatorias con distribución exponencial con parámetro 1, $h > 0$. La aproximación determinista al proceso (1.29) está dada por las ecuaciones:

$$\begin{aligned}
 x_{t+1} &= \alpha_t x_t a_t, \\
 \alpha_{t+1} &= h \alpha_t,
 \end{aligned} \tag{1.30}$$

$t = 0, 1, \dots$. Considere estados iniciales $x_0 = 1$ y $\alpha_0 = 1$. Se puede observar que la política $\pi_0^* = \{0, 0, \dots\}$ es óptima para el proceso determinista (1.30) y para $\epsilon > 0$, $\hat{V}_{\epsilon,\delta}((1,1),\pi_0^*) = \frac{1}{1-\beta}$. Mientras que

CAPÍTULO 1. APROXIMACIÓN DETERMINISTA DE PDMS CON DINÁMICA ACOPLADA

$\hat{V}_{\epsilon,\delta}((1, 1), \pi_1) = \frac{3}{1-\beta}$, con $\pi_1 = \{1, 1, \dots\}$. Por lo tanto,

$$\begin{aligned}\Delta_{\epsilon,\delta}((1, 1), \pi_1) &= \hat{V}_{\epsilon,\delta}((1, 1), \pi_1) - V_{\epsilon,\delta}((1, 1)) \\ &\geq \hat{V}_{\epsilon,\delta}((1, 1), \pi_1) - \hat{V}_{\epsilon,\delta}((1, 1), \pi_0^*) \\ &= \frac{3}{1-\beta} - \frac{1}{1-\beta} \\ &= \frac{2}{1-\beta}.\end{aligned}$$

En este ejemplo, la Suposición 1.3.3 no se verifica, en particular la función de costo c , no es una función Lipschitz. En este caso se concluye que $\Delta_{\epsilon,\delta}(1, 1) \geq \frac{2}{1-\alpha}$ aunque $\hat{\delta}_{\epsilon,\delta} \rightarrow 0$ cuando $\epsilon \rightarrow 0$ y $\delta \rightarrow 0$.

Tenga en cuenta que si $L_1 > 1$ en la Suposición 1.3.3 b), no hay garantía de encontrar una cota superior para $\Delta_{\epsilon,\delta}(y, \pi_0^*)$ ni encontrar una tasa de convergencia para la función de valor óptimo, cuando ocurre que $\beta L_1 > 1$.

En el siguiente capítulo se presenta una aplicación del enfoque de ecuaciones en diferencias acopladas, en el que se aborda un modelo de crecimiento económico [58]. El modelo original está dado por una ecuación en diferencias que no satisface las condiciones Lipschitz establecidas en las condiciones del modelo de control que se abordó en este capítulo. El modelo propuesto es perturbado por dos ruidos pequeños y mediante un enfoque de ecuaciones en diferencias acopladas, se presenta un teorema central del límite.

Capítulo 2

Comportamiento normal en un modelo de crecimiento económico

En este capítulo, se aborda un problema de crecimiento económico que se ha estudiado ampliamente en [58]. En tal artículo, los autores muestran que las trayectorias de acumulación de capital convergen a las trayectorias correspondientes de un modelo determinista. Otra de las conclusiones es un teorema del límite central funcional que muestra que a una tasa σ , las trayectorias del capital centrado son asintóticamente normales. Este resultado es válido tanto para el nivel de capital como para su logaritmo. La convergencia mencionada se logra bajo existencia de un estado estacionario para el sistema determinista.

Inicialmente se identificó el problema de crecimiento económico con un Proceso de Decisión de Markov que evolucionan través de dos ecuaciones en diferencias acopladas como las planteadas en el Capítulo 1 de este escrito, que se propusieron en [49]. Se abordó este problema de crecimiento económico considerando la tasa de depreciación del capital como una cantidad aleatoria, suponiendo que la depreciación del capital se desarrolla de acuerdo con una tasa que cuenta con una ecuación en diferencias para su evolución. Esto permitió resolver el problema usando Aprendizaje por Refuerzo (AR), específicamente se utilizó la metodología de Q -learning. Para lograr lo anterior, se discretizó el espacio de estados y el espacio de acciones asociados al PDM del problema de crecimiento económico. Se encontró la política casi-óptima que proporciona Q -learning y como consecuencia se presentan las realizaciones de la trayectoria óptima del sistema estocástico cuando el estado estacionario

CAPÍTULO 2. COMPORTAMIENTO NORMAL EN UN MODELO DE CRECIMIENTO ECONÓMICO

del sistema determinista se considera como el estado inicial del sistema aleatorio. Lo anterior representa una solución aproximada para el problema de crecimiento económico. Además, se presenta evidencia de la convergencia del sistema aleatorio al estado estable determinista, como el que se define en [15, 43]. Por otro lado, se abordó la convergencia de las trayectorias del capital. Como resultado se obtuvo un teorema límite en el contexto de dinámicas acopladas que involucra una función de los logaritmos de las dinámicas del capital en el modelo aleatorio y en el determinista. Este resultado se complementa con experimentos numéricos en donde se fijan los valores de los parámetros del modelo, y como resultado se presentan algunos histogramas y pruebas estadísticas que garantizan el comportamiento normal de la transformación de los logaritmos de los capitales de los sistemas determinista y aleatorio.

Lo anterior corresponde a un caso particular del estudio de pequeñas perturbaciones en PDMs, por lo que se resume brevemente el trabajo relacionado a esta línea. El estudio en [45], se consideró el problema de aproximación de un proceso de control estocástico por un proceso determinista en el caso continuo. En este artículo, los autores demostraron que el problema estocástico puede ser aproximado por un determinista cuando el ruido es pequeño y las fluctuaciones se vuelven rápidas. En este contexto, se demuestra que el control óptimo para el problema determinista es asintóticamente óptimo para problemas estocásticos. En el caso continuo, [23] aborda un problema similar, es decir, cuando los efectos del ruido en un sistema físico son pequeños, estos autores realizaron un análisis asintótico de aproximación de difusión y usaron esto para estimaciones deseadas del sistema original. Para PDMs a tiempo discreto, esta clase de problemas fueron estudiados por [17, 18], donde la dinámica del sistema está descrita por una sola ecuación en diferencias. La convergencia entre estos modelos también fue estudiada en [44]. Sin embargo, la convergencia fue estudiada usando sucesiones que pertenecen al conjunto de parejas de estado-acción admisibles, que se supone es un subconjunto de un espacio euclidiano. Además, este estudio se lleva a cabo bajo el supuesto de que el espacio de acciones es un conjunto compacto y que la función de costos es acotada. Ahora, cuando consideramos PDMs que se desarrollan con respecto a (1.1) y (1.2), los resultados que aparecen en [17] son generalizados.

El enfoque de ecuaciones acopladas puede aplicarse, por ejemplo, cuando se considera un factor de descuento aleatorio [25–28], donde la segunda ecuación en diferencias se refiere a la evolución del factor de descuento aleatorio. Recientemente, en el contexto de ecuaciones en diferencias con dinámicas acopladas, en [49] se imponen restricciones de continuidad de Lipschitz [39, 46] sobre las componentes del modelo de

control y se aplican técnicas de PD para obtener las conclusiones siguientes: una tasa de convergencia de la función de costo óptimo para el sistema aleatorio con respecto al sistema determinista y la convergencia uniforme de las políticas estocásticas óptimas para la política determinista cuando $\epsilon \rightarrow 0$ y $\delta \rightarrow 0$, sobre subconjuntos compactos del espacio de estados. Finalmente, aunque consideramos un enfoque de ecuaciones en diferencias acopladas, no contamos con condiciones Lipschitz para las componentes del modelo de crecimiento económico por lo que no podemos asegurar las conclusiones dadas en [49].

El capítulo está organizado como sigue. En la Sección 2.1, se establece la teoría relacionada a Q -learning que permite proporcionar una solución aproximada al problema de crecimiento económico. En la Sección 2.2, se presenta la formulación del problema de crecimiento económico a través de un PDM. En la Sección 2.3, se presenta un teorema límite en el contexto de dinámicas acopladas que involucra una función de los logaritmos de las dinámicas del capital en el modelo aleatorio y determinista. Posteriormente, en la Sección 2.4 se ilustran los experimentos numéricos en los que se muestran como resultados la trayectoria óptima del sistema aleatorio, histogramas y pruebas estadísticas que garantizan el comportamiento normal de la transformación de los logaritmos de los capitales del sistema determinista y del aleatorio.

2.1. Q -learning

En la literatura, es ampliamente reconocido que existen situaciones en las que los PDMs presentan inconvenientes, por ejemplo, cuando se trabaja con altas dimensiones o cuando se tiene problemas con el modelaje, lo que en conjunto se denomina la doble maldición de la Programación Dinámica [33]. Una salida que se puede proporcionar para evitar la doble maldición de la PD es la metodología de Aprendizaje por Refuerzo (AR). Esto es posible ya que el esfuerzo de modelado en AR es menor que el de PD. Una cuestión notable es que en PDMs a gran escala, el enfoque de PD no es factible, mientras que AR sigue siendo factible [33]. Además, para resolver problemas cuyas probabilidades de transición son difíciles de estimar, AR es un enfoque atractivo que produce soluciones casi óptimas. La principal herramienta empleada por AR es la simulación, la cual se utiliza para evitar el cálculo de las probabilidades de transición. Sin embargo, si se tiene acceso a las matrices de costos/recompensas y probabilidades de transiciones, se debe usar PD porque se garantiza que serán generadas soluciones óptimas, mientras que AR se mantiene con soluciones casi-óptimas. Como el problema de crecimiento económico se desarrolla en un entorno continuo, no contamos con una

CAPÍTULO 2. COMPORTAMIENTO NORMAL EN UN MODELO DE CRECIMIENTO ECONÓMICO

matriz de transición, por lo que usar Q -learning parece ser factible.

En [33] se establece el AR como una rama de PD, por lo que el algoritmo de AR se deriva de sus contrapartes de PD. En la metodología de PD, el primer paso es generar la matriz de probabilidad de transición y la matriz de transición de costos/recompensas, después se usan estas matrices en un algoritmo adecuado para generar la solución óptima. En AR no se estima ninguna de estas matrices, lo que se hace es simular el sistema utilizando las distribuciones de las variables aleatorias gobernantes. Posteriormente, dentro del simulador, se utiliza un algoritmo adecuado para obtener la solución. A continuación, se discutirán algunos de los conceptos clave relacionados con el AR: los llamados Q -factores, el algoritmo de Robbins-Monro, tamaños de paso y la mezcla de estas ideas para resolver PDMs dentro de los simuladores.

2.1.1. Q -factor

En AR, la función de valor se almacena en los denominados Q -factores. La función de valor de costo descontado se puede definir por la ecuación de optimización de Bellman que sigue:

$$J^*(i) = \max_{a \in A(i)} \left(\sum_{j=1}^n p_{ij}(a) [r(i, a, j) + \beta J^*(j)] \right), \quad (2.1)$$

para $i \in X$, donde $n = |X|$ es el número de estados de la cadena de Markov; $J^*(i)$ denota el i -ésimo elemento del vector de la función de valor; $A(i)$ es el conjunto de acciones admisibles en el estado i ; $p_{ij}(a)$ denota la probabilidad de transición en un paso de ir del estado i al estado j bajo la influencia de la acción a ; $r(i, a, j)$ denota la recompensa inmediata en el estado i cuando la acción a es seleccionada y como resultado el sistema transita al estado j y $\beta \in (0, 1)$ es el factor de descuento.

En PD, dado un estado $i \in X$, asociamos un solo elemento de la función de valor: $J^*(i)$. En cambio, en el enfoque de AR se utiliza una pareja de estado-acción para asociarlo a un vector llamado Q -factor. Para una pareja de estado-acción $(i, a) \in X \times A(i)$ se define el Q -factor mediante la expresión:

$$Q(i, a) = \sum_{j=1}^n p_{ij}(a) [r(i, a, j) + \beta J^*(j)]. \quad (2.2)$$

Observe que (2.1) y (2.2) en conjunto producen lo siguiente.

$$J^*(i) = \max_{a \in A(i)} Q(i, a). \quad (2.3)$$

La relación (2.3) se traduce en que si conocemos los valores de los Q -factores, es posible obtener la función de valor J^* . Usando la igualdad (2.3), la ecuación (2.2) se puede escribir como sigue.

$$Q(i, a) = \sum_{j=1}^n p_{ij}(a) \left[r(i, a, j) + \beta \max_{a \in A(i)} Q(i, a) \right], \quad (2.4)$$

para todo $(i, a) \in X \times A(i)$. La ecuación (2.4) puede interpretarse como la versión Q -factor de la ecuación de optimalidad de Bellman para el costo descontado de PDMs, ver ecuación (2.1). Por la expresión (2.4), es fácil ver que el algoritmo anterior es completamente equivalente al algoritmo de iteración de valor regular.

Es importante mencionar que los Q -factores deben ser actualizados en cada etapa del tiempo, por ello se super indexan con el número de iteración correspondiente. En PD, la actualización de Q se realiza de la siguiente manera:

$$Q^{k+1}(i, a) \leftarrow \sum_{j=1}^n p_{ij}(a) \left[r(i, a, j) + \beta \max_{b \in A(j)} Q^k(j, b) \right],$$

para $k = 1, \dots, k_{max}$, donde $k_{max} \in \mathbb{N}$ es el número máximo de iteraciones. Cuando se usa AR, se realiza un cambio en la regla de actualización de Q que propone PD. Este cambio se logra cuando se aplica el algoritmo de Robbins-Monro.

2.1.2. Q -factores y Robbins-Monro

El algoritmo de Robbins-Monro, es un algoritmo popular y ampliamente usado que puede ayudar a estimar la media de una variable aleatoria a partir de sus muestras. Considere x_i la i -ésima muestra independiente de una variable aleatoria X que tiene esperanza $E[X]$. Entonces, con probabilidad 1, ocurre que

$$E[X] = \lim_{k \rightarrow \infty} \frac{\sum_{i=1}^k x_i}{k}.$$

En esta expresión es posible descomponer la suma que aparece en el límite anterior. Denote la suma de las primeras k muestras de X por

CAPÍTULO 2. COMPORTAMIENTO NORMAL EN UN MODELO DE CRECIMIENTO ECONÓMICO

$S_k := \frac{\sum_{i=1}^k x_i}{k}$, $k \geq 1$. Con esta notación, se puede escribir

$$\begin{aligned}
 S_{k+1} &= \frac{\sum_{i=1}^{k+1} x_i}{k+1} \\
 &= \frac{\sum_{i=1}^k x_i + x_{k+1}}{k+1} \\
 &= \frac{kS_k + x_{k+1}}{k+1} \\
 &= \frac{kS_k + S_k - S_k + x_{k+1}}{k+1} \\
 &= \frac{(k+1)S_k - S_k + x_{k+1}}{k+1} \\
 &= S_k - \frac{S_k}{k+1} + \frac{x_{k+1}}{k+1} \\
 &= \left(1 - \frac{1}{k+1}\right)S_k + \frac{1}{k+1}x_{k+1}.
 \end{aligned}$$

Defina $\alpha_k := \frac{1}{k}$, para $k \geq 1$, entonces la expresión anterior se escribe como

$$S_{k+1} = (1 - \alpha_{k+1})S_k + \alpha_{k+1}x_{k+1},$$

la cual se denomina algoritmo de Robbins-Monro y α es llamada tamaño de paso o tasa de aprendizaje. En la literatura, se han estudiado diferentes reglas comúnmente usadas para α . Algunos tamaños de paso conocidos se comentan al final de esta subsección.

Ahora, se utilizará el algoritmo de Robbins-Monro para actualizar los Q -factores dentro de los simuladores. Se puede demostrar que cada Q -factor se puede expresar como un promedio de una variable aleatoria. Observe que con ayuda de la expresión (2.4) se consigue que

$$\begin{aligned}
 Q(i, a) &= \sum_{j=1}^n p_{ij}(a) \left[r(i, a, j) + \beta \max_{b \in A(j)} Q(j, b) \right] \\
 &= E \left[r(i, a, j) + \beta \max_{b \in A(j)} Q(j, b) \right] \\
 &= E[SAMPLE].
 \end{aligned}$$

Por lo tanto, si se generan muestras de la variable aleatoria que aparece entre corchetes en la expresión anterior, es posible usar el esquema que proporciona el algoritmo de Robbins-Monro para actualizar los Q -factores. De este modo, utilizando el algoritmo de Robbins-Monro se

tiene que

$$Q^{k+1}(i, a) \leftarrow (1 - \alpha_{k+1})Q^k(i, a) + \alpha_{k+1} \left[r(i, a, j) + \beta \max_{b \in A(j)} Q^k(j, b) \right], \quad (2.5)$$

para cada $(i, a) \in X \times A(i)$.

Esta última expresión tiene la característica de que no se necesitan las probabilidades de transición para la implementación del algoritmo. Note que en este caso, únicamente se abordó el criterio descontado ya que este es el adecuado para el problema de crecimiento económico en cuestión. En [50] se puede consultar el algoritmo de Q -learning en el que se utiliza el criterio promedio y con este enfoque se presentan resultados numéricos en un modelo de líneas de espera.

Para concluir esta subsección, se agregan algunos comentarios acerca de los tamaños de paso [32, 33]. En Q -learning, el tamaño de paso tiene gran impacto al desarrollar el algoritmo y obtener la solución. En los inicios de esta técnica, se probó con la tasa $\frac{1}{k}$, $k = 1, 2, \dots$. Sin embargo, se observó que la tasa de aprendizaje decae rápidamente a cero, lo cual podría no ser satisfactorio. Como consecuencia, se propuso una generalización de este tamaño de paso, en el que se agregaron valores constantes A y B , de tal manera que la nueva propuesta tiene la forma $\frac{A}{k+B}$, $k = 1, 2, \dots$. Otro tipo de paso conocido es la regla logarítmica, esta es $\frac{\log(k)}{k}$, $k = 1, 2, \dots$. Finalmente, se cuenta con el tamaño de paso dado por $\frac{A}{V(i,a)}$, donde A es una constante y $V(i, a)$ denota el número de visitas a la pareja estado-acción $(i, a) \in X \times A(i)$. En general, para obtener la convergencia al óptimo de las soluciones, es necesario que los tamaños de paso cumplan las siguientes condiciones

$$\sum_{k=1}^{\infty} \alpha^k = \infty \quad \text{y} \quad \sum_{k=1}^{\infty} (\alpha^k)^2 < \infty.$$

Adicionalmente, se han realizado algunos trabajos donde se comparan las soluciones obtenidas por Q -learning, cuando se prueban con las tasas antes mencionadas. Lo que se encontró es que el tamaño de paso $\frac{A}{k+B}$ produce las mejores soluciones para $A = 100$ y $B = 150$. Basado en lo anterior, la implementación numérica de este trabajo se desarrolló con la tasa de aprendizaje $\frac{A}{k+B}$, $k = 1, 2, \dots$, con A y B constantes.

CAPÍTULO 2. COMPORTAMIENTO NORMAL EN UN MODELO DE CRECIMIENTO ECONÓMICO

2.2. El modelo de crecimiento identificado con un PDM

El modelo de referencia para el análisis es una economía especializada de [10] con producción, acumulación de capital y crecimiento estocástico de la productividad.

Se supone que la producción se realiza por una función de producción Coob-Douglas de rendimientos constantes a escala con parámetro α ;

$$F(K, L) = K^\alpha (AL)^{1-\alpha},$$

donde K es el stock de capital, L es la oferta de mano de obra y A es el parámetro tecnológico de aumento de la mano de obra. Para simplificar se fija $L = 1$. Además, se supone que A evoluciona exógenamente como sigue:

$$\log(A_{t+1}) = \kappa + \log(A_t) + \sigma Z_{t+1},$$

donde $Z \sim N(0, 1)$ y $\kappa \geq 0$ es la tasa media de crecimiento tecnológico. Sea δ la tasa de depreciación del capital y C_t denota el consumo. Entonces la ecuación de la evolución del capital está dada por:

$$K_{t+1} = A_t^{1-\alpha} K_t^\alpha - C_t + (1 - \delta)K_t. \quad (2.6)$$

Resulta que los cocientes de capital y tecnología, $k_t = K_t/A_t$ y la del consumo a la tecnología, $c_t = C_t/A_t$, son estacionarios. Por lo tanto, se presenta el problema en términos de variables estacionarias. Normalizando el nivel de tecnología la expresión (2.6) produce la ecuación de la evolución del capital que sigue:

$$k_{t+1} = \theta Z_{t+1}^\sigma (k_t^\alpha - c_t + (1 - \delta)k_t), \quad (2.7)$$

donde $\theta = e^{-\kappa}$ y Z^σ tiene distribución lognormal. Un agente representativo tiene preferencias de consumo separables en el tiempo, con una utilidad periódica:

$$U(C) = \frac{C^{1-\gamma}}{1-\gamma} = A^{1-\gamma} \frac{c^{1-\gamma}}{1-\gamma},$$

con $\gamma \in (0, 1)$. El problema del planificador social consiste en elegir una sucesión de consumo que maximice la utilidad descontada esperada del agente representativo. Por lo tanto, resolvemos

$$\sup_{C_t} E \sum_{t=0}^{\infty} \beta^t U(C_t), \quad (2.8)$$

sujeto a (2.7).

2.2. EL MODELO DE CRECIMIENTO IDENTIFICADO CON UN PDM

En [58], se muestra que este problema de optimización de Markov tiene una solución que es un control de retroalimentación de la forma $c_t = c^\sigma(k_t)$. De esta manera se puede escribir la evolución del capital óptimo como

$$\begin{aligned} k_{t+1} &= \theta Z_{t+1}^\sigma (k_t^\alpha - c_t + (1 - \delta)k_t) \\ &\equiv \bar{f}^\sigma(Z_{t+1}^\sigma, k_t, c^\sigma), \end{aligned}$$

donde la notación en la segunda línea remarca la dependencia de la política de consumo desconocido c^σ . Esta política satisface la ecuación estocástica de Euler que sigue:

$$\begin{aligned} c^\sigma(k)^{-\gamma} &= \beta \int (\theta Z^\sigma)^\gamma c^\sigma \left(\bar{f}^\sigma(Z^\sigma, k, c^\sigma) \right)^{-\gamma} \left[\alpha \bar{f}^\sigma(Z^\sigma, k, c^\sigma)^{\alpha-1} + 1 - \delta \right] dG^\sigma(Z^\sigma), \end{aligned} \quad (2.9)$$

donde G^σ es la función de distribución log-normal.

El correspondiente modelo determinista está descrito por la evolución del capital dada por:

$$k_{t+1} = \theta (k_t^\alpha - c_t^0 + (1 - \delta)k_t) := \bar{f}^0(k_t)^2, \quad (2.10)$$

que se obtiene al reemplazar σ por 0 en (2.7). Cálculos muestran que este problema tiene una solución que es un control de retroalimentación de la forma $c_t = c^0(k_t)$. La política de consumo óptimo satisface la ecuación de Euler análoga a (2.9):

$$c^0(k)^{-\gamma} = \beta \theta^\gamma c^0 \left(\bar{f}^0(k, c^0) \right)^{-\gamma} \left[\alpha \bar{f}^0(k, c^0)^{\alpha-1} + 1 - \delta \right]. \quad (2.11)$$

Ahora, será derivada una expresión analítica para determinar el estado estable del sistema determinista (2.10). En la ecuación de Euler que aparece en la expresión (2.11), sustituya k^* en vez de k , entonces se obtiene que

$$c^0(k^*)^{-\gamma} = \beta \theta^\gamma c^0(k^*)^{-\gamma} \left[\alpha (k^*)^{\alpha-1} + 1 - \delta \right].$$

Despejando a k^* se consigue que el único estado estable determinista interior es:

$$k^* = \left(\frac{1 + (\delta - 1)\beta\theta^\gamma}{\alpha\beta\theta^\gamma} \right)^{\frac{1}{\alpha-1}}. \quad (2.12)$$

Finalmente, se puede observar en el caso determinista, que si se toma $k = k^*$ en la expresión (2.10) y se despeja a $c^0(k^*)$ se obtiene que

$$c^0(k^*) = (k^*)^\alpha + \frac{\theta(1 - \delta) - 1}{\theta} k^*.$$

CAPÍTULO 2. COMPORTAMIENTO NORMAL EN UN MODELO DE CRECIMIENTO ECONÓMICO

La expresión anterior y la representación del estado estable en (2.12) permitirán realizar una comparación entre el modelo determinista y el estocástico cuando $\sigma \rightarrow 0$. Tal comparación se realizará a través de simulación. En lo que sigue se identifica el problema de crecimiento económico con un PDM discreto.

El espacio de estados será considerado como $X_C = [0, K]$, el espacio de acciones es $A_C = [0, K]$ con $K > 0$. Además, el conjunto de acciones admisibles cuando el sistema se encuentra en el estado $k \in X$ es $A_C(k) = [0, k]$. La función de utilidad está definida por $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$, para $c \in A_C(k)$, $k \in X_C$, note que por esta condición c depende del estado actual k . La ley de transición denotada por Q_l se asume está inducida por la ecuación en diferencias (2.7). De esta manera, el PDM con el que se trabaja está definido por

$$\mathcal{M} = \{X_C, A_C, \{A_C(k) \mid k \in X\}, Q_l, u\}.$$

Como el modelo estudiado es continuo, se discretizan las componentes y se obtiene lo siguiente. Se considera el conjunto finito de estados $X = \{k_0, k_1, \dots, k_n\}$, con $k_0 < k_1 < \dots < k_n$ donde $k_0 = 0$ y $k_n = K$, $n \in \mathbb{N}$ fijo. El espacio de acciones se toma como $A = \{c_0, c_1, \dots, c_m\}$, con $c_0 < c_1 < \dots < c_m$ donde $c_0 = 0$ y $c_m = K$, $m \in \mathbb{N}$ fijo. De manera que el conjunto de acciones admisibles cuando el sistema se encuentra en el estado $k \in X$ es $A(k) = \{0, c_1, \dots, c_j\}$, con $c_j = k$. La función de utilidad está definida de la misma manera por $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$, $\gamma \in (0, 1)$ para $c \in A(k)$, $k \in X$. Para completar el PDM necesitamos de las probabilidades de transición controladas, denotadas por $p_{kl}(c)$, donde $k, l \in X$ y $c \in A(k)$, tales entradas corresponden a la probabilidad de encontrarse en el estado k aplicar la acción c y transitar al estado l . De manera clara se puede observar que no tenemos a la mano la matriz de transición, por lo que una opción factible es utilizar Q -learning. Por lo tanto, únicamente se necesita una trayectoria del proceso dado en (2.7).

2.3. Teorema del límite central

En esta sección presentamos una versión ligeramente generalizada del teorema del límite central en un contexto de ecuaciones en diferencias acopladas. Se inicia reescribiendo el modelo de crecimiento económico cuando la tasa de depreciación del capital es aleatoria.

Normalizando el nivel de tecnología y considerando a δ como un valor aleatorio, (2.6) produce las ecuaciones de la evolución del capital y de la evolución del parámetro τ , el cual interviene en la tasa de depreciación

2.3. TEOREMA DEL LÍMITE CENTRAL

del capital, que siguen

$$k_{t+1} = \theta Z_{t+1}^\sigma (k_t^\alpha - c_t + (1 - e^{-S_t})k_t), \quad (2.13)$$

$$\tau_{t+1} = \tau_t + \eta_{t+1}(\epsilon), \quad (2.14)$$

donde $\theta := e^{-\kappa}$ y Z^σ tiene distribución log-normal, Z^σ es un elemento genérico de $\{Z_t^\sigma\}$, e^{-S_t} es la tasa de depreciación del capital con $S_t = \sum_{i=0}^{t-1} \tau_i$ y $\tau_0 = \tau > 0$ y $\eta_t(\epsilon)$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas tales que $\eta_t(\epsilon) \rightarrow 0$ cuando $\epsilon \rightarrow 0$. La representación de la tasa de depreciación del capital se consideró como la mostrada en [28], en el que se aleatorizó el factor de descuento.

Las expresiones (2.7) y (2.16) se pueden identificar con las dinámicas de los x -estados y los α -estados, las cuales aparecen en las expresiones (1) y (2) de [49], respectivamente como sigue:

$$\begin{aligned} k_{t+1} &= F(k_t, \tau_t, c_t, \xi_{t+1}(\sigma)) \\ &:= \theta Z_{t+1}^\sigma (k_t^\alpha - c_t + (1 - e^{-S_t})k_t), \end{aligned} \quad (2.15)$$

$$\tau_{t+1} = G(\tau_t, \eta_{t+1}(\epsilon)) := \tau_t + \eta_{t+1}(\epsilon), \quad (2.16)$$

donde $\xi_{t+1}(\sigma) := \exp(-\sigma Z_{t+1}) = Z_{t+1}^\sigma$. Observe que cuando $\sigma \rightarrow 0$, se cumple que $\xi_t(\sigma) \rightarrow 1$.

En lo que sigue es conveniente descomponer la evolución del capital en su esperanza condicional y su componente martingala. De esta manera, se define la esperanza del lado derecho de (2.15) condicionada a $k_t = k$ como:

$$f^{\sigma, \epsilon}(k) := \theta e^{\frac{\sigma^2}{2}} \left(k^\alpha - c^\sigma(k) + (1 - e^{-\tau} E[e^{-n\xi(\epsilon)}])k \right), \quad (2.17)$$

y la componente aleatoria es

$$\nu^{\sigma, \epsilon}(k) := \theta e^{-\sigma Z_{t+1} - \frac{\sigma^2}{2}} \left(k^\alpha - c^\sigma(k) + (1 - e^{-\tau} E[e^{-n\xi(\epsilon)}])k \right). \quad (2.18)$$

Observe que en (2.17) y (2.18) se supone que la esperanza $E[e^{-n\xi(\epsilon)}]$ existe y es finita.

Si se define $(l_t, \iota_t) := (\log(k_t), \log(\tau_t))$, es posible reescribir las expresiones (2.15) y (2.16) como

$$l_{t+1} = g^{\sigma, \epsilon}(l_t) - \sigma Z_{t+1},$$

$$\iota_{t+1} = \log(\tau_t),$$

CAPÍTULO 2. COMPORTAMIENTO NORMAL EN UN MODELO DE CRECIMIENTO ECONÓMICO

donde

$$g^{\sigma, \epsilon}(l_t) = \log(e^{\alpha l_t} - c^{\sigma}(e^{l_t}) + (1 - e^{-S_t} e^{l_t})) - \kappa. \quad (2.19)$$

Además, las ecuaciones en diferencias acopladas para el correspondiente modelo determinista son:

$$\begin{aligned} k_{t+1} &= F(k_t, \tau_t, c_t, 1) \\ &:= \theta (k_t^{\alpha} - c_t^0 + (1 - e^{-\tau}) k_t) \\ &= f^{0,0}(k_t)^2, \end{aligned} \quad (2.20)$$

$$\tau_{t+1} = G(\tau_t, 0) := \tau_t, \quad (2.21)$$

con $\tau_0 = \tau > 0$, que se obtiene al reemplazar σ por 0 en (2.15) y $\xi_{t+1}(\epsilon)$ por 1 en (2.16). El problema de crecimiento óptimo es el análogo determinista de (2.8) con la ley de movimiento dada por (2.20) y (2.21). Además, la expresión correspondiente a (2.19) para $\epsilon = 0$ y $\sigma = 0$ es la siguiente.

$$l_{t+1} = g^{0,0}(l_t) = \log(e^{\alpha l_t} - c^0(e^{l_t}) + (1 - e^{-\tau} e^{l_t})) - \kappa. \quad (2.22)$$

En lo que sigue se busca comparar el modelo estocástico y el modelo determinista.

2.3.1. Comparación de los modelos estocásticos y deterministas

Para la comparación de modelos considere $\{(k_t^{\sigma}, \tau_t^{\epsilon})\}$ una realización de la trayectoria del stock del capital y del parámetro con el cual evoluciona la tasa de depreciación del capital para el modelo de crecimiento estocástico y $\{(k_t^0, \tau_t^0)\}$ una realización de la trayectoria del stock del capital y del parámetro con el cual evoluciona la tasa de depreciación del capital para el modelo determinista. Además, considere $\{(l_t^{\sigma}, l_t^{\epsilon})\}$ y $\{(l_t^0, l_t^0)\}$ los logaritmos de la trayectoria del stock del capital y del parámetro con el cual evoluciona la tasa de depreciación del capital, para el modelo estocástico y determinista, respectivamente. Cuando $\epsilon \rightarrow 0$ y $\sigma \rightarrow 0$ esperamos que, tomando el mismo valor inicial, las trayectorias de la solución del modelo de crecimiento estocástico podrían aproximarse a las del modelo determinista. En efecto esto es lo que ocurre.

En los resultados propuestos se considera la expresión de diferencias normalizadas que sigue:

$$X_t^{\sigma, \epsilon} := \frac{(l_t^{\sigma} - l_t^0) + (l_t^{\epsilon} - l_t^0)}{\sigma}. \quad (2.23)$$

Ahora veremos algunos de los supuestos sobre las componentes del modelo para presentar un teorema límite central funcional en el que intervenga

2.3. TEOREMA DEL LÍMITE CENTRAL

la expresión $X_t^{\sigma, \epsilon}$. A causa del amplio estudio de este modelo, se sabe que se requieren de algunas condiciones de suavidad, estabilidad y convergencia uniforme de las políticas, las cuales son adaptadas al contexto de una dinámica acoplada como sigue. Considere $l \in \mathcal{L} \subset \mathbb{R}$ y $k \in \mathcal{K} \subset \mathbb{R}_+$ con \mathcal{L}, \mathcal{K} compactos. En lo que sigue, denotamos la primera derivada de $g^{\sigma, \epsilon}$ y $f^{\sigma, \epsilon}$ por f_k y g_l .

Suposición 1. Sobre cualesquiera conjuntos compactos \mathcal{L}, \mathcal{K} , las funciones $g^{\sigma, 0}$ y $f^{\sigma, 0}$ son continuas, de clase C^2 y tienen primera y segunda derivadas acotadas para todo $\sigma \geq 0$.

Suposición 2. Sobre cualesquiera conjuntos compactos \mathcal{L}, \mathcal{K} , las funciones $g^{\sigma, 0} \rightarrow g^{0, 0}$ y $f^{\sigma, 0} \rightarrow f^{0, 0}$ uniformemente cuando $\sigma \rightarrow 0$.

Suposición 3. Sobre cualesquiera conjuntos compactos \mathcal{L}, \mathcal{K} , las funciones $g^{0, 0}$ y $f^{0, 0}$ tienen un único punto fijo l^* y k^* , respectivamente, el cual es estable, i.e. $|g_l^{0, 0}(l^*)| < 1$ y $|f_k^{0, 0}(k^*)| < 1$, y cuyo dominio de atracción incluye todo \mathcal{L} y \mathcal{K} , respectivamente.

Lo anterior, permite establecer los siguientes resultados.

Teorema 2.3.1. *Si las Suposiciones 1 y 2 se cumplen para $g^{\sigma, 0}$, entonces las diferencias normalizadas $\{X_t^{\sigma, \epsilon}\}$ convergen débilmente al proceso $\{X_t^{0, 0}\}$, cuando $\epsilon \rightarrow 0$ y $\sigma \rightarrow 0$. El proceso límite sigue la auto-regresión lineal gaussiana, dependiente del proceso determinista $\{l_t^0\}$:*

$$X_{t+1}^{0, 0} = g_l^{0, 0}(l_t^0)X_t^{0, 0} + Z_{t+1}, \quad (2.24)$$

donde $\{Z_{t+1}\}$ es una sucesión de variables aleatorias normales independientes e idénticamente distribuidas.

Además, si existe un único estado de equilibrio del modelo determinista, el Teorema 2.3.1 implica que las diferencias con respecto al estado estacionario son asintóticamente normales. Lo anterior se establece en el siguiente corolario.

Corolario 2.3.2. *Si las Suposiciones 1-3 se cumplen para $g^{\sigma, 0}$. Entonces, cuando $\epsilon, \sigma \rightarrow 0$ y $t \rightarrow \infty$ se cumple asintóticamente lo siguiente*

$$l_{t+1}^\sigma = g_l^{0, 0}(l^*)(l_t^\sigma - l^*) + \sigma Z_{t+1}.$$

Por lo tanto, cuando $\epsilon \rightarrow 0$, $\sigma \rightarrow 0$ y $t \rightarrow \infty$, $\{l_t^\sigma\}$ converge a un proceso gaussiano estacionario con media l^* y varianza $\frac{\sigma^2}{1 - g_l^{0, 0}(l^*)^2}$.

Este resultado muestra que el logaritmo del stock de capital sigue asintóticamente una auto-regresión lineal gaussiana centrada en el estado estacionario determinista.

CAPÍTULO 2. COMPORTAMIENTO NORMAL EN UN MODELO DE CRECIMIENTO ECONÓMICO

Demostración del Teorema 2.3.1. Por las ecuaciones de evolución para τ y l , que se encuentran en (2.16) y (2.22) respectivamente, la segunda diferencia que aparece en el numerador de la expresión (2.23) se escribe como sigue:

$$\begin{aligned}
 \iota_t^\epsilon - \iota_t^0 &= \log(\tau_t^\epsilon) - \log(\tau_t^0) \\
 &= \log\left(\tau + \sum_{i=1}^{t-1} \xi_i(\epsilon)\right) - \log(\tau) \\
 &= \log\left(1 + \frac{\sum_{i=1}^{t-1} \xi_i(\epsilon)}{\tau}\right).
 \end{aligned} \tag{2.25}$$

Cuando $\epsilon \rightarrow 0$ la expresión (2.25) converge a 0, por la continuidad de la función logaritmo y por la suposición de que $\xi(\epsilon) \rightarrow 0$, cuando $\epsilon \rightarrow 0$.

Por lo tanto, cuando $\epsilon \rightarrow 0$, se obtiene que

$$\{X_t^{\sigma, \epsilon}\} \rightarrow \{X_t^{\sigma, 0}\},$$

donde $X_t^{\sigma, 0} := \frac{l_t^\sigma - l_t^0}{\sigma}$. Por el Teorema 3.1 de [58], se sabe que cuando $\sigma \rightarrow 0$, $\{X_t^{\sigma, 0}\}$ converge débilmente al proceso $\{X_t^{0, 0}\}$, y la expresión (2.24) se cumple.

2.4. Experimentos numéricos

En esta sección se presentan los experimentos numéricos que utilizan el modelo discretizado de la sección anterior. Los experimentos numéricos se dividen en dos etapas. La primera consiste en determinar la solución del modelo de crecimiento económico cuando la tasa de depreciación del capital es aleatoria. La segunda etapa consiste en exhibir la conclusión del Teorema 2.3.1 a través de algunos histogramas que fueron generados por datos simulados. Adicionalmente, se realizaron pruebas estadísticas que garantizan, a un nivel de significancia $\alpha = 0.1$, que los datos generados por el proceso $X_t^{\sigma, \epsilon}$ tienen un comportamiento normal. Específicamente, nos interesa el comportamiento de la trayectoria óptima del proceso estocástico y el comportamiento de una transformación de los logaritmos de las expresiones que modelan la evolución del capital y la evolución de la tasa de depreciación del capital cuando $\sigma \rightarrow 0$ y $\epsilon \rightarrow 0$.

Para conseguir lo anterior se implementaron los procedimientos descritos en la Sección 2.1. Con AR, el procedimiento central para obtener la solución para el problema de optimización vía Q -learning consta de los siguientes 6 pasos.

1. *Inicialice los Q -factores en 0, es decir, establezca $Q(i, j) = 0$ para todo $i \in X$ y $j \in A(i)$. Defina el tamaño de paso α a utilizar, introduzca k_{max} : número de iteraciones máximo (suficientemente grande) y considere i el estado inicial. Iniciar con $k = 1$.*
2. *Simule una acción $a \in A(i)$ con probabilidad $\frac{1}{|A(i)|}$.*
3. *El siguiente estado, digamos j , se obtiene con la expresión (2.7) considerando los respectivos ajustes para obtener un elemento de X . De esta manera, obtenemos la utilidad $u(i, a, j)$.*
4. *Para $(i, a) \in X \times A(i)$, calcule*

$$Q^{k+1}(i, a) \leftarrow (1 - \alpha) Q^k(i, a) + \alpha \left[u(i, a, j) + \beta \max_{b \in A(j)} Q^k(j, b) \right].$$

5. *Actualice $k \leftarrow k + 1$. Si $k < k_{max}$ entonces $i \leftarrow j$ y regrese al paso 2, en otro caso vaya al punto 6.*
6. *Para cada $i \in X$, calcule*

$$d(i) = \arg \max_{b \in A(i)} Q(i, b),$$

donde d denota la ϵ -óptima política y pare.

CAPÍTULO 2. COMPORTAMIENTO NORMAL EN UN MODELO DE CRECIMIENTO ECONÓMICO

Los procedimientos desarrollados en este capítulo se muestran en los Algoritmos 1-3 que se encuentra en la Subsección 2.4.1 y fueron implementaron en el lenguaje de programación de R [52].

El Algoritmo 1 se presenta como una función principal que nos devuelve la trayectoria óptima del proceso, el capital promedio, genera un histograma y realiza una prueba estadística para los datos configurados mediante la expresión $X_{t+1}^{\sigma, \epsilon} - g_x(l_t^{0,0})X_t^{\sigma, \epsilon}$ que aparece en el Teorema 2.3.1.

El procedimiento numérico para obtener la solución al problema de crecimiento económico puede consultarse en el Algoritmo 2, aquí es donde se desarrolla la implementación de Q -learning para encontrar la política casi óptima. El Algoritmo 3 proporciona una trayectoria del proceso estocástico que modela el capital cuando la tasa de depreciación del capital es aleatoria.

Para los resultados de los experimentos se consideran los valores numéricos que se describen en la Tabla 2.1. La mayoría de estas estimaciones se tomaron del estudio [58], únicamente se ajustaron los valores del factor de descuento y del parámetro γ . Esta modificación se realizó ya que considerando el valor original de γ , la función de utilidad resultaba negativa.

Parámetro	Valor numérico
α	0.65
κ	0.0176
δ	0.0517
γ	0.4
β	0.7638761

Tabla 2.1: Valores numéricos de los parámetros.

Se puede observar que sustituyendo los valores de la Tabla 2.1 en la expresión (2.12) el valor del estado estable determinista es $k^* = 5$. En la Figura 2.1, se muestran las simulaciones de la trayectoria óptima del proceso conjunto dado por las expresiones (2.13) y (2.14) para un periodo de 43 unidades de tiempo. Los resultados arrojan aproximaciones del capital promedio de 7.77057 y 8.12322, para las representaciones (I) y (II), respectivamente. Además, en la Figura 2.2, se observa el comportamiento de la trayectoria óptima del capital en el modelo estocástico para un periodo de 300 unidades de tiempo. Las gráficas fueron obtenidas con los valores que siguen: iteración máxima 1000, la discretización del espacio de estados y el espacio de acciones se tomó con incrementos de 0.75 unidades para la Figura 2.1, mientras que para la Figura 2.2, los incrementos fueron de 0.1 unidades. En ambos casos se realizaron

2.4. EXPERIMENTOS NUMÉRICOS

1000 trayectorias para promediar y conseguir la trayectoria óptima promedio. Particularmente, en la Figura 2.2, se observa que cuando σ y ϵ , los parámetros que intervienen en el grado de estocasticidad del sistema convergen a cero, la trayectoria se comporta de manera determinista.

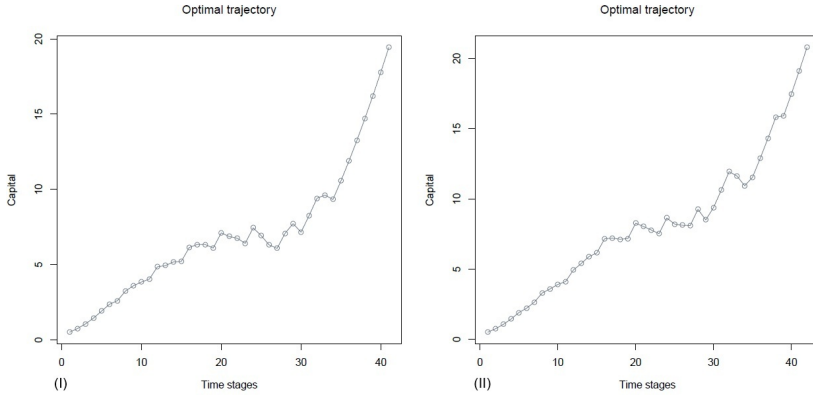


Figura 2.1: Realización de la trayectoria óptima del capital que se desarrolla mediante (2.13) y (2.14): (I) $\sigma = 4.92 \times 10^{-3}$ y $\epsilon = 1 \times 10^{-3}$, (II) $\sigma = 4.92 \times 10^{-6}$ y $\epsilon = 1 \times 10^{-6}$.

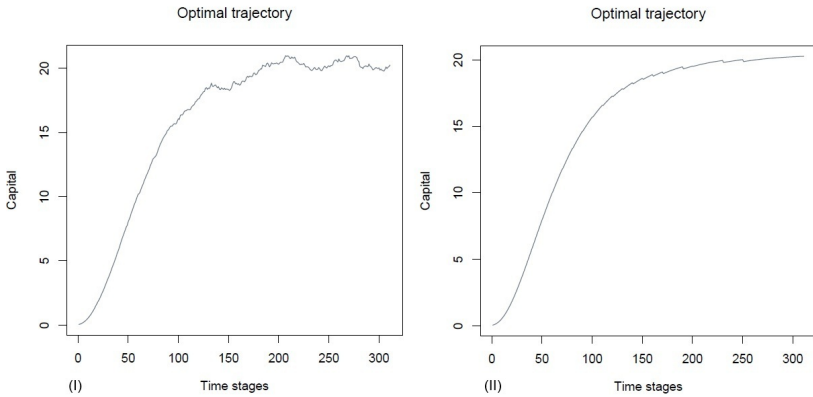


Figura 2.2: Realización de la trayectoria óptima del capital que se desarrolla mediante (2.13) y (2.14): (I) $\sigma = 4.92 \times 10^{-6}$ y $\epsilon = 1 \times 10^{-6}$, (II) $\sigma = 4.92 \times 10^{-11}$ y $\epsilon = 1 \times 10^{-11}$.

CAPÍTULO 2. COMPORTAMIENTO NORMAL EN UN MODELO DE CRECIMIENTO ECONÓMICO

En la figura 2.3, se muestran los histogramas generados por el proceso $X_{t+1}^{\sigma, \epsilon}$, para distintos valores de σ y ϵ . Adicionalmente, con los datos generados se realizó una prueba de Kolmogorov-Smirnov, para verificar la normalidad de $X_{t+1}^{\sigma, \epsilon} - g_x(l_t^{0,0})X_t^{\sigma, \epsilon}$. La prueba se realizó fijando el valor $\alpha = 0.1$, como resultado de la prueba se obtuvieron los p -valores que se muestran en la Tabla 2.4, de modo que la normalidad se verifica tal como lo establece el Teorema 2.3.1. En la última columna de la Tabla 2.4, podemos observar que cuando ϵ y σ convergen a cero, el valor promedio del capital se aproxima al valor del estado estable del sistema determinista que es 5. Esto fue posible ya que, los valores en Tabla 2.1, satisfacen que existe un único estado estable del sistema determinista.

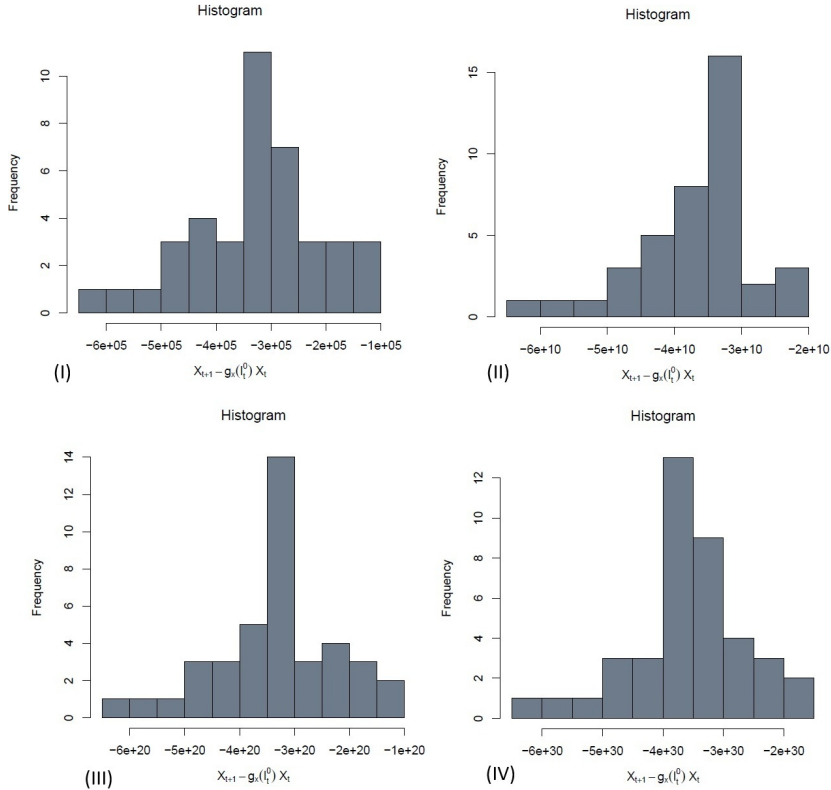


Figura 2.3: Histogramas: (I) $\sigma = 4.92 \times 10^{-6}$ y $\epsilon = 1 \times 10^{-6}$, (II) $\sigma = 4.92 \times 10^{-11}$ y $\epsilon = 1 \times 10^{-11}$, (III) $\sigma = 4.92 \times 10^{-21}$ y $\epsilon = 1 \times 10^{-21}$, (IV) $\sigma = 4.92 \times 10^{-31}$ y $\epsilon = 1 \times 10^{-31}$.

2.4. EXPERIMENTOS NUMÉRICOS

σ	ϵ	α	p -valor	Capital promedio
4.92×10^{-6}	1×10^{-6}	$\alpha = 0.1$	0.5765	7.51682
4.92×10^{-11}	1×10^{-11}	$\alpha = 0.1$	0.2091	5.11794
4.92×10^{-21}	1×10^{-21}	$\alpha = 0.1$	0.5764	7.104244
4.92×10^{-31}	1×10^{-31}	$\alpha = 0.1$	0.2023	5.352499

Tabla 2.2: Capital promedio y p -valores obtenidos para la prueba de Kolmogorov-Smirnov, para distintos valores de σ y ϵ .

CAPÍTULO 2. COMPORTAMIENTO NORMAL EN UN MODELO DE CRECIMIENTO ECONÓMICO

2.4.1. Algoritmos

Algorithm 1: Main function

Datos $\alpha, \beta, \delta, \gamma, \kappa, \text{max.ite}, \Delta_e, \Delta_a, k_0, \text{Max.aver}, \sigma, \epsilon$;
Resultado Average Capital, Optimal Capital trajectory and Histogram;
 $Pol = \text{GeneratePolicy}(\alpha, \beta, \delta, \gamma, \kappa, \text{max.ite}, \Delta_e, \Delta_a, k_0, \sigma, \epsilon)$;
 $N = \text{length}(Pol)$;
for $1 \leq j < M$ **do**
 $T = \text{matrix}(0, N, M)$;
 $Tau = \text{matrix}(0, N, M)$;
 $T[1, j] = k_0$;
 $Tau[1, j] = -\log(\delta)$;
 $T[2, j] = e^{-\sigma rnorm(1) - \kappa} (k_0^\alpha - pol[1] + (1 - e^{-\tau_0})k_0)$;
 $Tau[1, j] = -\log(\delta)$;
 for $2 \leq i < N$ **do**
 $Tau[i, j] = Tau[i - 1, j] + \epsilon rnorm(1)$;
 $T[i + 1, j] = e^{-\sigma rnorm(1) - \kappa} (k_0^\alpha - pol[i] + (1 - e^{-\tau_0})k_0)$;
for $1 \leq i \leq N$ **do**
 $Ta[i] = \text{mean}(T[i])$; $Tac[i] = \text{mean}(Tau[i])$;
 $\text{plot}(Ta)$: genera la trayectoria óptima;
 $X = \frac{\log(Ta) - \log(k_0) + \log(Tac) - \log(-\log(\delta))}{\sigma}$;
for $1 \leq i \leq N - 1$ **do**
 $Z[i] = X[i + 1] - g_x X[i]$;
 $\text{print}(\text{mean}(A))$: muestra el capital promedio;
 $\text{Histograma}(Z)$: crea el histograma;
 $\text{ks.test}(Z, \text{prnom}, \mu, \sigma)$ realiza la prueba de Kolmogorov-Smirnov con los datos generados.

2.4. EXPERIMENTOS NUMÉRICOS

Algorithm 2: Generate Policy

Datos $\alpha, \beta, \delta, \gamma, \kappa, max.ite, \Delta_e, \Delta_a, k_0, Max.aver, \sigma, \epsilon;$

Resultado Optimal policy;

$r(i, c) = \frac{e^{1-\gamma}}{1-\gamma}$: función de utilidad;

for $1 \leq j \leq 100$ **do**

$T = matrix(0, 2, max.ite);$

$T = Trajectory(100, 000, \alpha, \sigma, \epsilon, \kappa, \delta, \gamma, k_0);$

N_e : Número de estados;

N_a : Número de acciones;

$Q = matrix(0, N_e + 1, N_a + 1);$

$\alpha_t = 150/300;$

$i = (trunc(k_0) + 1) / \Delta_e;$

$p = array(1/(i + 1), i + 1);$

$a = sample(c(0 : i), 1, replace = T, prob = p)$: genera un número aleatorio entre 0 e i ;

$\tau_0 = -\log(\delta);$

$j = trunc\left(e^{-\sigma rnorm(1) - \kappa} (k_0^\alpha - a\Delta_a + (1 - e^{-\tau_0})k_0) + 1\right) / \Delta_e;$

$Q[i, a] = (1 - \alpha_t)Q[i, a] + \alpha_t (r(i\Delta_e, a\Delta_a) + \beta \max(Q[j,]));$

for $1 \leq s < max.ite$ **do**

$\tau_0 = \tau_0 + \epsilon rnorm(1);$

$i = j;$

$p = array(1/(i + 1), i + 1);$

$a = sample(c(0 : i), 1, replace = T, prob = p);$

$j =$

$trunc\left(e^{-\sigma rnorm(1) - \kappa} (k_0^\alpha - a\Delta_a + (1 - e^{-\tau_0})k_0) + 1\right) / \Delta_e;$

$\alpha_t = 150/(300 + s);$

$Q[i, a] = (1 - \alpha_t)Q[i, a] + \alpha_t (r(i\Delta_e, a\Delta_a) + \beta \max(Q[j,]));$

for $1 \leq s < max.ite$ **do**

$pol[s] = which.max(Q[s,]);$

return pol : genera la política óptima.

CAPÍTULO 2. COMPORTAMIENTO NORMAL EN UN MODELO DE CRECIMIENTO ECONÓMICO

Algorithm 3: Trajectory

Datos $N, \alpha, \sigma, \epsilon, \kappa, \gamma, k_0$;
Resultado Trajectory;
 $T = \text{array}(0, N)$;
 $A = \text{array}(0, N)$;
 $\tau_0 = -\log(\delta)$;
 $T[1] = k_0$;
 $A[1] = \text{runif}(1, 0, T[1])$;
genera un número aleatorio en el intervalo $(0, k_0)$;
 $T[2] = e^{-\sigma \text{norm}(1) - \kappa} (k_0^\alpha - A[1] + (1 - e^{-\tau_0})k_0)$;
 $A[2] = \text{runif}(1, 0, T[2])$;
genera un número aleatorio en el intervalo $(0, T[2])$
for $2 \leq i < N$ **do**
 $\tau_0 = \tau_0 + \epsilon \text{norm}(1)$;
 $T[i+1] = e^{-\sigma \text{norm}(1) - \kappa} (T[i]^\alpha - A[i] + (1 - e^{-\tau_0})k_0)$;
 $A[i+1] = \text{runif}(1, 0, T[i+1])$;
 $Ta = \text{matrix}(0, 2, N)$;
 $Ta[1,] = T$;
 $Ta[2,] = A$;
return(T): genera una trayectoria del proceso estocástico.

Capítulo 3

Cadenas de decisión de Markov promedio: caso propenso al riesgo

En este capítulo abordamos un tema independiente de procesos de decisión que evolucionan mediante un par de ecuaciones en diferencias acopladas y se da paso para trabajar con cadenas de decisión de Markov dotadas de un espacio de estados numerable, con función de costo acotada y criterio de rendimiento promedio propenso al riesgo. Las condiciones estructurales en la ley de transición garantizan que el costo promedio es constante, sin embargo, no es posible garantizar que la ecuación de optimalidad tenga solución. En esta situación se obtienen aproximaciones convergentes al costo promedio óptimo y se determinan aproximadamente las políticas estacionarias óptimas usando los puntos fijos de una familia de operadores contractivos. Esto representa una extensión del clásico enfoque descontado, en el caso neutral al riesgo. Las conclusiones son presentadas en el Teorema 3.2.1.

El estudio de cadenas de decisión de Markov con un criterio sensible al riesgo se remonta a los artículos proporcionados por Howard y Matheson [40], donde se analizaron cadenas de decisión de Markov con espacio de estados finito y el costo promedio óptimo fue caracterizado vía una ecuación de optimalidad. Modelos con espacio de estados finito o infinito son considerados, por ejemplo, en [11, 12], [55, 56] mientras que cadenas de decisión de Markov sobre espacios de estado de Borel fueron analizados en [20–22], [41, 42] y [54]. Juegos estocásticos con criterio sensible al riesgo son estudiados en [3].

CAPÍTULO 3. CADENAS DE DECISIÓN DE MARKOV PROMEDIO: CASO PROPENSO AL RIESGO

Este capítulo está organizado como sigue. En la Sección 3.1 se describe el modelo de decisión, se define el criterio promedio y se establecen las principales suposiciones sobre el modelo. En la Sección 3.2 se introduce una familia de operadores contractivos y se presenta el resultado principal de este capítulo en el Teorema 3.2.1. Las herramientas que se utilizarán para establecer dicho resultado se establecen en la Sección 3.3, y la demostración del resultado principal se presenta en la Sección 3.4.

Notación. En todo el capítulo, \mathbb{N} denota el conjunto de enteros no negativos y dado un espacio topológico S , el espacio de Banach de todas las funciones acotadas $H : S \rightarrow \mathbb{R}$ es denotado por $\mathcal{B}(S)$, la norma del supremo de $H \in \mathcal{B}(S)$ es denotada por $\|H\| := \sup_{x \in S} |H(x)|$. Por otro lado, cada (des) igualdad que involucra variables aleatorias es válida casi seguramente con respecto a la medida de probabilidad subyacente.

3.1. Modelo de decisión

Sea $\mathcal{M} := (S, A, \{A(x)\}_{x \in S}, C, [p_{x,y}(a)])$ una cadena de decisión de Markov, un modelo para un sistema dinámico cuyas componentes son las siguientes: El espacio de estados S es un conjunto numerable dotado con la topología discreta, el espacio métrico A es el conjunto de acciones, mientras que para cada estado $x \in S$, $A(x) \subset A$ es la clase de acciones admisibles (controles) en el estado x . Por otro lado $C : \mathbb{K} \rightarrow \mathbb{R}$ es la función de costo, donde $\mathbb{K} = \{(x, a) \mid x \in S, a \in A(x)\}$ es la familia de parejas admisibles y, finalmente $[p_{x,y}(a)]_{x,y \in S, a \in A(x)}$ es la ley de transición controlada. La interpretación de \mathcal{M} es la siguiente: en cada tiempo $t \in \mathbb{N}$ el controlador observa el estado del sistema $X_t = x \in S$, y entonces elige y aplica una acción $A_t = a \in A(x)$. Como consecuencia de esta intervención, (i) se incurre en un costo $C(x, a)$, y (ii) el sistema se mueve a un nuevo estado $X_{t+1} \in S$ donde, de acuerdo con los estados y acciones previos, el evento $[X_{t+1} = y]$ es observado con probabilidad $p_{x,y}(a)$, donde $\sum_{y \in S} p_{x,y}(a) = 1$; esta es la propiedad de Markov del proceso de decisión.

Suposición 3.1.1. (i) Para cada $x \in S$, $A(x)$ es un subconjunto compacto de A .

(ii) Para cada $x, y \in S$, el mapeo $a \mapsto p_{x,y}(a)$ y $a \mapsto C(x, a)$ son continuos en $a \in A(x)$.

(iii) La función de costo es acotada, i.e., $C \in \mathcal{B}(\mathbb{K})$.

Políticas. Una política de control es una regla para elegir acciones, la cual en cada tiempo de decisión $n \in \mathbb{N}$ podría depender del estado actual, así como de los estados y acciones anteriores. Formalmente, para cada

$n \in \mathbb{N}$ defina el espacio \mathcal{H}_n de posibles historias hasta el tiempo n por $\mathcal{H}_0 := S$ y $\mathcal{H}_n := \mathbb{K}^n \times S$ para $n = 1, 2, 3, \dots$; un elemento genérico de \mathcal{H}_n es denotado por $h_n = (x_0, a_0, x_1, a_1, \dots, x_{n-1}, a_{n-1}, x_n)$, donde $(x_k, a_k) \in \mathbb{K}$ para $k < n$ y $x_n \in S$. Con esta notación, una política de control $\pi = \{\pi_n\}$ es una sucesión de kernels estocásticos π_n sobre A dado \mathcal{H}_n , que satisfacen $\pi_n(A(x_n)|h_n) = 1$, para cada $h_n \in \mathcal{H}_n$ y $n \in \mathbb{N}$. La familia de todas las políticas es denotada por \mathcal{P} . Luego, establezca $\mathbb{F} := \prod_{x \in S} A(x)$, el cual es un espacio métrico compacto, por la Suposición 3.1.1, y consiste de todas las funciones $f : S \rightarrow A$ que satisfacen $f(x) \in A(x)$ para cada $x \in S$. Una política $\pi \in \mathcal{P}$ es estacionaria si existe $f \in \mathbb{F}$ tal que la igualdad $\pi_n(\{f(x_n)\}|h_n) = 1$ siempre es válida: la clase de políticas estacionarias es naturalmente identificada con \mathbb{F} , una convención que permite escribir $\mathbb{F} \subset \mathcal{P}$. Dado un estado inicial $X_0 = x$ y la política $\pi \in \mathcal{P}$ usada para manejar el sistema, la distribución del proceso estado-acción $\{(X_t, A_t)\}_{t \in \mathbb{N}}$ está determinado únicamente y es denotado por P_x^π [1, 35, 51], mientras que E_x^π denota el correspondiente operador esperanza. A lo largo del capítulo, se utilizará la siguiente notación, para cada $n \in \mathbb{N}$, establezca

$$H_n := (X_0, A_0, \dots, X_{n-1}, A_{n-1}, X_n) \text{ y } \mathcal{F}_n := \sigma(H_n), \quad (3.1)$$

mientras que para cada $F \subset S$ el primer tiempo de retorno al conjunto F está definido por

$$T_F := \min\{n \geq 1 \mid X_n \in F\}. \quad (3.2)$$

Cuando $F = \{x\}$ es un conjunto singular, se considera la siguiente notación

$$T_x := T_{\{x\}}. \quad (3.3)$$

Note que T_F es un tiempo de paro respecto a la filtración $\{\mathcal{F}_n\}$, es decir, $[T_F = n] \in \mathcal{F}_n$ para cada $n \in \mathbb{N}$.

Criterio promedio. En el desarrollo del trabajo, se supone que el controlador tiene un coeficiente de sensibilidad al riesgo λ que satisface $\lambda < 0$. Esto significa que el controlador evalúa un costo aleatorio Y , a través del valor esperado de $U_\lambda(Y)$, donde la función de (des-)utilidad $U_\lambda : \mathbb{R} \rightarrow (-\infty, 0)$ está definida por

$$U_\lambda(x) = -e^{\lambda x}, \quad x \in \mathbb{R}. \quad (3.4)$$

Note que $U_\lambda(\cdot)$ es estrictamente creciente y satisface la relación

$$U_\lambda(a + b) = e^{\lambda a} U_\lambda(b), \quad a, b \in \mathbb{R}. \quad (3.5)$$

Cuando el decisor se enfrenta a la posibilidad de elegir entre dos costos aleatorios C_0 y C_1 , el controlador preferirá pagar C_0 si $E[U_\lambda(C_1)] >$

CAPÍTULO 3. CADENAS DE DECISIÓN DE MARKOV PROMEDIO: CASO PROPENSO AL RIESGO

$E[U_\lambda(C_0)]$, mientras que es indiferente entre ambos costos si $E[U_\lambda(C_1)] = E[U_\lambda(C_0)]$. El $(\lambda-)$ equivalente seguro de un costo Y , es denotado por $\mathcal{E}_\lambda[Y]$ y está determinado por $U(\mathcal{E}_\lambda[Y]) = E[U_\lambda(Y)]$, de modo que el controlador es indiferente entre pagar la cantidad fija $\mathcal{E}_\lambda(Y)$ o hacer frente al costo aleatorio Y . Note que $U_\lambda(\cdot)$ es una función cóncava, así que por la desigualdad de Jensen se tiene que $\mathcal{E}_\lambda(Y) \leq E[Y]$. Ahora, observe que

$$\epsilon_\lambda[Y] = U^{-1}(E[U_\lambda(Y)]) = \frac{1}{\lambda} \log(\mathbb{E}[e^{\lambda Y}]), \quad (3.6)$$

una expresión que inmediatamente produce

$$P[|Y| \leq b] = 1 \Rightarrow |\epsilon(Y)| \leq b. \quad (3.7)$$

Luego, suponga que el controlador elige la acción $\pi \in \mathcal{P}$ iniciando en $X_0 = x \in S$. La aplicación de las primeras n acciones A_0, A_1, \dots, A_n genera el costo $\sum_{k=0}^{n-1} C(X_k, A_k)$ y por (3.6) el equivalente seguro asociado está dado por

$$J_n(\pi, x) := \frac{1}{\lambda} \log\left(\mathbb{E}_x^\pi\left[e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t)}\right]\right), \quad n = 1, 2, 3, \dots, \quad (3.8)$$

que representa un promedio de $J_n(\pi, x)/n$ por decisión. El (límite inferior λ -sensible) índice de rendimiento promedio de la política $\pi \in \mathcal{P}$ en el estado $x \in S$ bajo la política π está dado por

$$J(\pi, x) := \liminf_{n \rightarrow \infty} \frac{1}{n} J_n(\pi, x), \quad (3.9)$$

y

$$J_*(x) := \inf_{\pi \in \mathcal{P}} J(\pi, x), \quad x \in S, \quad (3.10)$$

es la correspondiente función de valor óptimo. Una política $\pi_* \in \mathcal{P}$ es $(\lambda-)$ promedio óptima si $J(\pi_*, x) = J_*(x)$ para cada $x \in S$.

Condiciones de recurrencia-comunicación. En el caso neutral al riesgo, se sabe que la *condición simultánea de Doeblin*, la cual se establece en la Suposición 2.2 (i) que aparece abajo, es suficiente para asegurar que el costo promedio óptimo es constante y es caracterizado vía una ecuación de optimalidad [1, 35, 51]. En el presente contexto sensible al riesgo, la ecuación de optimalidad promedio λ -sensible está dada por

$$U_\lambda(g + h(x)) = \inf_{a \in A(x)} \left[\sum_{y \in S} p_{x,y}(a) U_\lambda((C(x, a) + h(y))) \right], \quad x \in S, \quad (3.11)$$

donde g es un número real y $h : S \rightarrow \mathbb{R}$ es una función. Cuando esta ecuación admite una solución $(g, h(\cdot))$ y $h(\cdot)$ es un mapeo acotado, se sabe

que la función de costo λ -promedio óptima $J_*(\cdot)$ es constante e igual a g , y si $f \in \mathbb{F}$ es tal que para cada estado x , la acción $f(x)$ minimiza el término dentro del paréntesis en (3.11), entonces f es λ -promedio óptima; ver, por ejemplo, [11], [34], o [40]. Note que por (3.4) la ecuación de optimalidad anterior puede ser escrita de manera equivalente como

$$e^{\lambda g + \lambda h(x)} = \sup_{a \in A(x)} \left[e^{\lambda C(x,a)} \sum_{y \in S} p_{x,y}(a) e^{\lambda h(y)} \right], \quad x \in S. \quad (3.12)$$

En contraste con el contexto neutral al riesgo, en el presente contexto donde el controlador propenso al riesgo, las condiciones de Doeblin simultáneas no son suficientes para asegurar incluso que la función de costo promedio óptimo sea constante [11], [13]. Por esta razón, en este trabajo la condición de Doeblin simultánea será complementada con un requerimiento de comunicación.

Suposición 3.1.2. *Existe $z \in S$ tal que las propiedades (i) y (ii) que se muestran a continuación son válidas:*

(i) *[Condición de Doeblin Simultánea.] El primer tiempo de retorno T_z satisface*

$$\sup_{x \in S, f \in \mathbb{F}} E_x^f[T_z] < \infty. \quad (3.13)$$

(ii) *[Accesibilidad desde z] Bajo la acción de cualquier política estacionaria, cada estado $y \in S$ es accesible desde z , esto es*

$$P_z^f[T_y < \infty] > 0, \quad y \in S, \quad f \in \mathbb{F}. \quad (3.14)$$

Observación 3.1.3. *Teorema 4.1 en [12], las Suposiciones 3.1.1 y 3.1.2 implican las condiciones (i) y (ii):*

(i) *Para cada $y \in S$, existe una constante finita M_y tal que*

$$E_x^\pi[T_y] \leq M_y, \quad x \in S, \quad \pi \in \mathcal{P}. \quad (3.15)$$

(ii) *Si $x, y \in S$ con $x \neq y$, entonces $P_x^\pi[T_y < T_x] > 0$ para cada $\pi \in \mathcal{P}$.*

Observación 3.1.4. *La Suposición 3.1.2 es, sin duda, muy fuerte. Sin embargo, actualmente es la condición más general bajo la que está disponible una caracterización del costo promedio óptimo sensible al riesgo. El resultado en esta dirección puede verse en [12] e involucra una extensión de las relaciones de Collatz-Wielandt en la teoría de matrices positivas.*

A continuación, se presenta un ejemplo donde se cumplen las Suposiciones 3.1.1 y 3.1.2.

CAPÍTULO 3. CADENAS DE DECISIÓN DE MARKOV PROMEDIO: CASO PROPENSO AL RIESGO

Ejemplo 3.1.5. *Considere el siguiente sistema simple de colas con espacio de estados $S = \mathbb{N}$: El flujo de llegadas es $\{S_k\}_{k \in \mathbb{N}}$ es una sucesión de variables aleatorias i.i.d. con distribución Poisson con esperanza $E[S_k] = \lambda > 0$. Sean b_0, b_1 y c enteros positivos con $b_0 < b_1$, y cada estado $x \in S$ define el conjunto de acción por $A(x) = \{b_0, b_0 + 1, \dots, b_1\}$. El decisor controla las salidas, y si la acción $a \in A(x)$, el número de clientes que abandonan el sistema es $D(x, a) = \lambda + (x - c)$ si $x > c$, mientras $D(x, a) = 1 + \lambda$ cuando $1 \leq x \leq c$ y, por supuesto, $D(x, a) = 0$ en $x = 0$. El número de clientes X_t en el sistema al tiempo t satisface $X_{t+1} = X_t + S_t - D(X_t, A_t)$ y entonces*

$$E[X_{t+1} | \mathcal{F}_t] = X_t + \lambda - D(X_t, A_t) = X_t + \lambda - (\lambda + (X_t - c)) = c, \text{ si } X_t > c$$

$$E[X_{t+1} | \mathcal{F}_t] = X_t + \lambda - D(X_t, A_t) = X_t + \lambda - (\lambda + 1) = X_t - 1 \text{ si } 1 \leq X_t \leq c.$$

De estas relaciones se sigue que para cualquier política π el número esperado de transiciones para pasar de $x \in \mathbb{N}$ al estado $z = 0$ está acotado por arriba por $c + 1$. Como $X_{t+1} = S_t$, se puede llegar a cualquier estado desde $z = 0$ y entonces, dotando al sistema de una función de costo acotada, la Suposición 3.1.1 y 3.1.2 se cumplen en este contexto.

Bajo las Suposiciones 3.1.1 y 3.1.2 la función de costo promedio $J_*(\cdot)$ es constante, pero la ecuación de optimalidad (3.11) no necesariamente admite una solución; un ejemplo (no controlado) que ilustra este fenómeno se presentó en la Sección 9 de [12]. Este hecho proporciona la motivación para analizar el siguiente problema:

Obtener aproximaciones convergentes al costo promedio óptimo, así como políticas estacionarias *casi óptimas* a través de los puntos fijos de operadores contractivos.

La respuesta a este problema permite determinar aproximaciones al costo promedio óptimo así como una política estacionaria cuyo costo promedio es *cercano* al óptimo resolviendo la única ecuación que caracteriza el punto fijo de un operador contractivo. El resultado principal sobre el problema anterior se establece en la siguiente sección y representa una extensión del clásico *enfoque descontado* en el caso neutral al riesgo [35, 51] al presente contexto propenso al riesgo. En el resto del capítulo, incluso sin referencia explícita, se cumplen los Supuestos 3.1.1 y 3.1.2.

3.2. Aproximaciones contractivas

En esta sección, el resultado principal del artículo se establece en el Teorema 3.2.1. Para empezar, para cada $\alpha \in (0, 1)$ defina $T_\alpha : \mathcal{B}(S) \rightarrow$

$\mathcal{B}(S)$ como sigue: para cada $W \in \mathcal{B}(S)$, $T_\alpha[W]$ está implícitamente determinado por

$$U_\lambda(T_\alpha[W](x)) = \inf_{a \in A(x)} \left[\sum_{y \in S} p_{x,y}(a) U_\lambda(C(x, a) + \alpha W(y)) \right], \quad x \in S, \quad (3.16)$$

una expresión que por la expresión (3.4) conduce a

$$T_\alpha[W](x) = \frac{1}{\lambda} \log \left(\sup_{a \in A(x)} \left[e^{\lambda C(x,a)} \sum_{y \in S} p_{x,y}(a) e^{\lambda \alpha W(y)} \right] \right), \quad x \in S. \quad (3.17)$$

Utilizando la expresión (3.7) se consigue que $\|T_\alpha[W]\| \leq \|C\| + \alpha\|W\|$, así que T_α mapea $\mathcal{B}(S)$ en él mismo. Además, no es difícil verificar que T_α es un operador monótono y α -homogéneo, i.e., para cada $W, V \in \mathcal{B}(S)$ se tiene que

$$W \geq V \implies T_\alpha[W] \geq T_\alpha[V] \text{ y } T_\alpha[V+c] = T_\alpha[V] + \alpha c, \quad c \in \mathbb{R}. \quad (3.18)$$

Observe que $V \leq W + \|W - V\|$, entonces estas propiedades implican que $T_\alpha[V] \leq T_\alpha[W + \|W - V\|] = T_\alpha[W] + \alpha\|W - V\|$, e intercambiando los roles de V y W se obtiene que

$$\|T_\alpha[W] - T_\alpha[V]\| \leq \alpha\|W - V\|, \quad W, V \in \mathcal{B}(S), \quad (3.19)$$

así que T_α es un operador contractivo sobre $\mathcal{B}(S)$. Como $\mathcal{B}(S)$ dotado con la norma del supremo es un espacio de Banach, existe un único $V_\alpha \in \mathcal{B}(S)$ que satisfacen

$$V_\alpha = T_\alpha[V_\alpha], \quad (3.20)$$

una ecuación que, por (3.17), es equivalente a

$$e^{\lambda V_\alpha(x)} = \sup_{a \in A(x)} \left[e^{\lambda C(x,a)} \sum_{y \in S} p_{x,y}(a) e^{\lambda \alpha V_\alpha(y)} \right], \quad x \in S. \quad (3.21)$$

Adicionalmente, de la Suposición 3.1.1 no es difícil ver que existe $f_\alpha \in \mathbb{F}$ tal que, para cada $x \in S$, la acción $f_\alpha(x)$ maximiza el término en paréntesis en la expresión anterior, así que

$$e^{\lambda V_\alpha(x)} = e^{\lambda C(x, f_\alpha(x))} \sum_{y \in S} p_{x,y}(f_\alpha(x)) e^{\lambda \alpha V_\alpha(y)}, \quad x \in S. \quad (3.22)$$

El (α) -costo normalizado y las funciones (α) -relativas están definidas, respectivamente, por

$$g_\alpha(x) = (1 - \alpha)V_\alpha(x), \quad h_\alpha(x) = \alpha[V_\alpha(x) - V_\alpha(w)], \quad x \in S, \quad (3.23)$$

CAPÍTULO 3. CADENAS DE DECISIÓN DE MARKOV PROMEDIO: CASO PROPENSO AL RIESGO

donde, desde este punto en adelante, $w \in S$ es un estado arbitrario pero fijo. Cálculos directos que combinan estas definiciones con las dos expresiones anteriores dan como resultado

$$e^{\lambda g_\alpha(x) + \lambda h_\alpha(x)} = \sup_{a \in A(x)} \left[e^{\lambda C(x,a)} \sum_{y \in S} p_{x,y}(a) e^{\lambda h_\alpha(y)} \right], \quad x \in S, \quad (3.24)$$

y

$$e^{\lambda g_\alpha(x) + \lambda h_\alpha(x)} = e^{\lambda C(x, f_\alpha(x))} \sum_{y \in S} p_{x,y}(f_\alpha(x)) e^{\lambda h_\alpha(y)}, \quad x \in S. \quad (3.25)$$

Note que

$$\|V_\alpha - T_\alpha[0]\| = \|T_\alpha[V_\alpha] - T_\alpha[0]\| \leq \alpha \|V_\alpha - 0\| = \alpha \|V_\alpha\|.$$

Entonces, observando que

$$\|T_\alpha[0]\| \leq \|C\|,$$

por (3.17), se sigue que

$$\|V_\alpha\| - \|C\| \leq \|V_\alpha\| - \|T_\alpha[0]\| \leq \|V_\alpha - T_\alpha[0]\| \leq \alpha \|V_\alpha\|,$$

así que

$$\|g_\alpha\| = (1 - \alpha) \|V_\alpha\| \leq \|C\|. \quad (3.26)$$

El siguiente teorema es el resultado principal de este capítulo.

Teorema 3.2.1. *Sea $\lambda < 0$ arbitrario, pero fijo. Bajo las Suposiciones 3.1.1 y 3.1.2. Las siguientes afirmaciones son válidas.*

- (i) *El costo promedio óptimo es constante, digamos g^* , y $\lim_{\alpha \nearrow 1} g_\alpha(x) = g^* = J_*(x)$ para cada $x \in S$.*
- (ii) *Dado $\epsilon > 0$, para cada $x \in S$ existe $\alpha_{x,\epsilon} \in (0, 1)$ tal que la política f_α en (3.22) es ϵ -óptima en x para $\alpha \in (\alpha_{x,\epsilon}, 1)$, esto es,*

$$\alpha \in (\alpha_{x,\epsilon}, 1) \implies g^* + \epsilon \geq J(f_\alpha, x). \quad (3.27)$$

La demostración del Teorema 3.2.1 será presentada en la Sección 3.4, después de los resultados establecidos en la siguiente sección.

3.3. Herramientas auxiliares

En esta sección, se analizan los instrumentos técnicos básicos que se usarán para demostrar el Teorema 3.2.1. Estos preliminares se establecen en los Lemas 3.3.1-3.3.3 que siguen. El primero está relacionado con las propiedades de acotación de la familia de funciones de costo relativas presentadas en (3.8).

Lema 3.3.1. (i) Para cada $\alpha \in (0, 1)$,

$$h_\alpha(\cdot) \leq 2\|C\|M_w, \quad (3.28)$$

donde la constante finita M_w es como en (3.15).

(ii) Para cada $x \in S$, $\liminf_{\alpha \nearrow 1} h_\alpha(x) > -\infty$.

Demostración. (i) Dado $\alpha \in (0, 1)$, defina la sucesión $\{Y_n\}$ de variables aleatorias por $Y_0 = e^{\lambda h_\alpha(X_0)}$ y $Y_n = e^{\lambda \sum_{t=0}^{n-1} (C(X_t, A_t) - g_\alpha(X_t)) + \lambda h_\alpha(X_n)}$ para $n \geq 1$. Ahora, sea $x \in S$ un estado fijo, y observe que (3.25) implica que para cada $n \in \mathbb{N}$

$$\begin{aligned} e^{\lambda h_\alpha(X_n)} &= e^{\lambda (C(X_n, f_\alpha(X_n)) - g_\alpha(X_n))} \sum_{y \in S} p_{X_n, y}(f_\alpha(X_n)) e^{\lambda h_\alpha(y)} \\ &= E_x^{f_\alpha} \left[e^{\lambda (C(X_n, A_n) - g_\alpha(X_n)) + \lambda h_\alpha(X_{n+1})} \middle| \mathcal{F}_n \right], \quad P_x^{f_\alpha}\text{-c. s.}, \end{aligned} \quad (3.29)$$

donde, se usó que la relación $P_x^{f_\alpha}[A_t = f_\alpha(X_t)] = 1$ siempre es válida, la segunda igualdad es debido a la propiedad de Markov. Observe que $e^{\lambda \sum_{t=0}^{n-1} (C(X_t, A_t) - g_\alpha(X_t))}$ es \mathcal{F}_n -medible, por (3.1), la expresión anterior produce

$$\begin{aligned} Y_n &= e^{\lambda \sum_{t=0}^{n-1} (C(X_t, A_t) - g_\alpha(X_t)) + \lambda h_\alpha(X_n)} \\ &= e^{\lambda \sum_{t=0}^{n-1} (C(X_t, A_t) - g_\alpha(X_t))} E_x^{f_\alpha} \left[e^{\lambda (C(X_n, A_n) - g_\alpha(X_n)) + \lambda h_\alpha(X_{n+1})} \middle| \mathcal{F}_n \right] \\ &= E_x^{f_\alpha} \left[e^{\lambda \sum_{t=0}^n (C(X_t, A_t) - g_\alpha(X_t)) + \lambda h_\alpha(X_{n+1})} \middle| \mathcal{F}_n \right] = E_x^{f_\alpha} [Y_{n+1} | \mathcal{F}_n], \end{aligned}$$

así que $\{(Y_n, \mathcal{F}_n)\}$ es una martingala con respecto a $P_x^{f_\alpha}$. Como $P_x^{f_\alpha}[X_0 = x] = 1$, el Teorema de muestreo opcional concluye que, para cada estado inicial $x \in S$ y $n \in \mathbb{N}$

$$\begin{aligned} e^{\lambda h_\alpha(x)} &= E_x^{f_\alpha}[Y_0] \\ &= E_x^{f_\alpha}[Y_{n \wedge T_w}] = E_x^{f_\alpha} \left[e^{\lambda \sum_{t=0}^{n \wedge T_w - 1} (C(X_t, A_t) - g_\alpha(X_t)) + \lambda h_\alpha(X_{n \wedge T_w})} \right]. \end{aligned}$$

Ahora, observe que por (3.2) y (3.3), $h_\alpha(X_{T_w}) = h_\alpha(w) = 0$ sobre el evento $[T_w < \infty]$. Como $P_x^{f_\alpha}[T_w < \infty] = 1$, por (3.15), se sigue que

$$\begin{aligned} &\lim_{n \rightarrow \infty} e^{\lambda \sum_{t=0}^{n \wedge T_w - 1} (C(X_t, A_t) - g_\alpha(X_t)) + \lambda h_\alpha(X_{n \wedge T_w})} \\ &= e^{\lambda \sum_{t=0}^{T_w - 1} (C(X_t, A_t) - g_\alpha(X_t)) + \lambda h_\alpha(X_{T_w})} \\ &= e^{\lambda \sum_{t=0}^{T_w - 1} (C(X_t, A_t) - g_\alpha(X_t))}, \quad P_x^{f_\alpha}\text{-c. s.} \end{aligned}$$

CAPÍTULO 3. CADENAS DE DECISIÓN DE MARKOV PROMEDIO: CASO PROPENSO AL RIESGO

Por el Lema de Fatou y por la desigualdad de Jensen, las últimas dos expresiones implican que

$$\begin{aligned}
 e^{\lambda h_\alpha(x)} &= \liminf_{n \rightarrow \infty} E_x^{f_\alpha} \left[e^{\lambda \sum_{t=0}^{n \wedge T_w - 1} (C(X_t, A_t) - g_\alpha(X_t)) + h_\alpha(X_{n \wedge T_w})} \right] \\
 &\geq E_x^{f_\alpha} \left[e^{\lambda \sum_{t=0}^{T_w - 1} (C(X_t, A_t) - g_\alpha(X_t))} \right] \\
 &\geq e^{E_x^{f_\alpha} [\lambda \sum_{t=0}^{T_w - 1} (C(X_t, A_t) - g_\alpha(X_t))]} \\
 &\geq e^{E_x^{f_\alpha} [-\sum_{t=0}^{T_w - 1} |\lambda(C(X_t, A_t) - g_\alpha(X_t))|]} \geq e^{2\lambda \|C\| E_x^{f_\alpha} [T_w]},
 \end{aligned}$$

donde (3.26) y la negatividad de λ se usaron en el último paso. Entonces se sigue que $\lambda h_\alpha(x) \geq 2\lambda \|C\| E_x^{f_\alpha} [T_w]$, así que $h_\alpha(x) \leq 2\|C\| E_x^{f_\alpha} [T_w]$. Dado que x fue un estado arbitrario, (3.28) se obtiene por (3.15).

(ii) Sea $\tilde{f} \in \mathbb{F}$ fijo, y defina la sucesión $\{S_k\}$ de subconjuntos del espacio de estados S por

$$\begin{aligned}
 S_0 &:= \{w\}, \\
 S_k &:= \{y \in S : p_{x,y}(\tilde{f}(x)) > 0 \text{ para algún } x \in S_{k-1}\}, \quad k = 1, 2, 3, \dots
 \end{aligned}$$

Note que por la Observación 3.1.3 (ii) se tiene que $\bigcup_{k=0}^{\infty} S_k = S$. Por lo tanto, para establecer la parte (ii) es suficiente demostrar que, para cada $k \in \mathbb{N}$,

$$\liminf_{\alpha \nearrow 1} h_\alpha(x) > -\infty, \quad x \in S_k. \tag{3.30}$$

Tal afirmación será verificada por inducción. Sea $\tilde{f} \in \mathbb{F}$ una política fija y note que (3.24) implica que

$$\begin{aligned}
 e^{\lambda h_\alpha(x)} &\geq e^{\lambda C(x, \tilde{f}(x)) - g_\alpha(x)} \sum_{y \in S} p_{x,y}(\tilde{f}(x)) e^{\lambda h_\alpha(y)} \\
 &\geq e^{2\lambda \|C\|} \sum_{y \in S} p_{x,y}(\tilde{f}(x)) e^{\lambda h_\alpha(y)},
 \end{aligned} \tag{3.31}$$

donde la segunda desigualdad es debido a (3.26) y la negatividad de λ . Ahora, observe que (3.30) es claramente válida para $k = 0$ ya que $S_0 = \{w\}$ y $h_\alpha(w) = 0$ para cada $\alpha \in (0, 1)$. Luego, suponga que (3.30) es válida para algún $k \in \mathbb{N}$ y sea $\tilde{y} \in S_{k+1}$ arbitrario. Elija $\tilde{x} \in S_k$ tal que

$$p_{\tilde{x}, \tilde{y}}(\tilde{f}(\tilde{x})) > 0$$

y note que (3.31) implica que $e^{\lambda h_\alpha(\tilde{x})} \geq e^{2\lambda \|C\|} p_{\tilde{x}, \tilde{y}}(\tilde{f}(\tilde{x})) e^{\lambda h_\alpha(\tilde{y})}$, así que

$$h_\alpha(\tilde{x}) \leq 2\|C\| + \frac{1}{\lambda} \log(p_{\tilde{x}, \tilde{y}}(\tilde{f}(\tilde{x}))) + h_\alpha(\tilde{y}).$$

3.3. HERRAMIENTAS AUXILIARES

Como $\tilde{x} \in S_k$, la hipótesis de inducción produce que $\liminf_{\alpha \nearrow 1} h_\alpha(\tilde{x}) > -\infty$, y entonces las dos últimas expresiones implican que

$$\liminf_{\alpha \nearrow 1} h_\alpha(\tilde{y}) > -\infty.$$

Recordando que $\tilde{y} \in S_{k+1}$ es arbitrario, se sigue que (3.30) es válida para $k+1$. \square

En el desarrollo subsecuente $\{\alpha_n\} \subset (0, 1)$ es una sucesión fija tal que

$$\alpha_n \nearrow 1 \text{ cuando } n \rightarrow \infty. \quad (3.32)$$

Sin pérdida de generalidad tomar una subsucesión -si es necesario- se supone que el siguiente límite existe

$$g(x) := \lim_{n \rightarrow \infty} g_{\alpha_n}(x), \quad h^*(x) := \lim_{n \rightarrow \infty} h_{\alpha_n}(x), \quad x \in S, \quad (3.33)$$

donde, para cada $x \in S$,

$$g(x) \in [-\|C\|, \|C\|], \quad h^*(x) \in (-\infty, 2\|C\|M_w]; \quad (3.34)$$

ver (3.26) y el Lema 3.3.1. El siguiente lema establece propiedades fundamentales de los mapeos $g(\cdot)$ y $h^*(\cdot)$.

Lema 3.3.2. *Con la notación en (3.32)–(3.34) las siguientes afirmaciones son válidas.*

(i) *El mapeo $g(\cdot)$ en (3.33) es constante, digamos $g(x) = g^* \in \mathbb{R}$ para cada $x \in S$.*

(ii) *Para cada $x \in S$,*

$$e^{\lambda g^* + \lambda h^*(x)} \geq \sup_{a \in A(x)} \left[e^{\lambda C(x,a)} \sum_{y \in S} p_{x,y}(a) e^{\lambda h^*(y)} \right].$$

(iii) *Para cada entero positivo n ,*

$$ng^* + h^*(x) - 2\|C\|M_w \leq J_n(\pi, x), \quad x \in S, \quad \pi \in \mathcal{P}.$$

(iv) *$g^* \leq J_*(\cdot)$.*

Demostración. (i) Note que (3.18) produce que

$$g_{\alpha_n}(x) - g_{\alpha_n}(w) = \frac{1 - \alpha_n}{\alpha_n} h_{\alpha_n}(x),$$

para cada $x \in S$. Tomando el límite cuando n tiende a ∞ , (3.33) y (3.34) juntos producen que $g(x) = g(w)$ para cada $x \in S$.

(ii) Sea $(x, a) \in \mathbb{K}$ arbitrario y note que (3.23) implica que, para cada $n \in \mathbb{N}$,

$$e^{\lambda g_{\alpha_n}(x) + \lambda h_{\alpha_n}(x)} \geq e^{\lambda C(x,a)} \sum_{y \in S} p_{x,y}(a) e^{\lambda h_{\alpha_n}(y)}.$$

CAPÍTULO 3. CADENAS DE DECISIÓN DE MARKOV PROMEDIO: CASO PROPENSO AL RIESGO

Tomando el límite inferior cuando n tiende a ∞ en ambos lados de la desigualdad, (3.21) y la parte (i) juntos implican que

$$\begin{aligned} e^{\lambda g^* + \lambda h^*(x)} &\geq \liminf_{n \rightarrow \infty} e^{\lambda C(x,a)} \sum_{y \in S} p_{x,y}(a) e^{\lambda h_{\alpha_n}(y)} \\ &\geq e^{\lambda C(x,a)} \sum_{y \in S} p_{x,y}(a) \liminf_{n \rightarrow \infty} e^{\lambda h_{\alpha_n}(y)}, \end{aligned}$$

donde el Lema de Fatou se usó en la segunda desigualdad. Por lo tanto, (3.21) y la expresión anterior implican que

$$e^{\lambda g^* + \lambda h^*(x)} \geq e^{\lambda C(x,a)} \sum_{y \in S} p_{x,y}(a) e^{\lambda h^*(y)}, \quad (x, a) \in \mathbb{K}. \quad (3.35)$$

(iii) Un argumento de inducción iniciando en (3.35) y usando la propiedad de Markov concluye que

$$e^{\lambda ng^* + \lambda h^*(x)} \geq E_x^\pi \left[e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t) + \lambda h^*(X_{n+1})} \right],$$

para cada estado $x \in S$, $\pi \in \mathcal{P}$ y $n \in \mathbb{N} \setminus \{0\}$. De la relación anterior y como $\lambda < 0$, usando (3.34) se sigue que

$$e^{\lambda ng^* + \lambda h^*(x)} \geq E_x^\pi \left[e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t) + 2\lambda \|C\| M_w} \right] = e^{\lambda J_n(\pi, x) + 2\lambda \|C\| M_w},$$

donde (3.8) se usó para establecer la igualdad. Por lo tanto, $\lambda ng^* + \lambda h^*(x) \geq \lambda J_n(\pi, x) + 2\lambda \|C\| M_w$, y la conclusión se sigue, ya que λ es negativo.

(iv) Dividiendo por n en ambos lados de (3.29) y tomando el límite inferior cuando $n \nearrow \infty$ en la desigualdad resultante, (3.9) produce que $g^* \leq J(\pi, x)$ para cada $x \in S$ y $\pi \in \mathcal{P}$. Desde este punto, (3.10) dirige a $g^* \leq J_*(\cdot)$. \square

El siguiente resultado es el paso final antes de proceder a la demostración del teorema principal.

Lema 3.3.3. *Dado $\alpha \in (0, 1)$, sea $f_\alpha \in \mathbb{F}$ una política tal que (3.22) es válida.*

(i) *Para cada $x \in S$,*

$$g_\alpha(x) \geq (1 - \alpha)^2 \sum_{k=1}^{\infty} \alpha^{k-1} J_k(f_\alpha, x).$$

(ii) *Dado $\epsilon > 0$ y $x \in S$, existe $\tilde{\alpha}_{x,\epsilon} \in (0, 1)$ tal que*

$$g_\alpha + \epsilon/2 \geq J(f_\alpha, x), \quad \alpha \in (\tilde{\alpha}_{x,\epsilon}, 1).$$

(iii) *$g^* \geq J_*(\cdot)$.*

3.3. HERRAMIENTAS AUXILIARES

Demostración. (i) Sea $x \in S$ arbitrario, pero fijo. Siguiendo las ideas en [14], se probará por inducción que para cada entero positivo n

$$e^{\lambda V_\alpha(x)} \leq E_x^{f_\alpha} \left[e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t) + \lambda V_\alpha(X_n)} \right]^{\alpha^n} \prod_{k=1}^n e^{\lambda(1-\alpha)\alpha^{k-1} J_k(f_\alpha, x)}. \quad (3.36)$$

Para empezar, recuerde que la igualdad $P_x^{f_\alpha}[A_t = f_\alpha(X_t)] = 1$ siempre es válida, así que la propiedad de Markov y (3.22) producen que, para cada $x \in S$ y $n \in \mathbb{N}$,

$$e^{\lambda V_\alpha(X_n)} = E_x^{f_\alpha} \left[e^{\lambda C(X_n, A_n) + \lambda V_\alpha(X_{n+1})} \middle| \mathcal{F}_n \right], \quad P_x^{f_\alpha}\text{-c. s.}$$

Estableciendo $n = 0$ en esta relación y utilizando que $P_x^{f_\alpha}[X_0 = x]$, se sigue que

$$\begin{aligned} e^{\lambda V_\alpha(x)} &= E_x^{f_\alpha} \left[e^{\lambda C(X_0, A_0) + \lambda V_\alpha(X_1)} \right] \\ &= E_x^{f_\alpha} \left[\left(e^{\lambda C(X_0, A_0) + \lambda V_\alpha(X_1)} \right)^\alpha \left(e^{\lambda C(X_0, A_0)} \right)^{1-\alpha} \right] \\ &\leq E_x^{f_\alpha} \left[e^{\lambda C(X_0, A_0) + \lambda V_\alpha(X_1)} \right]^\alpha E_x^{f_\alpha} \left[e^{\lambda C(X_0, A_0)} \right]^{1-\alpha} \\ &= E_x^{f_\alpha} \left[e^{\lambda C(X_0, A_0) + \lambda V_\alpha(X_1)} \right]^\alpha e^{\lambda J_1(f_\alpha, x)(1-\alpha)}, \end{aligned}$$

donde la desigualdad de Hölder se aplicó en el tercer paso, y la última igualdad es debido a (3.8). Esto demuestra que (3.36) es válida para $n = 1$. Luego, suponga que (3.36) es válida para un entero positivo $n \geq 1$. Observe que la igualdad $A_t = f_\alpha(X_t)$ siempre es válida con probabilidad uno bajo f_α . Lo anterior y aprovechando que $\sum_{t=0}^{n-1} C(X_t, A_t)$ es \mathcal{F}_n -medible, por (3.1), por la propiedad de Markov se obtiene que

$$\begin{aligned} E_x^{f_\alpha} \left[e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t) + \lambda V_\alpha(X_n)} \middle| \mathcal{F}_n \right] &= e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t)} e^{\lambda V_\alpha(X_n)} \\ &= e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t)} E_x^{f_\alpha} \left[e^{\lambda C(X_n, A_n) + \lambda V_\alpha(X_{n+1})} \middle| \mathcal{F}_n \right] \\ &= E_x^{f_\alpha} \left[e^{\lambda \sum_{t=0}^n C(X_t, A_t) + \lambda V_\alpha(X_{n+1})} \middle| \mathcal{F}_n \right]. \end{aligned}$$

CAPÍTULO 3. CADENAS DE DECISIÓN DE MARKOV PROMEDIO: CASO PROPENSO AL RIESGO

Por lo tanto, de la desigualdad de Hölder y (3.8) se consigue que

$$\begin{aligned}
& E_x^{f_\alpha} \left[e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t) + \lambda V_\alpha(X_n)} \right] \\
&= E_x^{f_\alpha} \left[e^{\lambda \sum_{t=0}^n C(X_t, A_t) + \lambda \alpha V_\alpha(X_{n+1})} \right] \\
&= E_x^{f_\alpha} \left[\left(e^{\lambda \sum_{t=0}^n C(X_t, A_t) + \lambda V_\alpha(X_{n+1})} \right)^\alpha \left(e^{\lambda \sum_{t=0}^n C(X_t, A_t)} \right)^{(1-\alpha)} \right] \\
&\leq E_x^{f_\alpha} \left[e^{\lambda \sum_{t=0}^n C(X_t, A_t) + \lambda V_\alpha(X_{n+1})} \right]^\alpha E_x^{f_\alpha} \left[e^{\lambda \sum_{t=0}^n C(X_t, A_t)} \right]^{(1-\alpha)} \\
&= E_x^{f_\alpha} \left[e^{\lambda \sum_{t=0}^n C(X_t, A_t) + \lambda V_\alpha(X_{n+1})} \right]^\alpha \left(e^{\lambda J_{n+1}(f_\alpha, x)} \right)^{(1-\alpha)}.
\end{aligned}$$

Entonces

$$\begin{aligned}
& E_x^{f_\alpha} \left[e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t) + \lambda V_\alpha(X_n)} \right]^{\alpha^n} \\
&\leq E_x^{f_\alpha} \left[e^{\lambda \sum_{t=0}^n C(X_t, A_t) + \lambda V_\alpha(X_{n+1})} \right]^{\alpha^{n+1}} \left(e^{\lambda J_{n+1}(f_\alpha, x)} \right)^{(1-\alpha)\alpha^n} \\
&= E_x^{f_\alpha} \left[e^{\lambda \sum_{t=0}^n C(X_t, A_t) + \lambda V_\alpha(X_{n+1})} \right]^{\alpha^{n+1}} e^{\lambda(1-\alpha)\alpha^n J_{n+1}(f_\alpha, x)}.
\end{aligned}$$

Combinando esta relación con la hipótesis de inducción, se obtiene que (3.36) es válida para $n + 1$. Ahora, para establecer la parte (i) note que para $n = 1, 2, 3, \dots$ se tiene que

$$\left| \sum_{t=0}^{n-1} C(X_t, A_t) + V_\alpha(X_n) \right| \leq n\|C\| + \|V_\alpha(\cdot)\| \leq \|C\|(n + (1 - \alpha)^{-1}).$$

De este modo, $E_x^{f_\alpha} \left[e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t) + \lambda V_\alpha(X_n)} \right] \leq e^{|\lambda|\|C\|(n + (1 - \alpha)^{-1})}$, y por la expresión (3.36) se sigue que

$$e^{\lambda V_\alpha(x)} \leq e^{\alpha^n |\lambda| \|C\| (n + (1 - \alpha)^{-1})} \prod_{k=1}^n e^{\lambda(1 - \alpha) \alpha^{k-1} J_k(f_\alpha, x)},$$

una desigualdad que, recordando que $\lambda < 0$, es equivalente a

$$V_\alpha(x) \geq -\alpha^n \|C\| (n + (1 - \alpha)^{-1}) + \sum_{k=1}^n (1 - \alpha) \alpha^{k-1} J_k(f_\alpha, x).$$

Multiplicando por $(1 - \alpha)$ en ambos lados de esta relación y por (3.23) se obtiene que

$$g_\alpha(x) \geq -\alpha^n (1 - \alpha) \|C\| (n + (1 - \alpha)^{-1}) + \sum_{k=1}^n (1 - \alpha)^2 \alpha^{k-1} J_k(f_\alpha, x).$$

3.4. DEMOSTRACIÓN DEL RESULTADO PRINCIPAL

La conclusión deseada se obtiene tomando el límite cuando n tiende a ∞ .

(ii) Sean $x \in S$ y $\epsilon > 0$ arbitrarios. Por (3.9), elija $N_0(x, \epsilon) \in \mathbb{N}$ tal que

$$\frac{1}{k} J_k(f_\alpha, x) \geq J(f_\alpha, x) - \epsilon/4, \quad k \geq N_0(x, \epsilon).$$

Como $|J(f_\alpha, x)|, k^{-1}|J_k(f_\alpha, x)| \leq \|C\|$ y por la parte (i) se consigue que

$$\begin{aligned} g_\alpha(x) &\geq (1-\alpha)^2 \sum_{k=1}^{\infty} k \alpha^{k-1} \frac{J_k(f_\alpha, x)}{k} \\ &= J(f_\alpha, x) + (1-\alpha)^2 \sum_{k=1}^{\infty} k \alpha^{k-1} \left(\frac{1}{k} J_k(f_\alpha, x) - J(f_\alpha, x) \right) \\ &\geq J(f_\alpha, x) + (1-\alpha)^2 \sum_{k=1}^{N_0(x, \epsilon)-1} k \alpha^{k-1} \left(\frac{1}{k} J_k(f_\alpha, x) - J(f_\alpha, x) \right) - \epsilon/4 \\ &\geq J(f_\alpha, x) - 2(1-\alpha)^2 \|C\| \sum_{k=1}^{N(x_0, \epsilon)-1} k \alpha^{k-1} - \epsilon/4, \end{aligned}$$

donde la expresión anterior se usó para establecer la primera desigualdad.

Finalmente, seleccione $\tilde{\alpha}_{x, \epsilon}$ tal que $(1-\alpha)^2 \sum_{k=1}^{N(x_0, \epsilon)-1} k \alpha^{k-1} \leq \epsilon(8\|C\| + 1)^{-1}$ cuando $\alpha \in (\tilde{\alpha}_{x, \epsilon}, 1)$ para concluir que

$$g_\alpha(x) \geq J(f_\alpha, x) - \epsilon/2, \quad \alpha \in (\tilde{\alpha}_{x, \epsilon}, 1).$$

(iii) Sea $x \in S$ arbitrario. Dado $\epsilon > 0$, sea $\tilde{\alpha}_{x, \epsilon} \in (0, 1)$ como en la parte (ii) y observe que por (3.32) existe $\tilde{N}(x, \epsilon) \in \mathbb{N}$ tal que $\alpha_n > \tilde{\alpha}_{x, \epsilon}$ si $n > \tilde{N}(x, \epsilon)$. En este caso (3.30) implica que $g_{\alpha_n}(x) \geq J(f_{\alpha_n}, x) - \epsilon/2$, por lo tanto,

$$g_{\alpha_n}(x) \geq J_*(x) - \epsilon/2, \quad n > \tilde{N}(x, \epsilon).$$

Tomando el límite cuando n tiende a ∞ , esta relación conduce a $g^* \geq J_*(x) - \epsilon/2$, y la condición se sigue, ya que $\epsilon > 0$ es arbitrario. \square

3.4. Demostración del resultado principal

Después de los preliminares en la sección previa, las conclusiones principales se pueden establecer como sigue.

Demostración del Teorema 3.2.1. Sea $\{\alpha_n\}_{n \in \mathbb{N}}$ una sucesión arbitraria que satisface (3.32) y, como antes, tome una subsucesión si es necesario, sin pérdida de generalidad suponga que (3.33) es válido, así que $\lim_{k \rightarrow \infty} g_{\alpha_k}(\cdot) = g^* \in \mathbb{R}$, por el Lema 3.3.2(i).

(i) Combinando el Lema 3.3.2(iv) y el Lema 3.3.3(iii) se sigue que $J_*(\cdot) =$

CAPÍTULO 3. CADENAS DE DECISIÓN DE MARKOV PROMEDIO: CASO PROPENSO AL RIESGO

$g^* = \lim_{n \rightarrow \infty} g_{\alpha_n}(x)$ para cada $x \in S$. Por lo tanto, como la sucesión $\{\alpha_n\}$ es arbitraria y satisface (3.32), se sigue que $\lim_{\alpha \nearrow 1} g_\alpha(\cdot) = J_*(\cdot) = g^*$.

(ii) Sea $x \in S$ arbitrario, pero fijo. Dado $\epsilon > 0$, por la parte (i) seleccione $\hat{\alpha}_{x,\epsilon} \in (0, 1)$ tal que

$$g_\alpha(x) < g^* + \epsilon/2, \quad \alpha \in (\hat{\alpha}_{x,\epsilon}, 1).$$

Establezca $\alpha_{x,\epsilon} = \max\{\hat{\alpha}_{x,\epsilon}, \tilde{\alpha}_{x,\epsilon}\}$, esta última expresión y el Lema 3.3.3

(ii) producen que (3.27) es válida.

Conclusiones

Se finaliza este escrito resaltando algunas de las conclusiones encontradas en el estudio de la convergencia en PDMs y presentamos algunas propuestas de investigaciones futuras que siguen esta línea.

En el Capítulo 1, se proporcionaron condiciones bajo las cuales existe convergencia uniforme de las funciones de valor óptimo y política óptima de una familia de PDMs indexados por parámetros ϵ y δ a la función de valor óptima y política óptima de un adecuado PDM determinista cuando $\epsilon \rightarrow 0$ y $\delta \rightarrow 0$. Estos PDMs, evolucionan de acuerdo con dos ecuaciones en diferencias acopladas. La primera ecuación hace referencia a la evolución de los x -estados a través de una función F que aparece en la ecuación (1.1), mientras que la segunda ecuación está asociada a la evolución de algún parámetro del modelo por medio de una función G (ver ecuación (1.2)). Los resultados principales de este capítulo son los Teoremas 1.3.7 y 1.3.10. El Teorema 1.3.7 proporciona una cota superior para el índice de estabilidad. Por otra parte, el Teorema 1.3.10 establece la convergencia de las sucesiones $\{V_{\epsilon,\delta}\}$ y $\{f_{\epsilon,\delta}\}$ a V y f^* , respectivamente, cuando ϵ y δ tienden a cero. Finalmente, la teoría desarrollada fue ilustrada con dos ejemplos que muestran las conclusiones de los resultados principales. Una consecuencia directa del Teorema 1.3.7 es que la política óptima del problema determinista es asintóticamente óptima para el problema estocástico. La teoría desarrollada en este capítulo fue publicada en el artículo [49]. Por otro lado, los resultados presentados en el Teorema 1.3.10 permiten realizar aproximaciones para sistemas estocásticos usando el método de perturbación. Tal metodología se encuentra bien establecida en la literatura de modelos de crecimiento económico para sistemas estocásticos cuya dinámica está descrita solo por una ecuación de x -estados [43].

En el Capítulo 2 se utilizó el enfoque de PDMs para reescribir y dar solución aproximada a un modelo de crecimiento económico. Se exhibe la solución proporcionada por Q -learning y se presentan realizaciones de

CAPÍTULO 3. CADENAS DE DECISIÓN DE MARKOV PROMEDIO: CASO PROPENSO AL RIESGO

la trayectoria óptima promedio del capital. Adicionalmente, se presentó en una versión aleatorizada del modelo: se aleatorizó la tasa de depreciación del capital, con esto se logró abordar el modelo vía un sistema de ecuaciones en diferencias acopladas como las propuestas en el Capítulo 1. Con esta identificación, se logró mostrar que a medida que la desviación estándar del proceso de choque tecnológico se vuelve pequeña y cuando el ruido pequeño ϵ asociado a la tasa de depreciación del capital tiende a cero, el proceso de stock del capital converge a una auto-regresión lineal gaussiana. Esto fue posible al considerar la expresión (2.23) la cual es una función del logaritmo de la dinámica del capital y del logaritmo del parámetro con el que la tasa de depreciación del capital se desarrolla. Además, con los experimentos numéricos desarrollados, se logró ilustrar la normalidad de la expresión que establece el Teorema 2.3.1, esto mediante histogramas y pruebas de normalidad. Adicionalmente, bajo una suposición de unicidad del estado estacionario del modelo determinista, se tiene que para ruidos pequeños σ y ϵ , el proceso de stock del capital converge al estado estacionario determinista. Con los experimentos numéricos, se pudo observar que el capital promedio del sistema estocástico converge al estado estable del sistema determinista. Algunos trabajos futuros, consideran abordar teoremas centrales del límite en contextos más generales.

Finalmente, en el Capítulo 3 se estudiaron cadenas de decisión de Markov sobre espacio de estados numerable. Se asumió que el desempeño de una política de decisión se mide por el criterio promedio percibido por un controlador propenso al riesgo con sensibilidad al riesgo constante. Bajo condiciones que aseguren que el costo promedio óptimo sea constante, pero *no* que la ecuación de optimalidad admita una solución, los problemas de aproximar el costo promedio óptimo y determinar una política casi óptima se estudiaron a través de puntos fijos de una familia de operadores contractivos. Los resultados en esta dirección, los cuales están establecidos en el Teorema 3.2.1, proporcionan una extensión al marco presente del clásico *enfoque descontado* en la teoría de cadenas de decisión de Markov dotadas con un índice promedio neutral al riesgo. La teoría desarrollada en este capítulo fue publicada en el artículo [48]. Por otra parte, extender las conclusiones del Teorema 3.2.1 a contextos más generales, incluyendo costos no acotados o un espacio de estados más general, parece ser un problema interesante.

Bibliografía

- [1] Arapostathis, A., Borkar, V. S., Fernández-Gaucherand, E., Ghosh, M. K. & Marcus, S.I. (1993). *Discrete-time controlled Markov processes with average cost criterion: A survey*. SIAM J. Control Optim. 31(2), 282–334.
- [2] Barz, C. & Waldmann, K.H. (2011). *Risk-sensitive capacity control in revenue management*. Mathematical Methods of Operations Research 65(3), 565–579.
- [3] Basu, A., Ghosh, M.K. (2014). *Zero-sum risk-sensitive stochastic games on a countable state space*. Stoch. Proc. Appl. 124(1), 961–983.
- [4] Bäuerle, N. & Rieder, U. (2011). *Markov Decision Processes with Applications to Finance*. Springer, New York.
- [5] Bäuerle, N. & Rieder, U. (2014). *More risk-sensitive Markov decision processes*. Math. Oper. Res. 39(1), 105–120.
- [6] Bellman, R. (2003). *Dynamic Programming*. Dover.
- [7] Bertsekas, D. P., & Shreve, S. E. (1996). *Stochastic Optimal Control: The Discrete-Time Case*. Athena Scientific.
- [8] Borkar, V.S. & Meyn, S.P. (2002). *Risk-sensitive optimal control for Markov decision process with monotone cost*. Math. Oper. Res. 27(1), 192–209.
- [9] Boucherie, R. J., & Van Dijk, N. M. (Eds.). (2017). *Markov Decision Processes in Practice*. Springer.
- [10] Brock, W. A., & Mirman, L. J. (1972). *Optimal economic growth and uncertainty: the discounted case*. Journal of Economic Theory, 4(3), 479–513.

- [11] Cavazos-Cadena, R. (2009). *Solutions of the average cost optimality equation for finite Markov decision chains: risk-sensitive and risk-neutral criteria*. Math. Method Oper. Res. 70, 541–566.
- [12] Cavazos-Cadena, R. (2018). *Characterization of the optimal risk-sensitive average cost in denumerable Markov decision chains*. Math. Oper. Res. 43(3), 1025–1050.
- [13] Cavazos-Cadena, R. & Fernández-Gaucherand, E. (1999). *Controlled Markov chains with risk-sensitive criteria: Average cost, optimality equations, and optimal solutions*. Math. Method Oper. Res. 49, 299–324.
- [14] Cavazos-Cadena, R. & Salem-Silva, F. (2010). *The discounted method and equivalence of average criteria for risk-sensitive markov decision processes on borel spaces*. Appl. Math. Optim. 61(2), 167–190.
- [15] De la Fuente, A. (2000). *Mathematical methods and models for economists*. Cambridge University Press.
- [16] Cruz-Suárez, D., Montes-de-Oca, R., & Salem-Silva, F. (2004). *Conditions for the uniqueness of optimal policies of discounted Markov decision processes*. Mathematical Methods of Operations Research, 60(3), 415-436.
- [17] Cruz-Suárez, H., Gordienko, E., & Montes-de-Oca, R. (2009). *A note on deterministic approximation of discounted Markov decision processes*. Applied mathematics letters, 22(8), 1252-1256.
- [18] Cruz-Suárez, H., & Ilhuicatzí-Roldán, R. (2010). *Stochastic optimal control for small noise intensities: The discrete-time case*. WSEAS Transactions on Mathematics, 9(2), 120-129.
- [19] Cruz-Suárez, H., Montes-De-Oca, R., & Zacarías, G. (2011). *A consumption-investment problem modelled as a discounted Markov decision process*. Kybernetika, 47(6), 909-929.
- [20] Di Masi, G.B. & Stettner, L. (1999). *Risk-sensitive control of discrete time Markov processes with infinite horizon*. SIAM J. Control Optim. 38(1), 61–78.

- [21] Di Masi, G.B. & Stettner, L. (2000). *Infinite horizon risk sensitive control of discrete time Markov processes with small risk*. Syst. Control Lett. 40, 15–20.
- [22] Di Masi, G.B. & Stettner, L. (2007). *Infinite horizon risk sensitive control of discrete time Markov processes under minorization property*. SIAM J. Control Optim. 46(1), 231–252.
- [23] Dupuis P., Kushner H. J. (1987). *Stochastic systems with small noise, analysis and simulation; a phase locked loop example*, SIAM J. Appl. Math. 47(3), 643–661.
- [24] Feinberg, E. A., & Schwartz, A. (Eds.). (2012). *Handbook of Markov Decision Processes: Methods and Applications*. Springer Science & Business Media.
- [25] González-Hernández, J., López-Martínez, R. R., & Minjárez-Sosa, J. A. (2008). *Adaptive policies for stochastic systems under a randomized discount criterion*. Bol. Soc. Mat. Mex., 149–163.
- [26] González-Hernández, J., López-Martínez, R. R., & Minjárez-Sosa, J. A. (2009). *Approximation, estimation and control of stochastic systems under a randomized discounted cost criterion*. Kybernetika, 45(5), 737–754.
- [27] González-Hernández, J., López-Martínez, R. R., Minjárez-Sosa, J. A., & Gabriel-Arguelles, J. A. (2013). *Constrained Markov control processes with randomized discounted cost criteria: occupation measures and extremal points*. Risk and Decision Analysis, 4(3), 163–176.
- [28] González-Hernández, J., López-Martínez, R. R., & Pérez-Hernández, R. (2007). *Markov control processes with randomized discounted cost in Borel space*. Math. Meth. Oper. Res., 65(1), 27–44.
- [29] Gordienko, E.I. (1992). *An estimate of the stability of optimal control of certain stochastic and deterministic systems*. J. Soviet Math., 59(4), 891–899.
- [30] Gordienko, E.I., Lemus-Rodríguez, E., & Montes-de-Oca, R. (2008). *Discounted cost optimality problem: stability with respect to weak metrics*. Mathematical Methods of Operations Research, 68(1), 77–96.

- [31] Gordienko, E. I., & Salem, F. S. (1998). *Robustness inequality for Markov control processes with unbounded costs*. Systems & control letters, 33(2), 125-130.
- [32] Gosavi, A. (2008). *On step sizes, stochastic shortest paths, and survival probabilities in reinforcement learning*. In 2008 Winter Simulation Conference (pp. 525-531). IEEE.
- [33] Gosavi, A. (2015). *Simulation-based optimization* (Vol. 62). Berlin: Springer.
- [34] Hernández-Hernández, D. & Marcus, S.I. (1996). *Risk-sensitive control of Markov processes in countable state space*. Systems and Control Letters 29, 147–155.
- [35] Hernández-Lerma, O. (1989). *Adaptive Markov Control Processes*. Springer, New York.
- [36] Hernández-Hernández, D., & Minjárez-Sosa, J. A. (Eds.). (2012). *Optimization, Control, and Applications of Stochastic Systems*. Birkhäuser.
- [37] Hernández-Lerma, O., & Lasserre, J. B. (1996). *Discrete-Time Markov Control. Processes: Basic Optimality Criteria*. Springer-Verlag, New York.
- [38] Hernández-Lerma, O., & Lasserre, J. B. (1999). *Further Topics on Discrete-Time Markov Control Processes*. Springer-Verlag, New York.
- [39] Hinderer K. (2005). *Lipschitz continuity of value functions in Markovian decision processes*, Math. Meth. Oper. Res. 62, 3-22.
- [40] Howard, R.A. & Matheson, J.E. (1972). *Risk-sensitive Markov decision processes*. Manage. Sci. 18(7), 356–369.
- [41] Jaskiewicz, A. (2007). *Average optimality for risk sensitive control with general state space*. Ann. Appl. Probab. 17(2), 654–675.
- [42] Jaskiewicz, A. & Nowak, A.S. (2014). *Stationary Markov perfect equilibria in risk sensitive stochastic overlapping generations models*. Journal of Economic Theory 151, 411–447.

- [43] Judd. Kenneth L. (1998). *Numerical methods in economics*, MIT press.
- [44] Kara A. D., Yüksel S. (2020). *Robustness to incorrect system models in stochastic control*, SIAM J. Control Optim. 58(2), 1144-1182.
- [45] Liptser R. S., Runggaldier W. J., Taksar M. (1996). *Deterministic approximation for stochastic control problems*, SIAM J. Control Optim. 34(1), 161-178.
- [46] Miculescu R. (2003). *Approximations by Lipschitz functions generated by extensions*, Real Anal. Exchange. 28(1), 33-40.
- [47] Pitera, M. & Stettner, L. (2016). *Long run risk sensitive portfolio with general factors*. Math. Meth. Oper. Res 82(2), 265–293.
- [48] Portillo-Ramírez, G., Cavazos-Cadena, R., & Cruz-Suárez, H. (2023). *Contractive approximations in average Markov decision chains driven by a risk-seeking controller*. Mathematical Methods of Operations Research, 98(1), 75-91.
- [49] Portillo-Ramírez, G., Cruz-Suárez, H., López-Ríos, R., & Blancas-Rivera, R. (2023). *Markov decision processes approximation with coupled dynamics via Markov deterministic control systems*. Open Mathematics, 21(1), 20230129.
- [50] Portillo-Ramírez, G., Cruz-Suárez, H., & Velasco-Luna, F. (2024). Un panorama general del aprendizaje por refuerzo en líneas de espera controladas a tiempo discreto. CIENCIA ergo-sum, 32.
- [51] Puterman, M. L. (1994). *Markov Decision Processes*. Wiley Interscience.
- [52] R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- [53] Saucedo-Zul, J., Cavazos-Cadena, R. & Cruz-Suárez, H. (2020). *A discounted approach in communicating average markov decision chains under risk-aversion*. Journal of Optimization Theory and Applications 187, 585–606.

- [54] Shen, Y., Stannat, W. & Obermayer, K. (2013). *Risk-sensitive Markov control processes*. SIAM J. Control Optim. 51(5), 3652–3672.
- [55] Sladký K. (2008). *Growth rates and average optimality in risk-sensitive Markov decision chains*. Kybernetika 44(2), 205–226.
- [56] Sladký, K. (2018). *Risk-sensitive average optimality in Markov decision processes*. Kybernetika 54(6), 1218–1230.
- [57] Stettner, L. (1999). *Risk sensitive portfolio optimization*. Math. Meth. Oper. Res. 50(3), 463– 474.
- [58] Williams, N. (2004). *Small noise asymptotics for a stochastic growth model*. Journal of Economic Theory, 119(2), 271-298.