PROCESOS DE MARKOV CONTROLADOS EN ESPACIOS DE BOREL: CRITERIOS NO DESCONTADOS

TESIS QUE PRESENTA

OSCAR VEGA AMAYA

PARA OBTENER EL GRADO
DE
DOCTOR EN CIENCIAS

MARZO 1998

ASESORES: DR. RAUL MONTES DE OCA
DR. ONESIMO HERNANDEZ-LERMA

UNIVERSIDAD AUTONOMA METROPOLITANA
UNIDAD IZTAPALAPA
DIVISION DE CIENCIAS BASICAS E INGENIERIA
DEPARTAMENTO DE MATEMATICAS

Sólo podemos dar lo que ya hemos dado.

Sólo podemos dar lo que ya es de otro.

Jorge Luis Borges

Los Conjurados

Para Tere,

Teresita,

Natalia,

Sebastián Isaac,

por darle sentido a cada uno de mis esfuerzos.

A Doña Yoli, por su amor y su eterna espera.

Toda obra humana es deleznable, afirma Carlyle, pero su ejecución no lo es.
Jorge Luis Borges
Los Conjurados

Acaso la vida no es la concatenación infinita de hechos improbables ?

Para que este trabajo se iniciara y, sobre todo, se concluyera hubieron de coincidir múltiples circunstancias y voluntades. Deseo agradecer a los Doctores Onésimo Hernández Lerma y Raúl Montes de Oca por haber dirigido esta tesis, por el apoyo brindado y sus valiosos consejos. Agradezco también a los Doctores Guillermo Ferreyra, Evgueni Gordienko, Daniel Hernández, Roberto Quezada por la disponibilidad mostrada y las sugerencias hechas a este trabajo.

Mis agradecimientos también son para Mario Enrique Alvarez y Loly Bojórquez, Enrique Hugues, Juan Antonio Nido y Jesús Adolfo Minjárez, por su amistad y generosa hospitalidad brindada durante mis estancias en la Cd. de México.

También quiero manifestar mis agradecimientos a los controleros de la Uni : Fernando Luque "el sparring" Vásquez, María Teresa Robles y Jesús Adolfo Minjárez.

Finalmente quiero agradecer a los Departamentos de Matemáticas de la Universidad de Sonora y la Universidad Autónoma Metropolitana, Iztapalapa, por el apoyo y facilidades que proporcionaron para la realización de este trabajo, al Sindicato de Trabajadores Académicos de la Universidad de Sonora, y al CONACYT de quien recibí apoyo económico a través de los proyectos 1332-E9206 y 3115-E9608

Contenido

Capítulo 1

Introducción

En esta tesis se estudian procesos de Markov controlados en espacios de Borel, con costos no acotados y criterios de optimalidad no-descontados. Los resultados que se presentan son una recopilación de los artículos [38], [57], [77] y [79].

Los criterios no-descontados pueden definirse de múltiples maneras, pero nuestro estudio está dirigido hacia aquellos que proporcionan información sobre la tasa de crecimiento de los costos esperados en horizonte finito cuando éste crece arbitrariamente. Entre estos criterios, en un extremo del espectro, se incluyen el criterio en *costo promedio* (esperado) que es el menos sensible al crecimiento de los costos esperados en horizonte finito y, en el otro, la noción excesivamente selectiva de *políticas dominantes* (strong overtaking optimal policies) introducida por Ramsey [60]. Entre estos criterios se encuentran la noción de *políticas dominantes* (overtaking optimal or catching-up policies) de Gale [26] y von Weiszäcker [81], el *costo de oportunidad* (opportunity cost) y el *criterio en costo promedio de Flynn* [25], el *criterio de Dutta* [21], la *optimalidad en sesgo* (bias optimality) de Veinott [80], y la noción de *soluciones de la Ecuación de Optimalidad en Costo Promedio* o *ternas canónicas* (canonical triplets) *en la terminología de* Yushkevich [82]). En este trabajo se presentan condiciones para la existencia de políticas estacionarias óptimas para estos criterios y se discuten algunas relaciones entre los mismos. También se efectúa un análisis por trayectorias del criterio en costo promedio. La aplicabilidad de los resultados obtenidos se ilustran con problemas de control en sistemas de inventarios.

En el Capítulo 2, además de introducir los elementos básicos que se requieren para plantear un problema de control óptimo (el modelo de control, el conjunto de políticas de control admisibles, los índices de funcionamiento y criterios de optimalidad), se presenta una discusión breve, pero panorámica, de los distintos problemas de control en los que estamos interesados.

Para el criterio *en costo promedio* (esperado), sin duda alguna el más estudiado entre los criterios mencionados previamente, se han desarrollado varios enfoques como el enfoque de *Aproximaciones por Problemas Descontados*, el *Algoritmo de Iteración de Valores*, el *Algoritmo de Iteración de Políticas*, *Programación Lineal*, *Análisis Convexo*, *Enfoque Ergódico*, entre otros (ver [3], [6], [7], [10], [20], [30], [34], [59], y sus referencias). Sin embargo, la mayor parte de la literatura se concentra en el caso de *espacios discretos* o *espacios de Borel* y *costos acotados*, excepto para el enfoque de *Aproximaciones por Problemas Descontados* (APD), cuyo origen se remonta a Blackwell [9] y que en años recientes se ha extendido a contextos bastante generales (espacios de Borel, costos no acotados, condiciones de continuidad y compacidad muy débiles; ver, por ejemplo, [34], [54], [66]). En el Capítulo 3, basandonos en [79], presentamos una variante más del APD que extiende algunos trabajos previos y la cual nos permite resolver *explicitamente* el problema en costo promedio para un sistema de inventarios con controles no acotados que no se había incluido en la literatura.

Otra forma de abordar el problema en costo promedio es imponiendo condiciones que garanticen que los procesos controlados tienen buenas propiedades de estabilidad (recurrencia o ergodicidad). Uno de los procedimientos mas importantes para conseguir este tipo de propiedades es el de las *funciones de Lyapunov* ([17], [28], [36], [38] [40]). En el Capítulo 4, se introduce una condición del tipo *Lyapunov* y se demuestra, bajo condiciones usuales de continuidad y compacidad, la existencia de una solución de la Ecuación de Optimalidad en Costo Promedio y la existencia de una política estacionaria Flynn-óptima, entre otros resultados. Lo anterior se demuestra en dos etapas: en la primera, se usa el enfoque APD para garantizar la existencia de una política óptima en costo promedio en la clase de las políticas estacionarias y, en la segunda etapa, se usa el Algortimo de Iteración de Políticas para obtener una solución de la EOCP, con la característca de que la política de "inicialización" es óptima en la clase de las políticas estacionarias. Con este procedimiento se evita el uso del Teorema de Arzela-Ascoli, el cual requiere de condiciones de continuidad muy restrictvas sobre la función de costo por etapa y la ley de transición del sistema. Además, se ilustra como verificar la condición de Lyapunov con un ejemplo de inventarios. En el Capítulo 5, usando las mismas condiciones del Capítulo 4, se demuestra que existe una política óptima en el sentido de Dutta en la clase de las políticas estacionarias, y que esto es equivalente a cada una de las siguientes afirmaciones cuando el análisis se restringe a la clase de las políticas estacionarias: a) la política es dominante; b) la política es óptima en costo de oportunidad; c) la política es óptima en sesgo. Los ejemplos proporcionados en [11] y [57] muestran que los resultados anteriores *no se extienden* a la clase de todas las políticas si no se imponen condiciones adicionales a las de estabilidad. Los resultados de los Capítulos 4 y 5 se tomaron de [38] y [57].

En Capítulo 6 se retoma el estudio de los problemas en costo promedio, pero ahora se realiza un análisis por trayectorias. Los trabajos en los cuales se realiza un análisis de este tipo son escasos y se restringen al caso de espacios discretos o de Borel pero considerando costos acotados, bajo condiciones de recurrencia o ergodicidad bastante fuertes ([3], [47], [15]). En este capítulo se presentan los resultados desarrollados en [77] sobre la existencia de políticas estacionarias óptimas en costo promedio por trayectorias para el caso de espacios de Borel y costos estrictamente no-acotados, bajo hipótesis de continuidad débil y recurrencia de Harris. De nuevo, los resultados principales se ilustran con problemas de inventarios y, en un caso espécifico, se exhibe una política estacionaria óptima en costo promedio por trayectorias.

En los Apéndices se reunen algunas definiciones y resultados que se usan a lo largo de este trabajo con el ánimo de facilitar la lectura.

1.1 Notación y terminología

Para un espacio topológico $(X, \mathcal{T})$, usaremos la siguiente notación:

- $\mathcal{B}(X)$ es la $\sigma-$álgebra de Borel de X, es decir, la mínima $\sigma-$álgebra que contiene a $\mathcal{T}$, y a sus elementos les llamaremos conjuntos de Borel.
- X es un *espacio de Borel* si es un conjunto de Borel de un espacio métrico separable y completo.
- $M(X)$ es la clase de las funciones (Borel-) medibles $u : X \to \mathbf{R}$.
- $M_+(X) = \{u \in M(X) : u \geq 0\}$.

- $L(X)$ es la clase de las funciones $u : X \to \mathbf{R}$ semicontinuas inferiormente y acotadas por abajo.

- $C_a(X)$ es el espacio vectorial formado por las funciones $u : X \to \mathbf{R}$ continuas acotadas dotado con la *norma del supremo*, es decir, para cada $u \in C_a(X)$, se define $||u|| := \sup_x |u(x)|$.

- $M_a(X)$ denota la clase de las funciones $u : X \to \mathbf{R}$, medibles y acotadas.

- Si d es una métrica en X consistente con la topología $\mathcal{T}$, entonces $\mathcal{U}_d(X)$ denotará espacio de las funciones uniformemente continuas con respecto a la métrica d y la topología relativa del espacio de Banach $C_a(X)$.

Capítulo 2

Procesos de Markov Controlados

2.1 Introducción

Para la formulación de un *problema de control óptimo* (PCO) determinístico o estocástico se requieren tres elementos: (a) un modelo de control; (b) un conjunto de políticas de control admisibles; (c) un índice de funcionamiento. El objetivo en este capítulo es, por una parte, presentar estos elementos e introducir la notación y terminología que usaremos, y por otra, discutir de forma general los distintos índices de funcionamientos que serán estudiados en este trabajo.

2.2 Modelos de control markoviano

Definición 2.2.1. Un *modelo de control markoviano* (MCM) consta de los objetos $(\mathbf{X}, \mathbf{A}, \{A(x) : x \in A(x)\}, Q, C)$, donde:

(a) $\mathbf{X}$ representa el *espacio* (o conjunto) *de estados* posibles del sistema;

(b) $\mathbf{A}$ representa el *espacio* (o conjunto) *de controles* disponibles para modificar el comportamiento del sistema.

Suponemos que los conjuntos $\mathbf{X}$ y $\mathbf{A}$ son *espacios de Borel*, es decir, subconjuntos de Borel (no vacíos) de espacios métricos separables y completos.

(c) Para cada estado $x \in \mathbf{X}$, $A(x)$ es un subconjunto de $\mathbf{A}$ y representa el conjunto de *controles admisibles* para el estado x. Al conjunto

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

se le llama *conjunto de pares admisibles*; suponemos que pertenece a la σ-álgebra de Borel del producto cartesiano $\mathbf{X} \times \mathbf{A}$ y que contiene la gráfica de una funcion medible $f : \mathbf{X} \to \mathbf{A}$.

(d) Q representa la *ley de evolución* del sistema y es una probabilidad de transición sobre $\mathbf{X}$ dado $\mathbf{K}$, es decir, satisface las siguientes propiedades:

(d.1) $Q(\cdot|x, a)$ es una medida de probabilidad en $\mathbf{X}$ para cada $(x, a) \in \mathbf{K}$;

(d.2) $Q(B|\cdot, \cdot)$ es una función medible en $\mathbf{K}$ para cada $B \in B(\mathbf{X})$.

(e) C es una función medible definida en $\mathbf{K}$ y representa la función de *costo por etapa*.

Un *modelo de control markoviano* (MCM) es la representación de un sistema estocástico que es observado y sujeto a control en un conjunto discreto de tiempos, que sin pérdida de generalidad tomaremos como el conjunto de los enteros no-negativos $\mathbf{N}_0$. Denotaremos por x_t al *estado del sistema* y por a_t al *control* aplicado, en ambos casos en el tiempo t; nos referiremos a x_0 como el *estado inicial* del sistema. El *problema de control óptimo* consiste en modificar o influir sobre el comportamiento del *proceso de estados* $\{x_t\}$ por medio de selecciones adecuadas del *proceso de control* $\{a_t\}$, de manera que dicho comportamiento sea óptimo de acuerdo a uno o más criterios. Heurísticamente, podemos pensar que el proceso de control se efectúa secuencialmente de la siguiente forma. Supongamos que en el tiempo t se observa al sistema en el estado $x \in \mathbf{X}$, es decir, $x_t = x$; entonces, si se aplica el control $a_t = a \in A(x)$, se genera un costo $C(x, a)$ y el sistema visitará un nuevo estado en el tiempo $t + 1$ de acuerdo a la medida (o distribución) de probabilidad $Q(\cdot|x, a)$, es decir,

$$Q(x_{t+1} \in B|x_t = x, a_t = a) = \Pr[x_{t+1} \in B|x_t = x, a_t = a].$$

Una vez observado el nuevo estado del sistema, digamos $x_{t+1} = x' \in \mathbf{X}$, se elige un control $a_{t+1} = a' \in A(x')$ con un costo $C(x', a')$ y se repite el proceso anterior hasta cierto tiempo finito N o indefinidamente. Si el proceso de observación/control termina en un tiempo finito N, se dice que el problema de control es en *horizonte finito* y a N se le llama *horizonte de planeación*; por el contrario, si el proceso de observación/control se repite indefinidamente se habla de un problema de control en *horizonte infinito*.

2.3 Políticas de control admisibles

Considere los conjuntos

$$\mathbf{H}_0 := \mathbf{X}$$

$$\mathbf{H}_t := \mathbf{K}^t \times \mathbf{X}, \ t \in \mathbf{N} = \{1, 2, \cdots\}.$$

Observe que para cada $t \in \mathbf{N}_0$, el conjunto $\mathbf{H}_t$ contiene todas las formas posibles en que el sistema puede evolucionar hasta el tiempo t, razón por la cual le llamaremos el *espacio de las t-historias* y a sus elementos, denotados genéricamente por h_t, serán referidos como *t-historias*; más explicitamente, cada *t-historia* es un vector de la forma $h_t = (x_0, a_0, x_1, a_1, \cdots, x_{t-1}, a_{t-1}, x_t)$, donde $a_k \in A(x_k)$ para $k = 0, 1, \ldots, t-1$.

En la definición de las políticas de control admisibles se hará uso de la notación y terminología que se introduce a continuación

Definición 2.3.1(a) Por Φ denotaremos a la clase de las probabilidades de transición φ sobre $\mathbf{A}$ dado $\mathbf{X}$, que satisfacen la restricción $\varphi(A(x)|x) = 1$ para cada $x \in \mathbf{X}$;

(b) Un *selector* o *función de decisión* es una función medible $f : \mathbf{X} \to \mathbf{A}$, tal que $f(x) \in$

$A(x)$, $\forall x \in \mathbf{X}$. Denotaremos por $\mathbf{F}$ a la familia de los selectores.

Puesto que a cada selector $f \in \mathbf{F}$ le podemos asociar un elemento de Φ, a saber, $\varphi_f(B|x) := \mathrm{I}_B(f(x))$, $x \in \mathbf{X}$, $B \in \mathcal{B}(\mathbf{A})$, consideraremos que $\mathbf{F} \subset \mathbf{\Phi}$.

Definición 2.3.2(a) Una *política de control* (*admisible*) es una sucesión $\pi = \{\pi_t\}$ tal que, para cada t, π_t es una probabilidad de transición sobre $\mathbf{A}$ dado $\mathbf{H}_t$ que satisface la restricción

$$\pi_t(A(x_t)|h_t) = 1 \ \ \forall h_t \in \mathbf{H}_t.$$

(b) Una política de control $\pi = \{\pi_t\}$ se dice que es *determinista* si para cada t existe una función medible $f_t : \mathbf{H}_t \to \mathbf{A}$ tal que $\pi_t(C|h_t) = \mathrm{I}_C(f_t(h_t))$, $\forall C \in \mathcal{B}(\mathbf{A})$ y $h_t \in \mathbf{H}_t$, donde I_C es la función indicadora del conjunto C.

(c) Una política $\pi = \{\pi_t\}$ se dice que es *markoviana* si existe una sucesión $\{f_t\} \subset \mathbf{F}$, tal que para cada t, $\pi_t(\cdot|h_t)$ está concentrada en $f_t(x_t)$, es decir, $\pi_t(C|h_t) = \mathrm{I}_C(f_t(x_t)), \forall C \in \mathcal{B}(\mathbf{A}), h_t \in \mathbf{H}_t$.

La familia de todas las políticas de control admisibles será denotada por Π, mientras que las subclases de las políticas deterministas y markovianas se denotarán por Π_D y Π_M, repectivamente. De acuerdo a las definiciones anteriores se tiene que $\Pi_M \subset \Pi_D \subset \Pi$.

Para los problemas de control en horizonte infinito es importante considerar dos subclases más de políticas, a saber, las subclases constituidas por las *políticas relajadas* (o estacionarias aleatorizadas) y la clase de las *políticas estacionarias* (deterministas), que a continuación se definen.

Definición 2.3.3. Sea $\pi = \{\pi_t\} \in \Pi$ una política de control. Diremos que π es una política:

(a) *relajada* o *estacionaria aleatorizada* si existe $\varphi \in \Phi$ tal que $\pi_t(\cdot|h_t) = \varphi(\cdot|x_t)$ $\forall h_t \in \mathbf{H}_t$ y $t \in \mathbf{N}_0$;

(b) *estacionaria* (determinista) si existe $f \in \mathbf{F}$ tal que $\pi_t(\cdot|h_t)$ está concentrada en $f(x_t)$, es decir, $\pi_t(C|h_t) = \mathrm{I}_C(f(x_t))$ $\forall C \in \mathcal{B}(\mathbf{A})$, $h_t \in \mathbf{H}_t$ y $t \in \mathbf{N}_0$.

Observe que cada elemento φ en Φ define una política relajada y viceversa; de manera que abusando un poco de la notación identificaremos a φ con tal política y nos referirnos a Φ como la clase de las políticas relajadas (o estacionarias aleatorizadas). De manera similar, a la familia de los selectores $\mathbf{F}$ la identificaremos con la clase de las políticas estacionarias. Es claro que $\mathbf{F} \subset \Phi$.

Observación 2.3.4. Del Teorema de Ionescu-Tulcea ([1], Teorema 2.7.2, p.109), tenemos que para cada política $\pi \in \Pi$ y cada medida $\nu \in \mathbf{P}(\mathbf{X})$ existe una medida de probabilidad P_ν^π definida sobre el *espacio canónico* $(\Omega, \mathcal{F})$, donde $\Omega = (\mathbf{X} \times \mathbf{A})^\infty$ y $\mathcal{F}$ es la $\sigma-$álgebra producto correspondiente, que satisface las siguientes propiedades para cada $t \in \mathbf{N}_0$:

(a)$P_\nu^\pi[x_0 \in B] = \nu(B) \ \ \forall B \in \mathcal{B}(\mathbf{X})$;

(b)$P_\nu^\pi[a_t \in C|h_t] = \pi_t(C|h_t) \ \ \forall C \in \mathcal{B}(\mathbf{A})$;

(c)$P_\nu^\pi[x_{t+1} \in B|h_t, a_t] = Q(B|x_t, a_t) \ \ \forall B \in \mathcal{B}(\mathbf{X})$.

La esperanza con respecto a la medida de probabilidad P_ν^π será denotada por E_ν^π. En concordancia con la propiedad (a), nos referiremos a ν como la *distribución inicial* del proceso controlado; si la medida ν está concentrada en un estado inicial $x_0 = x \in \mathbf{X}$, es decir, $\nu(B) := \mathbf{I}_B(x), \ B \in \mathcal{B}(\mathbf{X})$, entonces en lugar de P_ν^π (E_ν^π) escribiremos P_x^π (E_x^π, respectivamente).

Cuando $\pi = \varphi \in \Phi$ es una *política relajada*, de las propiedades (a)-(c) se puede deducir que el proceso de estados $\{x_t\}$ es un *proceso de Markov* con espacios de estados $\mathbf{X}$ y probabilidad de transición dada por

$$Q(B|x, \varphi) := \int_{\mathbf{A}} Q(B|x, a)\varphi(da|x), \ \ x \in \mathbf{X}, B \in \mathcal{B}(\mathbf{X}),$$

y distribución inicial ν. Debido a esta propiedad, y por extensión, cuando se usa cualquier política $\pi \in \Pi$, se dice que el proceso $(\Omega, \mathcal{F}, P_\nu^\pi, \{x_t\})$ es un *proceso de Markov controlado* (PMC).

2.4 Indices de funcionamiento

Una vez especificados el modelo de control y el conjunto de políticas de control admisibles, para plantear un *problema de control óptimo* (PCO) sólo hace falta especificar el *índice de funcionamiento* o *función objetivo* mediante el cual se evaluarán los comportamientos posibles del sistema estocástico inducidos por la aplicación de las diferentes políticas de control. Generalmente, el índice de funcionamiento es una función

$$I : \Pi \times \mathbf{X} \to \mathbf{R},$$

y la *función de valor óptimo* correspondiente se define como

$$I^*(x) := \inf_{\pi \in \Pi} I(\pi, x), \quad x \in \mathbf{X}.$$

Entonces, el PCO consite en encontrar una política $\pi^* \in \Pi$ tal que

$$I^*(x) = I(\pi^*, x) \;\; \forall x \in \mathbf{X},$$

en cuyo caso se dice que π^* es *óptima* con respecto al índice I.

Los problemas de control óptimo pueden ser clasificados de muchas formas, las cuales generalmente dependen de aquellos aspectos del problema que se desean destacar. Por ejemplo, una primera clasificación que se hace en la literatura se refiere al tipo de espacios de estados, el cual puede ser *discreto* (es decir, un conjunto finito o infinito numerable) o un *espacio de Borel* (es decir, subconjunto de Borel de un espacio métrico separable y completo). Otra posibilidad consiste en clasificar los problemas de control de acuerdo al *horizonte de planeación*, el cual puede ser *finito* (es decir, el proceso de observación y control termina en un número finito de etapas) o *infinito* (es decir, el proceso de observación y control se repite indefinidamente). A su vez, los problemas en horizonte infinito se clasifican como *descontados* o *no descontados*, dependiendo si el índice de funcionamiento en cuestión incluye o no un *factor de descuento*, respectivamente. Para los problemas en horizonte infinito, desde el punto de vista técnico, también es importante saber si la función de costo por etapa es acotada o no, lo que permite hablar de problemas con *costos acotados* o con *costos no acotados*. En [30], [34] y [74] se proporcionan clasificaciones muy completas de los problemas de control.

Especificamente, nuestro trabajo está dirigido hacia el estudio de problemas de control óptimo en *horizonte infinito, espacios de Borel, con índices de funcionamiento no descontados y costos no acotados.*

En lo que resta de esta sección, introduciremos los índices de funcionamiento de interés y, en paralelo se discutirá brevemente la problemática asociada a dichos índices.

2.4.1 Problemas de control en horizonte finito

El *costo* (esperado) *en N-etapas,* $N \geq 1$, cuando se aplica la política $\pi \in \Pi$, dado que el estado inicial es $x_0 = x \in \mathbf{X}$, se define como

$$J_N(\pi, x) := E_x^\pi \sum_{t=0}^{N-1} C(x_t, a_t), \tag{2.1}$$

y la *función de costo óptimo en N-etapas* por

$$J_N^*(x) := \inf_{\pi \in \Pi} J_N(\pi, x) \ \ \pi \in \Pi, x \in \mathbf{X}. \tag{2.2}$$

El *problema de control óptimo* (PCO) en *N-etapas* consiste en encontrar una política $\pi^* \in \Pi$ que satisfaga

$$J_N^*(x) = J_N(\pi^*, x) \ \ \forall x \in \mathbf{X}. \tag{2.3}$$

A las políticas que satisfacen la relación (2.3) les llamaremos políticas *óptimas en N-etapas*, o brevemente, *N-óptimas*.

El *Algoritmo de Programación Dinámica* ([6], [30], [34], [59]) garantiza la existencia de políticas óptimas para el problema (2.1)-(2.3) bajo condiciones muy generales y proporciona caracterizaciones de las mismas; además, nos provee de un procedimiento secuencial para la determinación tanto de las funciones de costo óptimo como de las políticas óptimas.

Desde el punto de vista analítico el *Algoritmo de Programación Dinámica* proporciona una solución completa del PCO en horizonte finito, pero desde el punto de vista práctico surgen algunos problemas. Por ejemplo, las funciones de costo óptimo $J_N^*(\cdot)$, $N \in \mathbf{N}$, generalmente no pueden obtenerse explícitamente— lo cual dificulta la determinación de las políticas óptimas— o si es posible, se requiere de un gran esfuerzo "computacional" que se acentúa conforme crece el horizonte de planeación. Sin embargo, cuando el horizonte de planeación es suficientemente grande, es factible "aproximar" el PCO en horizonte finito por problemas en horizonte infinito. Una segunda motivación para el estudio de los problemas en horizonte infinito se debe a que en algunas aplicaciones no es posible especificar *a priori* el horizonte de planeacion y no existe una preferencia por optimizar a "corto plazo"—por ejemplo, en control en sistemas de espera o redes de comunicaciones ([22], [59], [73]).

Para los problemas descontados en horizonte infinito existe un significado muy específico: en cada etapa de control o decisión, el costo generado en el proceso de control es "actualizado" por medio de un *factor de descuento* $\alpha \in (0,1)$, y el problema de control consiste en minimizar el valor esperado de la suma de los costos descontados sobre la clase de todas las políticas admisibles. Esto es, el *costo α-descontado* generado por el uso de la politica $\pi \in \Pi$, dado que el estado inicial es $x_0 = x \in \mathbf{X}$, está dado por

$$V_\alpha(\pi, x) := E_x^\pi \sum_{t=0}^{\infty} \alpha^t C(x_t, a_t) \quad \pi \in \Pi, x \in \mathbf{X}, \tag{2.4}$$

mientras que la función

$$V_\alpha(x) := \inf_{\pi \in \Pi} V_\alpha(\pi, x) \quad x \in \mathbf{X}, \tag{2.5}$$

representa el *costo óptimo α-descontado*. Una política $\pi^* \in \Pi$ es *α-óptima* si satisface la relación

$$V_\alpha(x) = V_\alpha(\pi^*, x) \quad \forall x \in \mathbf{X}. \tag{2.6}$$

Entre los problemas de control, el PCO en costo descontado (2.4)-(2.6) es el mejor comprendido y para el cual existe un cuerpo teórico muy completo, el cual incluye caracterizaciones de las políticas estacionarias óptimas, algoritmos de aproximación—e.g., iteración de valores, iteración de políticas—que ya es de uso estándar en la literatura (ver por ejemplo, [6], [7], [20], [30], [34], [59], [63]).

Sin embargo, para el caso no-descontado la situación es muy distinta. Para comenzar, el adjetivo no-descontado puede tener significados muy diferentes. Por ejemplo, una posibilidad sería tomar $\alpha = 1$ en (2.4) y considerar el índice en *costo total esperado* definido por

$$\begin{aligned} V_1(\pi, x) \quad : &= \lim_{N\to\infty} J_N(\pi, x) \quad \pi \in \Pi, x \in \mathbf{X}; \\ &= E_x^\pi \sum_{t=0}^{\infty} C(x_t, a_t) \end{aligned} \tag{2.7}$$

Sin embargo, este índice puede resultar poco adecuado, ya que en muchos casos la sucesión $\{J_N(\pi,x)\}$ puede diverger a infinito o no tener límite, para algunas o todas las políticas ([59, Section 5.1]). Otra desventaja del índice (2.7) consiste en que no dice nada con respecto a la tasa de crecimiento de los costos en horizonte finito, lo cual no permite discriminar entre dos políticas π^* y π_* para las cuales $V_1(\pi^*,\cdot) = V_1(\pi_*,\cdot)$. Esto último es de mucha importancia si el objetivo es aproximar $J_N^*(\cdot)$, para N grande, por medio del índice en costo total esperado.(Para el estudio detallado del PCO con respecto al índice (2.7) se remite al lector a [4], [59], [61], [65], [67]). Desde esta perspectiva, parece más fructífero considerar *criterios sensibles a la tasa de variación de* $J_N(\pi,x)$ *con respecto al crecimiento de* N, pero este tipo de índices pueden definirse de múltiples maneras que van desde el *índice en costo promedio*, que es el de menor capacidad para discriminar políticas de acuerdo al comportamiento de los costos generados en horizontes finitos, hasta el criterio extremadamente selectivo introducido por Ramsey ([60]), al que nosotros llamaremos *fuertemente dominante*. Entre estos dos extremos existen un buen número de índices con grados distintos de sensibilidad al comportamiento en horizonte finito, entre los que podemos mencionar los siguientes: el criterio introducido por Gale ([26]) and von Weizsäcker ([81]), las nociones de *optimalidad fuerte en costo promedio* y *costo de oportunidad* introducidas por Flynn ([24], [25]), el criterio de Dutta ([21]), la *optimalidad en sesgo* de Veinott ([80]), y la noción de *ternas canónicas* de Yushkevich ([82]).

2.4.2 Problemas de control en horizonte infinito: criterios en costo promedio

Para cada $\pi \in \Pi$, el *costo promedio* (por etapa), dado que $x_0 = x$, se define como

$$J(\pi,x) := \limsup_{N\to\infty} \frac{1}{N} J_N(\pi,x), \tag{2.8}$$

y la *funcion de valor óptimo en costo promedio* por

$$J^*(x) := \inf_{\pi\in\Pi} J(\pi,x), \quad x \in \mathbf{X}. \tag{2.9}$$

Una politica $\pi^* \in \Pi$ es *óptima en costo promedio* si

$$J^*(x) = J(\pi^*,x), \quad \forall x \in \mathbf{X}. \tag{2.10}$$

El índice en costo promedio (2.8) parece ser bastante adecuado para evaluar el funcionamiento de sistemas controlados cuya función de costo no tiene un significado económico directo y que, por otro lado, realizan un número grande de transiciones en periodos cortos de tiempo real— como por ejemplo los sistemas de comunicaciones ([22], [73])— de manera que en tiempos razonables se espera que las sucesión $\{N^{-1}J_N(\pi,\cdot)\}$ ya se haya estabilizado alrededor de su "valor límite", lo que a su vez, permite interpretar a $J(\pi,x)$ como la tasa de crecimiento del costo generado en horizontes finitos.

Un concepto que ha jugado un papel muy importante para el estudio del PCO en costo promedio es el de soluciones de la ecuación de optimalidad en costo promedio: una terna formada por una constante ρ^*, una función medible $h^* : \mathbf{X} \to \mathbf{R}$ y una política estacionaria $f^* \in \mathbf{F}$, es una solución de la *Ecuación de Optimalidad en Costo Promedio* (EOCP) si satisface

$$\rho^* + h^*(x) = \min_{a\in A(x)} \left[C(x,a) + \int_{\mathbf{X}} h^*(y)Q(dy|x,a)\right] \quad \forall x \in \mathbf{X}, \tag{2.11}$$

$$= C(x,f^*) + \int_{\mathbf{X}} h^*(y)Q(dy|x,f^*). \tag{2.12}$$

La importancia de contar con una solución (ρ^*, f^*, h^*) a la EOCP radica en que, bajo condiciones muy débiles sobre la función h^*, se deduce que f^* es óptima en costo promedio y que el costo promedio óptimo es ρ^* para todo estado inicial, es decir,

$$J^*(x) = J(f^*,x) = \rho^* \quad \forall x \in \mathbf{X}. \tag{2.13}$$

Sin embargo, se sabe que la existencia de soluciones de la EOCP no es una condición *necesaria* para la existencia de políticas estacionarias óptimas (ver Ejemplo 5.3.9) y que en ciertos contextos, implica que el modelo de control posee una estructura de recurrencia muy restrictiva ([12], [36]). De hecho, bajo ciertas condiciones, basta con tener una solución a la *Desigualdad de Optimalidad en Costo Promedio* (DOCP), es decir, una terna (ρ^*, f^*, h^*) que satisfaga la relación

$$
\begin{aligned}
\rho^* + h^*(x) &\geq \min_{a \in A(x)} \left[C(x,a) + \int_{\mathbf{X}} h^*(y) Q(dy|x,a) \right] \quad \forall x \in \mathbf{X} &(2.14)\\
&= C(x,f^*) + \int_{\mathbf{X}} h^*(y) Q(dy|x,f^*) \quad \forall x \in \mathbf{X}, &(2.15)
\end{aligned}
$$

para concluir que (2.13) se cumple ([14], [18], [31], [33], [54], [66], [68], [70], [79]).

Un método muy popular para obtener soluciones de la EOCP, o bien de la DOCP, es el enfoque de *Aproximaciones por Problemas Descontados* (APD) cuyas ideas básicas se remontan a Blackwell ([9]) y Taylor ([75]), el cual posteriormente fue afinado y extendido a contextos muy generales por otros autores ([14], [18], [31], [33], [54], [66], [68], [70], [79]). En el Capítulo 3 de esta tesis se desarrolla una variante más del enfoque APD para obtener una solución de la DOCP y se aplica a un problema de inventarios que no había sido cubierto por los resultados obtenidos en trabajos previos.

Otra forma importante de obtener soluciones de la EOCP es suponer que los procesos controlados tienen buenas propiedades de recurrencia y ergodicidad, las cuales pueden interpretarse como propiedades de *estabilidad* ([12], [17], [36]). Un esquema común e importante para conseguir propiedades de este tipo es por medio de la existencia de *funciones de Lyapunov*, tanto para procesos no controlados ([27], [48], [50]) como para procesos controlados ([15], [28], [29], [38], [40], [49]). En el Capítulo 4, se estudia el PCO en costo promedio estableciendo la existencia de una solución de la EOCP bajo una condición del tipo *Lyapunov*, la cual implica que los procesos de Markov inducidos por las políticas estacionarias son *geométricamente ergódicos.* En este marco se demuestra la existencia de una política estacionaria óptima en costo promedio y que las siguientes afirmaciones son equivalentes: a) f es óptima en costo promedio; b) f proviene de una solución de la ecuación de optimalidad; c) f es fuertemente óptima; d) f es Flynn óptima (las nociones de optimalidad en (c) y (d) se introducen párrafos adelante; ver (2.18) y (2.19)). En el Capítulo 5 se hace uso de la misma condición de *Lyapunov* para estudiar la existencia de políticas estacionarias *sensibles al horizonte de planeación.*

Observe que en la formulación del PCO en costo promedio (2.8)-(2.10), al considerar el límite superior se adopta una actitud "conservadora" ya que se busca minimizar el "peor" comportamiento posible; por el contrario, si se toma límite inferior en (2.8) en lugar de límite superior, la actitud hacia el problema de optimización sería optimista.

Ambas alternativas, la optimista y la pesimista, pueden combinarse para formular un criterio de optimalidad más fuerte que el anterior: una política $\pi^* \in \Pi$ es *fuertemente óptima* (en costo promedio) si

$$J(\pi^*, x) \leq \liminf_{N\to\infty} \frac{1}{N} J_N(\pi, x) \quad \forall x \in \mathbf{X}, \pi \in \Pi. \tag{2.16}$$

Por otra parte, note que

$$\limsup_{n\to\infty} \frac{1}{N} J_N^*(x) \leq J(\pi, x) \quad \forall x \in \mathbf{X}, \pi \in \Pi; \tag{2.17}$$

entonces

$$\limsup_{n\to\infty} \frac{1}{N} J_N^*(x) \leq J^*(x) \quad \forall x \in \mathbf{X}. \tag{2.18}$$

La desigualdad en (2.18) implica que una política óptima en costo promedio tendrá en el mejor de los casos una tasa de crecimiento similar a la de los costos óptimos en horizonte finito. Ésta es la idea subyacente en el siguiente criterio de optimalidad introducido por Flynn ([24], [25]). Diremos que una política $\pi^* \in \Pi$ es *Flynn-óptima* (en costo promedio) si

$$\lim_{N\to\infty} \frac{1}{N} [J_N(\pi^*, x) - J_N^*(x)] = 0 \quad \forall x \in \mathbf{X}. \tag{2.19}$$

Es fácil verificar de las definiciones que tanto el criterio de optimalidad fuerte como el de Flynn implican al criterio (estandar) de optimalidad en costo promedio.

Una posibilidad más para el estudio de los problemas en costo promedio consiste en hacer un análisis por trayectorias, es decir, estudiar los costos promedio por etapa conforme estos son *observados* por el controlador (note que los criterios de optimalidad previamente introducidos se formulan en términos de *costos esperados*, no en términos de costos observados). Diremos que una constante $\widehat{\rho}$ es el *costo promedio óptimo por trayectorias* y que una política $\pi^* \in \Pi$ es *óptima por trayectorias* (en costo promedio) si se satisfacen las dos condiciones siguientes:

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{t=0}^{n-1}C(x_t,a_t)=\widehat{\rho}\quad P_x^{\pi^*}-\text{casi seguramente},\ \forall x\in\mathbf{X}; \tag{2.20}$$

y

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{t=0}^{n-1}C(x_t,a_t)\geq\widehat{\rho}\quad P_x^{\pi}-\text{casi seguramente},\ \forall x\in\mathbf{X},\pi\in\Pi. \tag{2.21}$$

Aun cuando este tipo de análisis es muy importante desde el punto de vista práctico, pocas veces se realiza y cuando se hace se restringe a uno de los siguientes casos: a) espacios de estados y controles *finitos* ([47]); b) espacios de estados *numerable*, bajo condiciones restrictivas de estabilidad ([10], [15]); c) espacios de *Borel* y costos *acotados*, bajo la hipótesis de que existe una solución *acotada* de la Ecuación de Optimalidad en Costo Promedio ([3]). En el Capítulo 6, apoyandonos en el análisis proporcionado en [32], se estudia la existencia de políticas estacionarias óptimas por trayectorias para problemas de control en *espacios de Borel* y *costos estrictamente no acotados*, bajo condiciones débiles de continuidad y recurrencia (*Harris recurrencia*). Además, se ilustran los resultados principales con ejemplos de control de inventarios.

Una descripción suscinta y completa de los distintos enfoques estudiados para resolver los problemas en costo promedio puede encontrarse en [3]. Otras fuentes que también proporcionan información relevante son [6], [34], [59].

2.4.3 Problemas de control en horizonte infinito: criterios sensibles al horizonte de planeación

Aun cuando es cierto que los criterios en costo promedio pueden usarse para "aproximar" problemas de control en horizonte finito, tienen la desventaja de ser poco selectivos para este tipo de horizontes; por ejemplo, para ilustrar la afirmación anterior, supongamos que $J_n(\pi_*,x)-J_n^*(x)=n^p$, $\forall n\in\mathbf{N}$, donde $p\in(0,1)$. Es claro que π_* es Flynn-óptima, pero no proporciona una buena aproximación para los problemas de control a largo plazo, es decir, para valores grandes de n. Buscando rodear este obstáculo se han introducido algunos criterios con un mayor grado de sensibilidad al crecimiento del horizonte de planeación.

Uno de estos criterios es el considerado por Ramsey ([60]): una política π^* es *fuertemente* *dominante* si para cada $x \in \mathbf{X}$ y $\pi \in \Pi$ existe un número natural $N = N(\pi^*, \pi, x)$ tal que

$$J_n(\pi^*, x) \leq J_n(\pi, x) \ \ \forall n \geq N. \tag{2.22}$$

Aunque es cierto que la noción de políticas *fuertemente dominantes* está diseñada para detectar políticas con muy buen comportamiento (casi óptimo) en horizontes finitos, resulta ser extremadamente selectiva como para ser de alguna utilidad ([11], [57], [59]). Una versión más débil de este criterio la proporcionan Gale ([26]) y von Weiszacker ([81]): una política π^* es *dominante* si

$$\limsup_{n\to\infty} \left[J_n(\pi^*, x) - J_n(\pi, x)\right] \leq 0 \ \ \forall x \in \mathbf{X}, \pi \in \Pi. \tag{2.23}$$

Observe que una política *fuertemente dominante* necesariamente es *dominante*

En [19] se proporcionan condiciones suficientes para la existencia de políticas *estacionarias fuertemente dominantes*, pero su análisis se restringe a procesos controlados con espacio de estados y controles *finitos*. Por otra parte, en [11] y [57] se exhiben ejemplos sencillos con buenas propiedades de estabilidad en el cual no existe una política estacionaria dominante, pero en ambos casos se violan las condiciones dadas en [19].

Otro índice alternativo es el *costo de oportunidad* introducido por Flynn ([24], [25]), definido como

$$CO(\pi, x) := \limsup_{n\to\infty} \left[J_n(\pi, x) - J_n^*(x)\right], \ \ x \in \mathbf{X}, \pi \in \Pi; \tag{2.24}$$

el *costo de oportunidad óptimo* está definido por

$$CO(x) := \inf_{\pi\in\Pi} CO(\pi, x). \tag{2.25}$$

Una política π^* es *óptima en costo de oportunidad* si

$$CO(x) = CO(\pi^*, x) \ \ \forall x \in \mathbf{X}. \tag{2.26}$$

Observe que si $CO(x) = +\infty$, no existe política alguna que dé una aproximación adecuada a los problemas de control a largo plazo, es decir, conforme el horizonte de planeación crece los costos generados por cualquier política "divergen" de los costos óptimos. Reciprocamente, si para alguna política $\pi \in \Pi$ se tiene que $CO(\pi, x) < +\infty$, podemos decir que dicha política es "buena" para los problemas en horizonte finitos. (Note que en este caso, la política π es Flynn-óptima.)

Una variante más es proporcionada por Dutta ([21]), quien considera el siguiente índice de funcionamiento: para cada $\pi \in \Pi$, el *costo de Dutta* se define por

$$D(\pi, x) := \limsup_{n\to\infty} [J_n(\pi, x) - nJ^*(x)] \quad x \in \mathbf{X}; \tag{2.27}$$

la función de costo óptimo correspondiente está dada por

$$D(x) := \inf_{\pi\in\Pi} D(\pi, x), \quad x \in \mathbf{X}. \tag{2.28}$$

Una política $\pi^* \in \Pi$ es *Dutta-óptima* si

$$D(x) = D(\pi^*, x) \;\; \forall x \in \mathbf{X}. \tag{2.29}$$

Note que en las definiciones de políticas dominantes (2.23), costo de oportunidad (2.24) y costo de Dutta (2.27) las cantidades involucradas, en principio, toman valores en los reales extendidos. Ésto provoca que en general (es decir, sin condiciones adecuadas) no pueda establecerse un orden jerárquico entre las nociones de optimalidad correspondientes. Por ejemplo, Flynn ([25]) proporciona un ejemplo en el cual existe una política dominante (en su terminología, *overtaking optimal*) la cual no es "buena" con respecto al costo de oportunidad. Es fácil verficar que en este ejemplo se tiene que $CO(x) = +\infty \; \forall x \in \mathbf{X}$.

En el Capítulo 5, bajo las condiciones de estabilidad introducidas en el Capítulo 4, se demuestra que existe una política estacionaria $f^* \in \mathbf{F}$ que es Dutta-óptima en la clase de las políticas estacionarias $\mathbf{F}$, es decir,

$$D(f^*, x) = \inf_{f\in\mathbf{F}} D(f, x) \;\; \forall x \in \mathbf{X},$$

y que las siguientes afirmaciones son equivalentes: i) $f \in \mathbf{F}$ es Dutta-óptima en $\mathbf{F}$; ii) $f \in \mathbf{F}$ es dominante en $\mathbf{F}$; iii) $f \in \mathbf{F}$ es óptima en costo de oportunidad en $\mathbf{F}$ y $CO(f, \cdot) < +\infty$; iv) $f \in \mathbf{F}$ es óptima en sesgo en $\mathbf{F}$. (Esta noción de optimalidad se introduce en la Definición 5.3.1.) Los ejemplos dados en [11] y [57] muestran que las afirmaciones anteriores no se extienden a la clase de todas las políticas sin imponer sobre el modelo de control condiciones adicionales a las de estabilidad.

Capítulo 3

Indice en Costo Promedio: Aproximación por Problemas Descontados

3.1 Introducción

En este capítulo se estudia el *problema de control óptimo* (PCO) *en costo promedio* por medio del enfoque de *Aproximaciones por Problemas Descontados* (APD). Quizás, este enfoque sea el más popular y, posiblemente, parte de esta popularidad se deba a su flexibilidad para adaptarse a distintos contextos sin dificultades importantes ya que, al contrario de la mayoria de los enfoques—si no es que todos— se ha mostrado que igual es aplicable cuando los espacios de estados y controles son discretos o espacios de Borel, los costos son acotados o no, o bien si se tienen o no propiedades de recurrencia/ergodicidad en el modelo de control.

El enfoque APD ha sido estudiado intensamente en años recientes—para el caso *discreto* ver por ejemplo, [14], [16], [18], [41], [68], [70]; y para el caso de *Borel*, [31], [33], [34], [54], [51], [52], [53], [66]—pero las ideas básicas se remontan a Blackwell [9], quién consideró espacios de estados y controles *finitos*. Posteriormente, el APD fue afinado y extendido a espacios discretos y costos acotados por Taylor [75] y Ross [62]; a espacios numerables y costos no acotados por Sennott [68], y más recientemente a espacios de Borel por Hernández-Lerma and Lasserre [33], Hernández-Lerma [32], Schäl [66]. (Para la aplicación del enfoque APD en problemas de control *semi-Markovianos* ver, por ejemplo, [46], [69], [76], [78].) No obstante la gran cantidad de literatura existente sobre el enfoque APD y la generalidad de los resultados que se han obtenido, aun se excluyen aplicaciones importantes de inventarios, para cuya solución se requiere extender o complementar algunos de los trabajos previos. Especificamente los resultados que se presentan en este capítulo extienden a espacios de Borel el analisis proporcionado por Sennott [70] para espacios discretos. Casi la totalidad del material que se presenta en el resto del capítulo se tomó de [79].

3.2 El problema de control óptimo en costo promedio

Aun cuando el PCO en costo promedio ya fue formulado en el capítulo anterior, se hace nuevamente aquí para facilitar la lectura. Consideraremos un MCM fijo $(\mathbf{X}, \mathbf{A}, \{A(x), x \in \mathbf{X}\}, Q, C)$ y para cada natural N, definimos el *costo* (esperado) *en N-etapas* cuando se usa la política $\pi \in \Pi$, dado que el estado inicial es $x_0 = x \in \mathbf{X}$, por

$$J_N(\pi, x) := E_x^\pi \sum_{t=0}^{N-1} C(x_t, a_t), \tag{3.1}$$

y el *costo promedio* (por etapa) como

$$J(\pi, x) := \limsup_{N\to\infty} \frac{1}{N} J_N(\pi, x). \tag{3.2}$$

También definimos la *función de costo promedio óptimo*

$$J^*(x) := \inf_{\pi\in\Pi} J(\pi, x) \quad x \in \mathbf{X}. \tag{3.3}$$

Una política $\pi^* \in \Pi$ es *óptima en costo promedio* si

$$J^*(x) = J(\pi^*, x) \ \ \forall x \in \mathbf{X}. \tag{3.4}$$

Además, recuerde que para cada $\alpha \in (0,1)$, el *índice en costo α-descontado* está dado por

$$V_\alpha(\pi, x) := E_x^\pi \sum_{t=1}^{\infty} \alpha^t C(x_t, a_t), \tag{3.5}$$

y la *función de costo óptimo $\alpha -$ descontado* se define por

$$V_\alpha(x) := \inf_{\pi \in \Pi} V_\alpha(\pi, x) \ \ x \in \mathbf{X}. \tag{3.6}$$

El PCO en costo α-descontado se define en forma análoga al problema en costo promedio; naturalmente, a una política π_* que satisface

$$V_\alpha(x) = V_\alpha(\pi_*, x) \ \ x \in \mathbf{X}, \tag{3.7}$$

le llamaremos *óptima en costo $\alpha -$descontado*, o más brevemente, *α-óptima*. Finalmente, defina

$$V_\alpha^n(x) := \inf_{\pi \in \Pi} \ E_x^\pi \sum_{t=1}^{n-1} \alpha^t C(x_t, a_t) \ \ x \in \mathbf{X}, n \in \mathbf{N}. \tag{3.8}$$

En el desarrollo del enfoque APD haremos uso de las siguientes condiciones de compacidad/continuidad, cuyo objetivo es garantizar la existencia de minimizadores medibles.

H3.2.1(a) C es no-negativa e inferiormente compacta en $\mathbf{K}$; es decir, para cada $x \in \mathbf{X}$ y cada número real $r \in \mathbf{R}$, el subconjunto $\{a \in A(x) : C(x,a) \leq r\}$ es compacto;

(b) $Q(\cdot|x,a)$ es fuertemente continua en $a \in A(x)$ para cada $x \in \mathbf{X}$, es decir, la función

$$a \mapsto \int_{\mathbf{X}} u(y) Q(dy|x,a),$$

es continua en $a \in A(x)$, para cada función medible y acotada $u \in M_a(\mathbf{X})$.

Algunas consecuencias de H3.2.1 se presentan a continuación.

Observación 3.2.2.(a) Note que H3.2.1(a) implica que $C(x, \cdot)$ es semicontinua inferiormente, para cada $x \in \mathbf{X}$. Entonces la función

$$C(x, \cdot) + \int_{\mathbf{X}} u(y) Q(dy|x, \cdot)$$

es semicontinua inferiormente en $A(x)$, para cada $x \in \mathbf{X}$ y cada función $u \in L(\mathbf{X})$ semicontinua inferiormente acotada por abajo . Por lo tanto, del Teorema de Selección Medible (Teorema A.B.5) existe un selector $f \in \mathbf{F}$, tal que $\forall x \in \mathbf{X}$:

$$C(x, f) + \int_{\mathbf{X}} u(y) Q(dy|x, f) = \inf_{a \in A(x)} \left[C(x, a) + \int_{\mathbf{X}} u(y) Q(dy|x, a) \right].$$

(b) Finalmente, note que no se gana generalidad si, en lugar de suponer que C es no-negativa, se supone que es acotada inferiormente por una constante M, ya que el problema de control en costo promedio para la funcion no-negativa $C' := C - M$ es equivalente al problema de control original. Esto mismo también es cierto para el problema en costo descontado.

En el siguiente teorema se recopilan algunos resultados sobre el PCO en costo descontado. En la demostración de este resultado se usan argumentos estándar en la literatura relacionada y su demostración puede consultarse, por ejemplo, en [34] (Theorem 4.2.3, p. 46).

Teorema 3.2.3. Sea $\alpha \in (0, 1)$ un factor de descuento fijo. Si H3.2.1 se satisface y $V_\alpha(x) < +\infty \ \forall x \in \mathbf{X}$, entonces

(a) $V_\alpha^n(\cdot) \uparrow V_\alpha(\cdot)$ cuando $n \to \infty$;

(b) $V_\alpha(\cdot)$ satisface la *Ecuación de Optimalidad en Costo α-Descontado* (α-EOCD):

$$V_\alpha(x) = \min_{a \in A(x)} \left\{ C(x, a) + \alpha \int_{\mathbf{X}} V_\alpha(y) Q(dy|x, a) \right\} \quad \forall x \in \mathbf{X}; \tag{3.9}$$

(c) una política $f \in \mathbf{F}$ es óptima en costo α-descontado si y sólo si $f(x)$ alcanza el mínimo en (3.9) para cada $x \in \mathbf{X}$;

(d) existe un selector $f_\alpha \in \mathbf{F}$ tal que

$$V_\alpha(x) = C(x, f_\alpha) + \int_{\mathbf{X}} V_\alpha(y) Q(dy|x, f_\alpha) \quad \forall x \in \mathbf{X};$$

por lo tanto, de (c), tenemos que f_α es óptima en costo α-descontado.

Los índices en costo promedio (3.2) y descontado (3.5) están relacionados por el siguiente Teorema Abeliano. La demostración puede encontrarse en [34] (Lemma 5.3.1, p. 84; Note 5.3.2, p. 85).

Teorema 3.2.4. Para cada política $\pi \in \Pi$ y cada estado $x \in \mathbf{X}$ se cumple lo siguiente:

(a) $\liminf_{n\to\infty} \frac{1}{n} E_x^{\pi} \sum_{t=0}^{n-1} C(x_t, a_t) \leq \liminf_{\alpha\uparrow 1} (1-\alpha)V_\alpha(\pi, x)$

$$\leq \limsup_{\alpha\uparrow 1} (1-\alpha)V_\alpha(\pi, x) \leq J(\pi, x);$$

(b) $\limsup_{\alpha\to 1^-} (1-\alpha)V_\alpha(x) \leq J^*(x)$.

3.3 La desigualdad de optimalidad en costo promedio

Para cada $\alpha \in (0,1)$ se definen las funciones

$$h_\alpha(x) := V_\alpha(x) - V_\alpha(z),\ x \in \mathbf{X}, \quad \text{y} \quad \rho_\alpha := (1-\alpha)V_\alpha(z), \tag{3.10}$$

donde $z \in \mathbf{X}$ es fijo. Con esta notación, la α-EOCD (3.9) puede reescribirse equivalentemente como

$$\rho_\alpha + h_\alpha(x) = \min_{a\in A(x)} \left\{ C(x,a) + \alpha \int_{\mathbf{X}} h_\alpha(y) Q(dy|x,a) \right\} \quad \forall x \in \mathbf{X}. \tag{3.11}$$

El enfoque de *Aproximaciones por Problemas Descontados* (APD) consiste en proporcionar condiciones que garanticen que al tomar límite en (3.11) cuando $\alpha \to 1$, se obtenga la *Desigualdad de Optimalidad en Costo Promedio*:

$$\rho^* + h^*(x) \geq C(x, f_*) + \int_{\mathbf{X}} h^*(y) Q(dy|x, f_*) \ \ \forall x \in \mathbf{X}, \tag{3.12}$$

donde ρ^* es una constante, $f_* \in \mathbf{F}$ y h^* es una función medible con valores reales. En una segunda etapa, bajo condiciones adicionales sobre h^*, se demuestra que f_* es costo promedio óptima y que la función de costo óptimo es igual a ρ^* para todo estado inicial $x \in \mathbf{X}$.

Observe que de (3.12) se obtiene

$$\rho^* + h^*(x) \geq \min_{a\in A(x)} \left\{ C(x,a) + \int_{\mathbf{X}} h^*(y)Q(dy|x,a) \right\} \quad \forall x \in \mathbf{X}; \tag{3.13}$$

$$= C(x,f^*) + \int_{\mathbf{X}} h^*(y)Q(dy|x,f^*). \tag{3.14}$$

Si (3.13) se cumple con igualdad, se dice que (ρ^*, f^*, h^*) es una solución de la *Ecuación de Optimalidad en Costo Promedio* (EOCP). Contar con una solución de la EOCP es más deseable que tener una solución para la DOCP ya que la primera proporciona mayor información sobre el problema de control. Sin embargo, en ([13]) se da un ejemplo que muestra que la desigualdad en (3.13) puede ser estricta; por otra parte, en ciertos contextos, la existencia de soluciones de la EOCP es "equivalente" a que los procesos controlados posean propiedades restrictivas de recurrencia ([12], [36]). El estudio de existencia de soluciones de la EOCP será retomado en el Capítulo 4 bajo condiciones de estabilidad.

Para la solución del PCO en costo promedio usaremos el siguiente conjunto de condiciones.

H3.3.1(a) $\rho_\alpha \leq M \quad \forall \alpha \in (0,1)$, donde M es una constante positiva;

(b) existe una función no-negativa $N{:}\mathbf{X} \to \mathbf{R}$ y una sucesión $\alpha(n) \uparrow 1$ tales que

(i) $-N(x) \leq h_{\alpha(n)}(x) \quad \forall x \in \mathbf{X}, n \in \mathbf{N}$;

(ii) la función

$$h^*(x) := \liminf_{n\to\infty} h_{\alpha(n)}(x) \quad x \in \mathbf{X}$$

sólo toma valores reales;

(c) la función

$$a \mapsto \int_{\mathbf{X}} N(y)Q(dy|x,a)$$

es continua para cada $x \in \mathbf{X}$;

(d) $\overline{N}(x) := \sup_{a\in A(x)} \int_{\mathbf{X}} N(y)Q(dy|x,a) < +\infty \quad \forall x \in \mathbf{X}$.

Observación 3.3.2. El papel de la condición H3.3.1(d), como puede verificarse en la demostración del Teorema 3.3.3, consiste en garantizar la existencia de puntos de acumulación de las políticas α-descontadas cuando α se aproxima a uno. Entonces, uno puede sustituir esta condición por cualquier otra que dé el mismo resultado. Por ejemplo, podría suponerse que los conjuntos $A(x)$ son compactos para cada $x \in \mathbf{X}$, y las conclusiones en el Teorema 3.3.3 no serían afectadas. En el ejemplo de inventarios que se discute en la Sección 3.4 es más fácil verificar que las políticas α-descontadas tienen puntos de acumulación que verificar que H3.3.1(d) se satisface.

Teorema 3.3.3. Si las hipótesis H3.2.1 y H3.3.1 se satisfacen, entonces

(a) existe una constante ρ^* y una política estacionaria $f^* \in \mathbf{F}$ tales que

$$\rho^* + h^*(x) \geq C(x, f^*) + \int_{\mathbf{X}} h^*(y)Q(dy|x, f^*) \quad \forall x \in \mathbf{X}, \tag{3.15}$$

donde h^* es la función definida en H3.3.1(b);

(b) si

$$\lim_{n\to\infty} \frac{1}{n} E_x^{f^*} N(x_n) = 0 \quad \forall x \in \mathbf{X}, \tag{3.16}$$

entonces f^* es óptima en costo promedio y $J^*(x) = J(f^*, x) = \rho^*$ $\forall x \in \mathbf{X}$; además

$$\rho^* := \lim_{\alpha\uparrow 1} \rho_\alpha = \lim_{\alpha\uparrow 1}(1-\alpha)V_\alpha(x) \quad \forall x \in \mathbf{X}. \tag{3.17}$$

Para la demostración del Teorema 3.3.3 necesitamos el siguiente resultado preliminar.

Lema 3.3.4. Si H3.3.1 se satisface, entonces $V_\alpha(x) < +\infty$ $\forall x \in \mathbf{X}$ y $\alpha \in (0, 1)$.

Demostracion del Lema 3.3.4. Sea $\{\alpha(n)\}$ la sucesión en H3.3.1(b) y $x \in \mathbf{X}$ un estado arbitrario fijo. Tómese una subsucesión $\{\beta(n)\} \subset \{\alpha(n)\}$ tal que

$$h_{\beta(n)}(x) \to h^*(x) \quad \text{cuando } n \to \infty.$$

Puesto que

$$(1-\beta(n))V_{\beta(n)}(x) = \rho_{\beta(n)} + (1-\beta(n))h_{\beta(n)}(x) \quad \forall n \in \mathbf{N},$$

las condiciones H3.3.1(a)-(b) implican que $\{(1-\beta(n))V_{\beta(n)}(x)\}$ es una sucesión acotada. Entonces, $V_{\beta(n)}(x) < +\infty\ \forall n$, lo cual implica que $V_\alpha(x) < +\infty\ \forall x \in \mathbf{X}$, ya que $\beta(n) \to 1$. ∎

Demostración del Teorema 3.3.3.(a) Sea $x \in \mathbf{X}$ un estado arbitrario fijo. Del Lema 3.3.4, Teorema 3.2.3 y (3.11), existe una sucesion $\{a_n\} \subset A(x)$ tal que

$$\rho_{\alpha(n)} + h_{\alpha(n)} = C(x,a_n) + \alpha(n)\int_{\mathbf{X}} h_{\alpha(n)}(y)Q(dy|x,a_n) \ \ \forall n \geq 1, \tag{3.18}$$

donde $\{\alpha(n)\}$ es la sucesión en H3.3.1(b). Ahora defina

$$\rho^* := \liminf_{n\to\infty} \rho_{\alpha(n)} \quad \text{y} \quad H_n(y) := \inf_{k\geq n} h_{\alpha(n)}(y), \ \ y \in \mathbf{X}, n \in \mathbf{N}. \tag{3.19}$$

Entonces, se tiene que

$$C(x,a_n) + \alpha(n)\int_{\mathbf{X}} H_n(y)Q(dy|x,a_n) \leq \rho_{\alpha(n)} + h_{\alpha(n)}(x) \ \ \forall n \in \mathbf{N}.$$

Además, (3.19) y H3.3.1(b) garantizan que para cada $\varepsilon > 0$ existe una sucesión creciente de enteros no-negativos $\{n'\}$ y un número real M_ε tales que

$$C(x,a_{n'}) + \alpha(n')\int_{\mathbf{X}} H_{n'}(y)Q(dy|x,a_{n'}) \leq \rho^* + h^*(x) + \varepsilon \ \ \forall n \geq M_\varepsilon.$$

Entonces, usando las condiciones en H3.3.1(a)-(c), se obtiene que

$$\begin{aligned} C(x,a_{n'}) + \alpha(n')\int_{\mathbf{X}} [H_{n'}(y) + N(y)]Q(dy|x,a_{n'}) \ &\leq \ \rho^* + h^*(x) + \\ &\int_{\mathbf{X}} N(y)Q(dy|x,a'_n) \qquad (3.20) \\ &\leq \ \rho^* + h^*(x) + \overline{N}(x). \qquad (3.21) \end{aligned}$$

Ahora, como consecuencia de (3.21) y H3.2.1(b), se tiene que los conjuntos

$$D_{n'}(x) := \left\{ a \in A(x) : C(x, a_{n'}) + \alpha(n') \int_{\mathbf{X}} [H_{n'}(y) + N(y)] Q(dy|x, a_{n'}) \leq r \right\},$$

donde $r := \rho^* + h^* + \overline{N}(x) + \varepsilon$, forman una sucesión no-decreciente de subconjuntos compactos no vacíos; en consecuencia, existe una subsucesión de $\{a_{n'}\}$, que denotaremos por $\{a_{n'}\}$ de nuevo, la cual converge a algún $a_x \in A(x)$. Entonces, usando la semi-continuidad de C (Observación 3.2.2(a)) y aplicando el Lema de Fatou ([64], Ch.11, Prop. 17) a (3.20) se obtiene

$$C(x, a_x) + \int_{\mathbf{X}} [h^*(y) + N(y)] Q(dy, a_x) \leq \rho^* + h^*(x) + \int_{\mathbf{X}} N(y) Q(dy|x, a_x).$$

Puesto que ε y x son arbitrarios, la desigualdad anterior implica que

$$\inf_{a \in A(x)} \left\{ C(x, a) + \int_{\mathbf{X}} h^*(y) Q(dy|x, a) \right\} \leq \rho^* + h^*(x) \quad \forall x \in \mathbf{X}.$$

Por otra parte, puesto que $h^*(\cdot) + N(\cdot) \geq 0$, tenemos que

$$C(x, \cdot) + \int_{\mathbf{X}} [h^*(y) + N(y)] Q(dy|x, \cdot)$$

es semi-continua en $A(x)$. Por lo tanto, de H3.3.1(c)-(d), se concluye que

$$C(x, \cdot) + \int_{\mathbf{X}} h^*(y) Q(dy|x, \cdot)$$

es semicontinua inferiormente y acotada inferiormente por $-\overline{N}(x)$, para cada $x \in \mathbf{X}$. Por lo tanto, del Teorema de Selección (Teorema A.B.5), tenemos que la desigualdad (3.15) se cumple, es decir, existe $f^* \in \mathbf{F}$ tal que

$$\rho^* + h^*(x) \geq C(x, f^*) + \int_{\mathbf{X}} h^*(y) Q(dy|x, f^*) \quad \forall x \in \mathbf{X}.$$

(b) De la iteración de (3.15) resulta

$$n\rho^* + h^*(x) \geq J_n(f^*, x) + E_x^{f^*} h^*(x_n).$$

Sumando $E_x^{f^*} N(x_n)$ en ambos lados de la desigualdad y usando que $h^*(\cdot) + N(\cdot) \geq 0$, se obtiene

$$\begin{aligned} n\rho^* + h^*(x) + E_x^{f^*} N(x_n) &\geq J_n(f^*, x) + E_x^{f^*}\left[h^*(x_n) + N(x_n)\right] \\ &\geq J_n(f^*, x); \end{aligned}$$

entonces, multiplicando por $\frac{1}{n}$ y tomando límite cuando $n \to \infty$, se obtiene

$$\rho^* \geq J(f^*, x) \ \ \forall x \in \mathbf{X},$$

lo cual, combinado con el Teorema 3.2.4(b), implica

$$J^*(x) = J(f^*, x) = \rho^* \ \ x \in \mathbf{X}. \tag{3.22}$$

Finalmente, falta por demostrar que (3.17) se cumple, es decir,

$$\rho^* = \limsup_{\alpha \to 1^-} (1 - \alpha) V_\alpha(x) \ \ \forall x \in \mathbf{X}.$$

Para hacerlo, observe que $(1 - \alpha)V_\alpha(x) = \rho_\alpha + (1 - \alpha)h_\alpha(x)$ $\forall x \in \mathbf{X}$, de manera que

$$\limsup_{\alpha \to 1^-} (1 - \alpha) V_\alpha(x) \geq \rho^* \ \ \forall x \in \mathbf{X},$$

lo cual, cmbinado con (3.22) y el Teorema 3.2.4, implica que (3.17) se satisface. ■

3.4 Un problema de control de inventarios

Ahora aplicaremos los resultados de la sección anterior a un problema de inventarios. Denotaremos por x_t a la cantidad de producto en inventario y por a_t la cantidad solicitada a la unidad de producción— la que se supone disponible inmediatamente después de solicitarse—, en ambos casos, al iniciar el t−ésimo periodo. La cantidad de producto demandada durante el t−ésimo periodo es representada por w_t, la cual suponemos que es una variable aleatoria no-negativa. El sistema de inventario evoluciona en $\mathbf{X} = [0, +\infty)$ de acuerdo a

$$
\begin{aligned}
x_{t+1} &= \max(x_t + a_t - w_t, 0) \;\; t = 0, 1, \cdots \\
x_0 &= x,
\end{aligned}
\tag{3.23}
$$

donde $\{a_t\} \subset \mathbf{A} = A(x) = [0, +\infty)$ $\forall x \in \mathbf{X}$. Suponemos también que la demanda satisface lo siguientes condiciones.

H3.4.1(a) El proceso $\{w_t\}$ está formado por variables aleatorias *i.i.d;* la función de distribución es denotada por $G(\cdot)$;

(b) $G(\cdot)$ tiene densidad $\mu(\cdot)$ continua y acotada.

La esperanza con respecto a la distribución conjunta de las variables $\{w_t\}$ se denotará por E.

Para la evaluación de las políticas de control consideraremos que la función de costo por etapa tiene la siguiente forma

$$
C(x, a) = F(x + a) + ba \;\; \forall (x, a) \in \mathbf{K},
\tag{3.24}
$$

donde b es una constante positiva y F es una función definida en $\mathbf{X} = [0, +\infty)$ que satisface lo siguiente.

H3.4.2(a) F es convexa y continua en $y = 0$;

(b) $\lim_{y \to +\infty} F(y) = +\infty$.

En el enunciado de la solución del PCO en costo promedio definido por (3.23)-(3.24) se usarán las siguientes funciones:

$$
L(y) \;\; : = F(y) + bE\min(y, w_0), \quad y \geq 0,
\tag{3.25}
$$

$$
L_\alpha(y) \;\; : = F(y) + \alpha bE\min(y, w_0) + (1 - \alpha)by, \quad y \geq 0,
\tag{3.26}
$$

Teorema 3.4.3. Si se satisface H3.4.1 y H3.4.2, entonces la política

$$f^*(x) := \begin{cases} R^* - x & \text{si} \quad 0 \leq x \leq R^* \\ 0 & \quad\quad x > R^* \end{cases}$$

es óptima en costo promedio y el costo óptimo es $\rho^* = L(R^*)$, donde R^* es una constante que satisface la relación

$$L(R^*) = \inf_{y \geq 0} L(y).$$

Para la demostración del Teorema 3.4.3 se requieren algunos resultados preliminares sobre los problemas descontados, los cuales seran presentados con sus demostraciones después de discutir brevemente el siguiente caso particular de (3.24).

Ejemplo 3.4.4. Supongamos que se satisface H3.4.1 y que

$$F(y) = p \min(0, w_0 - y) + h_c y - sE \min(y, w_0) \quad y \geq 0, \tag{3.27}$$

donde p, h_c y s son constantes positivas. Es fácil verificar que esta elección específica para F satisface H3.4.2.

Observe que en este caso, la función de costo por etapa está dada por la fórmula siguiente:

$$\begin{aligned} \text{costo por etapa} \quad = \quad & \text{costo de penalización por demanda no satisfecha} + \\ & \text{costo por manejo de inventario} \ + \ \text{costo de producción} \\ & - \text{ingreso por ventas.} \end{aligned}$$

Con cálculos directos se obtienen los siguientes resultados:

(A) Si $s + p > b + h_c$, entonces R^* es la *única* constante que satisface la ecuación

$$G(y) = \frac{(s+p) - (b + h_c)}{s + p - h_c} \tag{3.28}$$

y el costo promedio óptimo es

$$\rho^* = pE\min(0, w_0 - R^*) + h_c R^* + (s-b)E\min(R^*, w_0). \tag{3.29}$$

(B) Si $s + p \leq b + h_c$, entonces $R^* = \rho^* = 0$.

A continuación se presentan algunas consecuencias inmediatas de H3.4.1 y H3.4.2 que se usan en la demostración del Teorema 3.4.3.

Observación 3.4.5(a) $C(x, a)$ es una función convexa en (x, a) y, además, tiende a infinito si $x \to +\infty$ o $a \to +\infty$; por lo tanto, C es acotada inferiormente y satisface H3.2.1(a), es decir, C es compacta inferiormente en $\mathbf{K}$.

(b) La ley de evolución (3.23) puede expresarse equivalentemente por medio de la probabilidad de transición

$$Q(B|x, a) = \int_{\mathbf{X}} \mathbf{I}_B[(x + a - w)^+]G(dw) \quad \forall B \in \mathcal{B}(\mathbf{X}) \text{ y } (x, a) \in \mathbf{K}.$$

Además,

$$\int_{\mathbf{X}} u(y)Q(dy|x, a) = Eu[(x + a - w)^+] \quad \forall u \in M_a(\mathbf{X}),$$

de donde se deduce, usando H3.4.1(b), que la función

$$(x, a) \mapsto \int_{\mathbf{X}} u(y)Q(dy|x, a)$$

es continua en $\mathbf{K}$. Por lo tanto se satisface la condición H3.2.1(b).

(c) Observe que las funciones $L(\cdot)$ y $L_\alpha(\cdot)$, $\alpha \in (0, 1)$, son continuas y que ambas tienden a infinito cuando $y \to +\infty$. Además, es fácil verificar que $L_\alpha(\cdot) \downarrow L(\cdot)$ si $\alpha \uparrow 1$.

En el siguiente teorema se presenta la solución del PCO en costos $\alpha-$descontados.

Teorema 3.4.6. Supongan que se satisface H3.4.1 y H3.4.2. Entonces para cada $\alpha \in (0, 1)$:

(a) $V_\alpha(\cdot)$ es una función convexa;

(b) la política

$$f_\alpha(x) := \begin{cases} R_\alpha - x & \text{si} \quad 0 \leq x \leq R_\alpha \\ 0 & \text{si} \quad x > R_\alpha \end{cases} \tag{3.30}$$

es *óptima en costo α-descontado*, donde R_α es una constante no-negativa que satisface

$$L_\alpha(R_\alpha) = \inf_{y \geq 0} L_\alpha(y), \tag{3.31}$$

y $L_\alpha(\cdot)$ es la función definida en (3.26);

(c) existe una sucesión $\alpha(n) \uparrow 1$ tal que la sucesión de políticas $\{f_{\alpha(n)}\}$ convergen puntualmente a la política

$$f^*(x) := \begin{cases} R^* - x & \text{si} \quad 0 \leq x \leq R^* \\ 0 & \text{si} \quad x > R^* \end{cases}, \tag{3.32}$$

con R^* tal

$$L(R^*) = \inf_{y \geq 0} L(y), \tag{3.33}$$

donde $L(\cdot)$ es la función definida en (3.25).

Ejemplo 3.4.7. Para el caso específico (3.27), la constante R_α en (3.31) queda especificada de la siguiente forma:

(A) Si $p + s > b + h_c$, entonces R_α es la *única* solución de la ecuación

$$G(y) = \frac{(s+p) - (h_c + b)}{s + p - \alpha b}.$$

(B) Si $p + s \leq b + h_c$, entonces $R_\alpha = 0$.

Demostración del Teorema 3.4.6(a). Primero mostraremos que para cada $\alpha \in (0,1)$, la función $V_\alpha(\cdot)$ is finita. Para hacer lo anterior, fijemos α arbitrario y una constante no-negativa K. Consideremos la política

$$f_K(x) := \begin{cases} K - x & \text{si} \quad 0 \leq x \leq K \\ 0 & \text{si} \quad x > K \end{cases} \tag{3.34}$$

y denotemos por $\{x_t\}$ al proceso de Markov definido por f. Ahora, observe que si $x_0 = x$, entonces el proceso nunca sale del intervalo $[0, K]$, es decir, dicho intervalo es un conjunto absorbente. Así que

$$x_{t+1} = (K_t - w_t)^+ \quad t = 1, 2, \cdots;$$

entonces $\{x_t\}$ es una sucesion de variables aleatorias *i.i.d.* De manera que de H3.4.2(c) y con cálculos directos se obtiene que

$$V_\alpha(f_K, x) = \frac{1}{1-\alpha}\{F(K) + \alpha b E \min(K, w_0) + b(1-\alpha)(K - x)\} < +\infty \tag{3.35}$$

para todo $x \leq K$; entonces, $V_\alpha(x) < +\infty$ $\forall x \in [0, K]$. Puesto que K es arbitraria, se concluye que $V_\alpha(\cdot) < +\infty$.

Ahora demostraremos que $V_\alpha(\cdot)$ es convexa para cada $\alpha \in (0,1)$. Del Teorema 3.2.3(a) tenemos que

$$V_\alpha^n(\cdot) \uparrow V_\alpha(\cdot) \tag{3.36}$$

cuando $n \to \infty$, donde $V_\alpha^n(\cdot)$ es la función definida en (3.8). Por otro lado, puesto que $V_\alpha^1(x) = \min_{a \in A(x)} C(x, a)$, $x \in \mathbf{X}$, es una función convexa, usando inducción se, puede mostrar que $V_\alpha^n(\cdot)$ es convexa para cada $n \geq 1$, lo cual combinado con (3.36), implica que $V_\alpha(\cdot)$ también es convexa.

(b) Note que del Teorema 3.2.3(b) y de la Observación 3.4.5 se tiene

$$V_\alpha(x) = \inf_{a \geq 0}\{C(x, a) + \alpha E V_\alpha[(x + a - w_0)^+]\} \quad \forall x \in \mathbf{X}, \alpha \in (0,1). \tag{3.37}$$

Ahora bien, definiendo

$$T_\alpha(y) := F(y) + by + \alpha E V_\alpha[(y - w_0)^+] \quad \alpha \in (0,1),\ y \geq 0, \tag{3.38}$$

la ecuación (3.37) puede re-escribirse como

$$V_\alpha(x) = \inf_{a\geq 0}\{T_\alpha(x+a)\} - bx \quad \forall x \in \mathbf{X},\, \alpha \in (0,1). \tag{3.39}$$

Por otra parte, note que para cada $\alpha \in (0,1)$, la función $T_\alpha(\cdot)$ es convexa y que $\lim_{y\to+\infty} T_\alpha(y) = +\infty$. En consecuencia, existe una constante R_α tal que

$$T_\alpha(R_\alpha) = \inf_{y\geq 0} T_\alpha(y),$$

de manera que la política

$$f_\alpha(x) := \begin{cases} R_\alpha - x & \text{si} \quad 0 \leq x \leq R_\alpha \\ 0 & \text{si} \quad x > R_\alpha \end{cases},$$

alcanza el mínimo en la Ecuación de Optimalidad en Costo α-Descontado (3.39); por lo tanto, del Teorema 3.2.3(c), tenemos que la política f_α es óptima en costo α-descontado.

Para demostrar la segunda afirmación en (b), note que (3.35) implica que

$$V_\alpha(0) = V_\alpha(f_\alpha, 0) = \frac{1}{1-\alpha} L_\alpha(R_\alpha) \leq V_\alpha(f_K, 0) = \frac{1}{1-\alpha} L_\alpha(K),$$

donde $L_\alpha(\cdot)$ es la función definida en (3.26) y f_K es la política en (3.34). Por lo tanto,

$$L_\alpha(R_\alpha) = \inf_{y\geq 0} L_\alpha(y).$$

(c) De la Observación 3.4.5(c), tenemos que existe $\overline{R}$ tal que $L(\overline{R}) = \inf_{y\geq 0} L(y)$ y que $L_\alpha(\cdot) \downarrow L(\cdot)$ cuando $\alpha \uparrow 1$. De manera que

$$L_\alpha(\overline{R}) \geq L_\alpha(R_\alpha) \geq L(R_\alpha) \geq L(\overline{R}), \tag{3.40}$$

y entonces

$$\lim_{\alpha\uparrow 1} L_\alpha(\overline{R}) = \lim_{\alpha\uparrow 1} L_\alpha(R_\alpha) = \lim_{\alpha\uparrow 1} L(R_\alpha) = L(\overline{R}). \tag{3.41}$$

La última igualdad en (3.41) implica que el conjunto $\{R_\alpha : \alpha \in (0,1)\}$ es acotado, pues de no ser así, para alguna sucesión $\{\beta(n)\} \subset (0,1)$ tendríamos que $R_{\beta(n)} \to +\infty$ cuando $n \to \infty$, de modo que $\lim_{n\to\infty} L(R_{\beta(n)}) = L(\overline{R}) = +\infty$, lo cual es una contradicción. Por lo tanto, existe $\{\alpha(n)\} \subset (0,1)$ y una constante R^* para las cuales $\lim_{n\to\infty} R_{\alpha(n)} = R^*$, lo cual implica que para todo $x \in \mathbf{X}$:

$$\lim_{n\to\infty} f_{\alpha(n)}(x) = f^*(x) := \begin{cases} R^* - x & \text{si} \quad 0 \le x \le R^* \\ 0 & \text{si} \quad x > R^*. \end{cases}$$

Finalmente, de la continuidad de $L(\cdot)$ y (3.40), se concluye que

$$L(R^*) = \inf_{y\ge 0} L(y). \blacksquare$$

Para cada $\alpha \in (0,1)$, denotamos por $\{x_t^\alpha\}$ al proceso de Markov correspondiente a la política f_α y definimos los tiempos de paro

$$\tau_\alpha := \min\{t \ge 1 : x_t^\alpha = 0\} \text{ y } \sigma_\alpha := \min\{t \ge 1 : x_t^\alpha \le R_\alpha\} \tag{3.42}$$

El siguiente lema contiene un resultado referente a la funciones definidas en (3.10) con $z = 0$, es decir,

$$h_\alpha(x) := V_\alpha(x) - V_\alpha(0) \text{ y } \rho_\alpha := (1-\alpha)V_\alpha(0) \quad x \in \mathbf{X}, \alpha \in (0,1). \tag{3.43}$$

Lema 3.4.8. Si se satisface H3.4.1, entonces para cada $x \in \mathbf{X}$:

(a) $B_1(x) := \sup\left\{E_x^{f_\alpha}\sigma_\alpha : \alpha \in (0,1)\right\} < +\infty$;

(b) $B_2(x) := \sup\left\{E_x^{f_\alpha}\tau_\alpha : \alpha \in (0,1)\right\} < +\infty$;

(c) $|h_\alpha(x)| \le [\rho_\alpha - M]B_2(x)$, donde M es una cota inferior de la función de costo por etapa C introducida en (3.24).

Demostración del Lema 3.4.8(a). Sea $\alpha \in (0,1)$ arbitrario y fijo. Primero note que $E_x^{f_\alpha}\sigma_\alpha = 1 \ \forall x \le R_\alpha$. Ahora considere $x > R_\alpha$ y observe que $[\sigma_\alpha > n] = [x_n > R_\alpha] = [S_n < x - R_\alpha] \ \forall n \in \mathbf{N}$, donde $S_n := \sum_{k=0}^{n-1} w_k$. Entonces

$$
\begin{aligned}
E_x^{f_\alpha}\sigma_\alpha &= \sum_{n=1}^{\infty} P_x^{f_\alpha}[\sigma_\alpha \geq n] \qquad (3.44)\\
&= 1+\sum_{n=1}^{\infty} P_x^{f_\alpha}[\sigma_\alpha > n]\\
&= 1+\sum_{n=1}^{\infty} G_n(x-R_\alpha)\\
&\leq 1+\sum_{n=1}^{\infty} G_n(x), \qquad (3.45)
\end{aligned}
$$

donde $G_n(\cdot)$ denota la función de distribución de S_n. Por otro lado,

$$
\begin{aligned}
\frac{G_{n+1}(x)}{G_n(x)} &= P[S_{n+1} \leq x | S_n \leq x]\\
&= \int_0^x G(x-y)G_n(dy)\\
&\leq G_n(x) \to 0 \text{ cuando } n \to \infty, \qquad (3.46)
\end{aligned}
$$

lo cual implica que la serie en (3.45) es convergente. Por lo tanto,

$$
B_1(x) = \sup_{\alpha\in(0,1)} E_x^{f_\alpha}\sigma_\alpha \leq 1+\sum_{n=1}^{\infty} G_n(x) < +\infty \ \ \forall x \in \mathbf{X}.
$$

(b) Primero consideremos $x_0 = x \leq R_\alpha$, de manera que

$$
x_{t+1} = (R_\alpha - w_t)^+ \ \ \forall t \geq 0.
$$

Entonces

$$
\begin{aligned}
E_x^{f_\alpha}\tau_\alpha &= \sum_{n=1}^{\infty} nP_x^{f_\alpha}[\tau_\alpha = n]\\
&= [1-G(R_\alpha)\} \sum_{n=1}^{\infty} n[G(R_\alpha)]^{n-1}\\
&= \frac{1}{1-G(R_\alpha)}.
\end{aligned}
$$

Ahora consideremos $x_0 = x > R_\alpha$ y observe que

$$E_x^{f_\alpha}\tau_\alpha \leq \frac{1}{1-G(R_\alpha)} + E_x^{f_\alpha}\sigma_\alpha,$$

lo cual combinado con la parte (a) de este lema y que el conjunto $\{R_\alpha : \alpha \in (0,1)\}$ es acotado, implica que

$$B_2(x) = \sup_{\alpha\in(0,1)} E_x^{f_\alpha}\tau_\alpha < +\infty \;\; \forall x \in \mathbf{X}. \blacksquare$$

(c) Sea M una cota inferior de C, y defina

$$\overline{C}(x,a) := C(x,a) - M \;\text{ y }\; \overline{V}_\alpha(x) := V_\alpha(x) - (1-\alpha)^{-1}M$$

$\forall(x,a) \in \mathbf{K}$ y $\alpha \in (0,1)$. De la propiedad (fuerte) de Markov, tenemos

$$\begin{aligned}
\overline{V}_\alpha(x) &= E_x^{f_\alpha}\sum_{t=1}^{\infty}\alpha^t\overline{C}(x_t,a_t) \\
&= E_x^{f_\alpha}\sum_{t=1}^{\tau_\alpha-1}\alpha^t\overline{C}(x_t,a_t) + E_x^{f_\alpha}\sum_{t=\tau_\alpha}^{\infty}\alpha^t\overline{C}(x_t,a_t) \\
&\geq E_x^{f_\alpha}\sum_{t=\tau_\alpha}^{\infty}\alpha^t\overline{C}(x_t,a_t) \\
&= E_x^{f_\alpha}[\alpha^{\tau_\alpha}]\overline{V}_\alpha(0).
\end{aligned} \tag{3.47}$$

De la misma forma se obtiene

$$\overline{V}_\alpha(x) \leq E_x^{f_\alpha}\sum_{t=1}^{\tau_\alpha-1}\alpha^t\overline{C}(x_t,a_t) + E_x^{f_\alpha}[\alpha^{\tau_\alpha}]\overline{V}_\alpha(0). \tag{3.48}$$

Puesto que

$$1-\alpha^{\tau_\alpha} \leq (1-\alpha)\tau_\alpha \;\; \forall\alpha \in (0,1),$$

usando (3.47) se tiene que

$$
\begin{aligned}
h_\alpha(x) &\geq E_x^{f_\alpha}[\alpha^{\tau_\alpha} - 1]\overline{V}_\alpha(0) \\
&= -E_x^{f_\alpha}\left[\frac{1-\alpha^{\tau_\alpha}}{1-\alpha}\right](1-\alpha)\overline{V}_\alpha(0) \\
&\geq -E_x^{f_\alpha}[\tau_\alpha](1-\alpha)\overline{V}_\alpha(0) \\
&\geq -E_x^{f_\alpha}[\tau_\alpha][\rho_\alpha - M]. \qquad (3.49)
\end{aligned}
$$

Por otra parte, usando argumentos de "renovación", se obtiene

$$
J(f_\alpha, x) - M = \left[E_x^{f_\alpha}\tau_\alpha\right]^{-1} E_x^{f_\alpha}\sum_{t=0}^{\tau_\alpha - 1}\overline{C}(x_t, a_t), \quad \forall x \in \mathbf{X}, \qquad (3.50)
$$

mientras que con cálculos directos se obtiene

$$
J(f_\alpha, x) = L(K_\alpha), \quad \forall x \in \mathbf{X}. \qquad (3.51)
$$

Puesto que $\alpha^{\tau_\alpha} < 1$, (3.48), (3.50) y (3.51) implican que

$$
\begin{aligned}
\overline{V}_\alpha(x) &\leq \left[E_x^{f_\alpha}\tau_\alpha\right][J(f_\alpha, x) - M] + \overline{V}_\alpha(0) \\
&= \left[E_x^{f_\alpha}\tau_\alpha\right][L(R_\alpha) - M] + \overline{V}_\alpha(0).
\end{aligned}
$$

Entonces, como $L_\alpha(\cdot) \geq L(\cdot)$, tenemos

$$
\begin{aligned}
h_\alpha(x) &\leq \left[E_x^{f_\alpha}\tau_\alpha\right][L_\alpha(R_\alpha) - M] \\
&= \left[E_x^{f_\alpha}\tau_\alpha\right][\rho_\alpha - M]. \qquad (3.52)
\end{aligned}
$$

Por lo tanto, (3.49) y (3.52) demuestran que

$$|h_\alpha(x)| \leq \left[E_x^{f_\alpha}\tau_\alpha\right][\rho_\alpha - M] \leq B_2(x)[\rho_\alpha - M]. \blacksquare$$

Demostración del Teorema 3.4.3. Para demostrar que la política f^* definida en (3.30)-(3.31) es óptima en costo promedio, de las Observaciones 3.3.2 y 3.4.5, sólo hace falta verificar que tanto H3.3.1(a)-(c) como la condición (3.16) [en el Teorema 3.3.3(b)] se cumplen para el modelo de inventarios.

Cálculos directos muestran que $\rho_\alpha = (1-\alpha)V_\alpha(f_\alpha, 0) = L(R_\alpha)$ $\forall \alpha \in (0,1)$, de manera que por el Teorema 3.4.6 (c) y (3.41), tenemos que

$$\rho^* = L(R^*) = \lim_{\alpha \to 1} \rho_\alpha.$$

Entonces, H3.3.1(a) se cumple, es decir$\{\rho_\alpha : \alpha \in (0,1)\}$ es un conjunto acotado.

Ahora, tomemos $\{\alpha(n)\} \subset (0,1)$ como en el Teorema 3.4.6(c) y definamos

$$N(x) := K \sup_n E_x^{f_{\alpha(n)}} \tau_{\alpha(n)}, \quad x \in \mathbf{X}, \tag{3.53}$$

donde K es una cota superior del conjunto $\{\rho_\alpha - M : \alpha \in (0,1)\}$. Del Lema 3.4.9 se sigue que

$$|h_{\alpha(n)}(x)| \leq N(x) < +\infty \ \forall x \in \mathbf{X},\ n \in N_0;$$

es decir, H3.3.1(b) se cumple. Además, note que

$$\int_{\mathbf{X}} N(y)Q(dy|x,a) = \int_0^{x+a} N(r)\mu(x+a-r)dr + N(0)[1 - G(x+a)],$$

para todo par $(x,a) \in \mathbf{K}$, lo cual combinado con H3.4.1(b), implica que H3.3.1(c) se satisface.

Usando los mismos argumentos dados en la demostración del Teorema 3.3.3(a), se concluye que la terna (ρ^*, f^*, h^*) satisface

$$\rho^* + h^*(x) \geq C(x, f^*) + \int_{\mathbf{X}} h^*(y)Q(dy|x, f^*) \ \forall x \in \mathbf{X},$$

donde ρ^*, f^* son como en el Teorema 3.4.3 y

$$h^*(\cdot) := \liminf_{n\to\infty} h_{\alpha(n)}(\cdot).$$

Finalmente, sólo resta verificar que la política f^* y la función (3.53) satisfacen la condición (3.16) del Teorema 3.3.3, es decir,

$$\lim_{n\to\infty} \frac{1}{n} E_x^{f^*} N(x_n) = 0 \ \ \forall x \in \mathbf{X}.$$

Para ver que esto efectivamente ocurre, primero note que

$$Q^*(B) := \int_B [(R^* - w)^{-1}] G(dw), \quad B \in \mathcal{B}(\mathbf{X}),$$

es la medida invariante de el proceso $\{x_n^*\}$ inducida por la política f^*, y que dicha medida se concentra en el intervalo $[0, K^*]$. Entonces,

$$\int_{\mathbf{X}} N(y) Q^*(dy) < +\infty;$$

por lo tanto (3.16) se cumple. ■

3.5 Conclusiones

En este capítulo se demostró la existencia de políticas óptimas en costo promedio por medio del enfoque de *Aproximaciones por Problemas Descontados* (APD) en contextos más generales a los considerados en trabajos previos (ver, por ejemplo, [31], [33], [54], [70]), en los cuales se supone que una o más de las siguientes condiciones se satisfacen: (i) $A(x)$ es compacto para cada $x \in \mathbf{X}$; (ii) $-N \leq h_\alpha(\cdot)\ \forall \alpha \in (0,1)$, para alguna constante $N \geq 0$, (iii) $h_\alpha(\cdot) \leq B(\cdot)\ \forall\alpha \in (0,1)$, para alguna función no negativa $B(\cdot)$.

Por otra parte, el análisis que se presentó del enfoque APD permitió resolver *explícitamente* el problema en costo promedio para un sistema de inventarios con controles no acotados, lo cual es poco frecuente en la literatura.

Capítulo 4

Indice en Costo Promedio: Condiciones de Estabilidad

4.1 Introducción

Ahora estudiaremos el PCO en costo promedio pero, a diferiencia del capítulo anterior, las condciones que se consideran se imponen directamente sobre el modelo de control, excepto la condición H4.3.8. Estas condiciones garantizarán que los procesos controlados tienen buenas propiedades de estabilidad, las cuales a su vez, permitirán mostrar que existe una solución de la EOCP. Esto último se hace en dos etapas: primero se muestra que existe una política estacionaria óptima en costo promedio en la clase de las políticas estacionarias por medio del enfoque de *Aproximaciones por Problemas Descontados*; posteriormente, se obtiene una solución de la EOCP usando el *Algoritmo de Iteración de Políticas* con la particularidad de que la política de "inicialización" es una política óptima en costo promedio en la clase de las políticas estacionarias.

Para procesos de Markov (controlados o no controlados) las propiedades de estabilidad generalmente se expresan como propiedades de recurrencia o ergodicidad y su uso abunda en la literatura (para procesos no controlados ver, e.g., [50]; para procesos controlados ver, e.g., [12], [17], [28], [36], [40], [42], [56]), pero en la mayor parte se considera espacios numerables o costos acotados. Por otra parte, se sabe que bajo ciertas condiciones, la existencia de soluciones de la EOCP es "equivalente" a que los procesos controlados posean propiedades muy fuertes de recurrencia ([12], [36]).

Uno de los caminos más importantes para establecer la estabilidad de sistemas determinísticos o estocásticos es por medio del uso de *funciones de Lyapunov*. Generalmente, se atribuye a Foster ([23]) la introducción de este tipo de funciones para el análisis de sistemas estocásticos no controlados y a Hordijk ([40]) para el caso de sistemas estocásticos controlados.

En este capítulo estudiaremos el PCO en costo promedio suponiendo que el modelo de control satisface una condición del tipo *Lyapunov* que garantizará, por una parte, que los procesos de Markov asociados a las políticas estacionarias son *geométricamente ergódicos* ([50])— es decir, que se "estabilizan" con rapidez geométrica— y, por otra parte, la existencia de soluciones a la *Ecuación de Poisson* ([27]) correspondiente a cada uno de dichos procesos. Los resultados principales se encuentran en el Teorema 4.5.2, el cual establece la existencia de una solución de la EOCP y de una política estacionaria óptima en costo promedio, y que las afirmaciones "$f^* \in \mathbf{F}$ es óptima en costo promedio", "f^* es Flynn-óptima" y " f^* es fuertemente óptima", son equivalentes.

El resto del capítulo está organizado de la siguiente forma. En la Sección 4.2 se presentan los criterios de optimalidad. En la Sección 4.3 se introduce una *condición del tipo Lyapunov* y se discute su relación con propiedades de ergodicidad geométrica y otras consecuencias importantes. La Sección 4.4 contiene únicamente resultados técnicos, y la Sección 4.5 contiene los resultados principales del capítulo. Para finalizar, en la Sección 4.6 se presenta un sistema de inventarios, variante del introducido en el capítulo anterior, que satisface las condiciones que se usan en este capítulo (ver H4.3.1, H4.3.2 y H4.3.8). Tanto los resultados como la presentación que se hace de ellos se tomaron de [38].

4.2 Indices en costo promedio

Dado que en este capítulo consideraremos varios criterios de optimalidad en "costo promedio", algunas de las definiciones introducidas previamente se repetiran de nuevo aquí con el ánimo de facilitar la lectura y evitar posibles confusiones.

Para cada $n \in \mathbf{N}$, el *costo esperado en N-etapas* cuando se usa la política $\pi \in \Pi$, dado que el estado inicial del sistema es $x_0 = x \in \mathbf{X}$, está dado por

$$J_N(\pi, x) := E_x^\pi \sum_{t=0}^{N-1} C(x_t, a_t), \tag{4.1}$$

y el *costo promedio* (por etapa) se define como

$$J(\pi, x) := \limsup_{n\to\infty} \frac{1}{n} J_N(\pi, x) \tag{4.2}$$

Además, la funciones de costo óptimo correspondientes son

$$J_N^*(x) := \inf_{\pi\in\Pi} J_N(\pi, x) \ \text{ y } \ J^*(x) := \inf_{\pi\in\Pi} J(\pi, x), \quad x \in \mathbf{X}. \tag{4.3}$$

Definición 4.2.1. Una política $\pi^* \in \Pi$ es :

(a) *óptima en costo promedio* si

$$J^*(x) = J(\pi^*, x) \ \ \forall x \in \mathbf{X}; \tag{4.4}$$

(b) *óptima en el sentido de Flynn*, o brevemente, *Flynn-óptima* si

$$\lim_{N\to\infty} \frac{1}{N} \left[J_N(\pi^*, x) - J_N^*(x)\right] = 0 \ \ \forall x \in \mathbf{X}; \tag{4.5}$$

(c) *fuertemente* (F-) *óptima* si

$$J(\pi^*, x) \leq \liminf_{N\to\infty} \frac{1}{N} J_N(\pi^*, x) \ \ \forall x \in \mathbf{X}. \tag{4.6}$$

Observación 4.2.2(a) Es fácil verificar que si una política es Flynn-óptima o fuertemente óptima, entonces es óptima en costo promedio;

(b) Por otra parte, si π^* es Flynn-óptima y

$$J(\pi^*, x) = \lim_{N\to\infty} \frac{1}{N} J_N(\pi^*, x) \ \ \forall x \in \mathbf{X},$$

entonces π^* es fuertemente óptima. En este caso, de (a) se sigue que se cumplen las siguientes implicaciones:

$$\pi^* \text{ es Flynn-ópt.} \Rightarrow \ \pi^* \text{ es F-ópt.} \Rightarrow \ \pi^* \text{ es ópt. en costo promedio.}$$

4.3 Condición de Lyapunov y ergodicidad geométrica

Consideraremos un modelo $(\mathbf{X}, \mathbf{A}, \{A(x) : x \in \mathbf{X}\}, Q, C\}$ de acuerdo a la Definición 2.2.1 y adicionalmente suponemos que $\mathbf{X}$ es un *espacio localmente compacto.*

Ahora introducimos las condiciones de compacidad y continuidad que usaremos en nuestros desarrollos.

H4.3.1. Para cada $x \in \mathbf{X}$:

(a) $A(x)$ es un subconjunto compacto de $\mathbf{A}$;

(b) $C(x, \cdot)$ is semicontinua inferiormente en $A(x)$;

(c) $Q(\cdot|x, a)$ es fuertemente continua en $a \in A(x)$, es decir, la función

$$a \mapsto \int_{\mathbf{X}} u(y) Q(dy|x, a)$$

es continua para cada función medible acotada $u \in M_a(\mathbf{X})$;

(d) existe una función medible $V \geq 1$ definida en $\mathbf{X}$ tal que

(d.1) la función $a \mapsto \int_{\mathbf{X}} V(y) Q(dy|x, a)$ es continua;

(d.2) $|C(x, a)| \leq V(x) \ \ \forall a \in A(x)$.

También suponempos que el modelo de control satisface la siguiente *condición del tipo Lyapunov* ([50], [27]).

H4.3.2. Para cada $f \in \mathbf{F}$:

(a) (*Condición de Lyapunov*) existe un subconjunto *petite* C_f de $\mathbf{X}$, constantes $b_f < +\infty$ and $B_f < 1$ tales que

$$\int_{\mathbf{X}} V(y)Q(dy|x,f) \leq B_f V(x) + b_f \mathbf{I}_{C_f}(x) \quad \forall x \in \mathbf{X}, \tag{4.7}$$

donde $\mathbf{I}_{C_f}(\cdot)$ denota la función indicadora del conjunto C_f;

(b) existe una medida σ-finita ψ, tal que el proceso de Markov $\{x_t\}$—inducido por la política f—es ψ-irreducible y aperiódica.

Definición 4.3.3. Sea $V(\cdot) \geq 1$ la función en H4.3.1(d). Denotaremos por L_V^∞ al espacio de Banach formado por todas las funciones medibles $u : \mathbf{X} \to \mathbf{R}$ tales que

$$||u||_V := \sup_{x \in \mathbf{X}} \frac{|u(x)|}{V(x)} < +\infty.$$

El siguiente resultado es esencial en los desarrollos posteriores. La demostración se proporciona en [50].

Teorema 4.3.4.(*V-Ergodicidad Geométrica*). Si se satisface H4.3.2, entonces para cada $f \in \mathbf{F}$:

(a) el proceso de Markov $\{x_t\}$, correspondiente a la política f, es *Harris (recurrente) positivo.* Denotamos por Q_f a la única medida de probabilidad invariante;

(b) $\int_{\mathbf{X}} V dQ_f < +\infty$;

(c) el proceso de Markov $\{x_t\}$ es *V-uniformente ergódico*, es decir, existen constantes $\gamma_f < 1$ y $M_f < +\infty$ tales que

$$\left| \int_{\mathbf{X}} u(y)Q^t(dy|x,f) - \int_{\mathbf{X}} u(y)Q_f(dy) \right| \leq ||u||_V M_f \gamma_f^t V(x), \tag{4.8}$$

para todo $x \in \mathbf{X}, t \in \mathbf{N}_0$ y $u \in L_V^\infty$.

Observación 4.3.5(a) Es importante notar que para cada $f \in \mathbf{F}$, bajo las hipótesis de (ψ–) irreducibilidad y aperiodicidad, la condiciones (4.7) y (4.8) son *equivalentes* ([50], Theorem 16.0.1, p. 383).

(b) Un caso particular e importante de H4.3.2, especialmente cuando se considera problemas con *costos acotados*, resulta al suponer que V es una función acotada. En este caso, el espacio L_V^∞ es el *espacio de las funciones acotadas* y la $V-$ergodicidad uniforme (4.8) se reduce a la condición usual de *ergodicidad uniforme*

$$||Q^n(\cdot|x,f) - Q_f(\cdot)||_{TV} \leq M_f\gamma_f^n \quad \forall n \in \mathbf{N},$$

donde $||\cdot||_{TV}$ es la *norma de la variación total* para medidas con signo. Además, para cada política estacionaria $f \in \mathbf{F}$, las siguientes condiciones son equivalentes ([50], Theorem 16.0.2, p. 384):

(i) el proceso $\{x_t\}$ es uniformemente ergódico;

(ii) el proceso $\{x_t\}$ es aperiódico y existe una solución acotada $V \geq 1$ a la desigualdad (4.7);

(iii) el proceso $\{x_t\}$ es aperiódico y la *Condición de Doeblin* se cumple; es decir, existe una medida de probabilidad $\lambda_f(\cdot)$ en $\mathcal{B}(\mathbf{X})$, constantes positivas $\varepsilon < 1$, δ, y un número natural m tales que

$$\inf_{x\in\mathbf{X}} Q(B|x,f) > \delta \quad \text{siempre que} \quad \lambda_f(B) > \varepsilon.$$

En la siguiente proposición se proporciona otra consecuencia importante de H4.3.2. En el enunciado hacemos uso de la siguiente notación: para cada $f \in \mathbf{F}$,

$$J(f) \quad : \quad = \int_{\mathbf{X}} C(y,f)Q_f(dy), \tag{4.9}$$

$$\widehat{h}_f(x) \quad : \quad = E_x^f \sum_{t=0}^{\infty}[C(x_t,f) - J(f)], \quad \forall x \in \mathbf{X}. \tag{4.10}$$

Note que (4.8) garantiza que la función $\widehat{h}_f(\cdot)$ en (4.10) está bien definida y que pertenece a L_V^∞ (c.f. Observación 4.39).

Proposición 4.3.6. Si H4.3.2 se satisface, entonces para cada $f \in \mathbf{F}$, el par $(J(f), \widehat{h}_f)$ satisface la *Ecuación de Poisson* (correspondiente a f)

$$J(f) + \widehat{h}_f(x) = C(x, f) + \int_{\mathbf{X}} \widehat{h}_f((y)Q(dy|x, f) \quad \forall x \in \mathbf{X}. \tag{4.11}$$

Demostración de la Proposición 4.3.6. La demostración se sigue directamente de (4.9)-(4.10) y la Propiedad de Markov.

Observación 4.3.7(a) Note que iterando (4.11) se obtiene

$$nJ(f) + \widehat{h}_f(x) = J_n(f, x) + E_x^f \widehat{h}_f(x_n) \quad \forall x \in \mathbf{X}, n \in \mathbf{N}, \tag{4.12}$$

lo cual implica que

$$J(f, x) = \lim_{n\to\infty} \frac{1}{n} J_n(f, x) = J(f) = \int_{\mathbf{X}} C(y, f)Q_f(dy) \quad \forall x \in \mathbf{X}. \tag{4.13}$$

De hecho, las dos primeras igualdades se siguen de (4.12) y (4.8) al multiplicar por $\frac{1}{n}$ y tomar límite cuando $n \to \infty$. La tercera igualdad se obtiene integrando (4.12) con respecto a la medida de probabilidad invariante $Q_f(\cdot)$.

(b) Por otra parte, observe que para cada constante k, la función $h_f^k(\cdot) := \widehat{h}_f(\cdot) + k$ también es solución de la Ecuación de Poisson. Reciprocamente, si $h_1, h_2 \in L_V^\infty$ son soluciones de la Ecuación de Poisson correspondiente a una política f, es decir,

$$J(f) + h_i(x) = C(x, f) + \int_{\mathbf{X}} h_i(y)Q(dy|x, f) \quad \forall x \in \mathbf{X}, i = 1, 2, \tag{4.14}$$

entonces $H(\cdot) := h_1(\cdot) - h_2(\cdot)$ es una función constante. De hecho, de (4.14) se tiene que $H(\cdot)$ es una *función armónica*, es decir,

$$H(x) = \int_{\mathbf{X}} H(y)Q(dy|x, f) \quad \forall x \in \mathbf{X},$$

lo cual implica que

$$H(x) = \int_{\mathbf{X}} H(y)Q^n(dy|x, f) \quad \forall x \in \mathbf{X}, n \in \mathbf{N}.$$

Entonces, de (4.8) se obtiene

$$H(x) = \int_{\mathbf{X}} H(y) Q_f(dy) \quad \forall x \in \mathbf{X}.$$

(c) De (a), tenemos que las soluciones a la Ecuación de Poisson *no son únicas*. Sin embargo, pidiendo que las soluciones a la Ecuaciones de Poisson satisfagan una condición adicional es posible garantizar la unicidad. Por ejemplo, la función $\widehat{h}_f(\cdot)$ definida en (4.9)-(4.10)es la única solución de la *Ecuación de Poisson* (correspondiente a la política $f \in \mathbf{F}$), que satisface la condición

$$\int_{\mathbf{X}} \widehat{h}_f(y) Q_f(dy) = 0. \tag{4.15}$$

También es fácil verificar que la función

$$h_f^z(x) := \widehat{h}_f(x) - \widehat{h}_f(z), \quad x \in \mathbf{X}, \tag{4.16}$$

donde $z \in \mathbf{X}$ es un estado fijo arbitrario, es la *única solución* que se anula en z, es decir, $h_f^z(z) = 0$. Finalmente note que

$$\widehat{h}_f(x) = h_f^z(x) - \int_{\mathbf{X}} h_f^z(y) Q_f(dy), \quad x \in \mathbf{X}. \tag{4.17}$$

La última hipótesis general que se usará en este capítulo involucra a las constantes en (4.8) y se introduce a continuación.

H4.3.8. Las constantes $M := \sup_f M_f$ y $\gamma := \sup_f \gamma_f$ satisfacen lo siguiente:

$$M < +\infty \quad \text{and} \quad \gamma < 1. \tag{4.18}$$

En la Sección 4.5 se presenta una variante del ejemplo de inventarios estudiado en el capítulo anterior, el cual satisface las condiciones en H4.3.1, H4.3.2 y H4.3.8.

La propiedades "ergódicas" clave para los desarrollos que se presentarán en el resto del capítulo son las propiedades (4.8), (4.11) y (4.18). En [30], [36], y [56] se presentan otros enfoques para obtener estas propiedades para costos acotados, y en [28], [49] para costos no acotados. Es importante mencionar que en las dos últimas referencias se imponen condiciones sobre la función de costo mucho más restrictivas que la nuestras, que pueden catalogarse como de uso estándar en la literatura relacionada. Por ejemplo, además de las condiciones de estabilidad, en [28] se supone que la función de costo y la ley de transición satisfacen una condición *"tipo Lipschitz"*, mientras que en [49] se supone que la función de costo es *estrictamente no acotada*, lo cual de entrada ya penaliza los comportamientos inestables de los procesos controlados (c.f. Capítulo 6). Por otra parte, es posible demostrar que bajo ciertas hipótesis las condiciones de estabilidad en [28] implican a H4.3.2. De hecho, en la Sección 4.5 se "demuestra" ésto para el caso especifico de un sistema de inventarios.

Observación 4.3.9. Considere las funciones (4.10) y (4.16). Note que H4.3.8 implica que

$$\sup_{f\in\mathbf{F}} ||\widehat{h}_f||_V \leq M(1-\gamma)^{-1},$$

lo cual a su vez implica que

$$\sup_{f\in\mathbf{F}} ||h_f^z||_V \leq M(1-\gamma)^{-1}.$$

Observación 4.3.10. Sean ψ y Q_f, $f\in\mathbf{F}$,como en H4.3.2(a) y Teorema 4.3.4(a), respectivamente. De la Observación A.D.9, se tiene que $\psi \ll Q_f$, para cada $f\in\mathbf{F}$, es decir, $Q_f(B)=0$ implica que $\psi(B)=0$.

4.4 Resultados preliminares

Para cada $u\in L_V^\infty$ se define la función

$$Tu(x) := \inf_{a\in A(x)} \left[C(x,a) + \int_{\mathbf{X}} u(y)Q(dy|x,a) \right], \quad x\in\mathbf{X}, \tag{4.19}$$

donde $u\in L_V^\infty$. En el siguiente teorema se demuestra que bajo las condiciones H4.3.1 y H4.3.2(a) existe un selector que alcanza el ínfimo en el lado derecho de (4.19), de manera que el operador T manda el espacio L_V^∞ en si mismo, es decir, $T(L_V^\infty)\subset L_V^\infty$.

Proposición 4.4.1. Suponga que se satisface H4.3.1 y H4.3.2(a). Entonces,

(a) Para cada $u \in L_V^\infty$ existe un selector $f \in \mathbf{F}$—que depende de u—tal que

$$Tu(x) = C(x, f) + \int_{\mathbf{X}} u(y)Q(dy|x, f) \quad \forall x \in \mathbf{X}; \tag{4.20}$$

(b) $T(L_V^\infty) \subset L_V^\infty$;

(c) existe $C \in \mathcal{B}(\mathbf{X})$, constantes $B < 1$ y $b < +\infty$ tales que

$$\int_{\mathbf{X}} V(y)Q(dy|x, a) \leq BV(x) + b\mathbf{I}_C(x) \quad \forall (x, a) \in \mathbf{K}. \tag{4.21}$$

Demostración de la Proposición 4.4.1. Sea $u \in L_V^\infty$ arbitraria fija y defina $m := ||u||_V$.

(a) Considere la función

$$u_V(x) := u(x) + mV(x)$$

y observe que dicha función es nonegativa y que pertenece al espacio L_V^∞, de manera que existe una sucesión $\{u_V^n\}$ de funciones medibles y acotadas tales que $u_V^n(\cdot) \uparrow u_V(\cdot)$ cuando n tiende a infinito. De H4.3.1(c), del Teorema de Convergencia Monótona se concluye que la función

$$\int_{\mathbf{X}} u_V(y)Q(dy|x, \cdot)$$

es semicontinua inferiormente sobre $A(x)$, para cada $x \in \mathbf{X}$. Combinando ésto con H4.3.1(b) y (d.1), se observa que

$$C(x, \cdot) + \int_{\mathbf{X}} u(y)Q(dy|x, \cdot) = C(x, \cdot) + \int_{\mathbf{X}} u_V(y)Q(dy|x, \cdot) - m\int_{\mathbf{X}} V(y)Q(dy|x, \cdot)$$

es semicontinua inferiormente para cada $x \in \mathbf{X}$. Finalmente, de H4.3.1(a) y del Teorema de Selección (Teorema A.B.5), se concluye que existe $f \in \mathbf{F}$ tal que

$$Tu(x) = C(x, f) + \int_{\mathbf{X}} u(y)Q(dy|x, f) \quad \forall x \in \mathbf{X}.$$

(b) Puesto que

$$|\int_{\mathbf{X}} u(y)Q(dy|x,a)| \leq m\int_{\mathbf{X}} V(y)Q(dy|x,a) \;\; \forall (x,a) \in \mathbf{K},$$

de (a), H4.3.1(a) y H4.3.2(a), se concluye que $Tu(\cdot) \in L_V^\infty$;

(c) De H4.3.1(a), (d.1) y aplicando de nuevo el Teorema de Selección, existe $g \in \mathbf{F}$ tal que

$$\sup_{a \in A(x)} \int_{\mathbf{X}} V(y)Q(dy|x,a) = \int_{\mathbf{X}} V(y)Q(dy|x,g) \;\; \forall x \in \mathbf{X},$$

lo cual implica que (4.21) se cumple con $B = B_g$, $b = b_g$ y $C = C_g$. ■

Proposición 4.4.2. Suponga que se satisfacen H4.3.1 y H4.3.2 y sean b, B las constantes de la Proposición 4.4.1(c). Entonces, para cada $x \in \mathbf{X}$, $\pi \in \Pi$, $u \in L_V^\infty$ y $n \in \mathbf{N}$ las propiedades siguientes se cumplen:

(a) $E_x^\pi V(x_n) \leq B^n V(x) + b\sum_{k=0}^{n-1} B^k$

(b) $E_x^\pi |u(x_n)| \leq ||u||_V \left[B^n V(x) + b\sum_{k=0}^{n-1} B^k\right] \;\; \forall n \in \mathbf{N}$;

(c) $\limsup_{n\to\infty} \left[\sup_{\pi\in\Pi} E_x^\pi |u(x_n)|\right] \leq b||u||_V (1-B)^{-1}$;

(d) $\lim_{n\to\infty} \frac{1}{n} \left[\sup_{\pi\in\Pi} E_x^\pi |u(x_n)|\right] = 0.$

Demostración de la Proposición 4.4.2. Observe que (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (d). Por otra parte, (a) se obtiene iterando (4.21). ■

Recuerde que para cada factor descuento $\alpha \in (0,1)$, el costo α-descontado generado por la aplicación de una política $\pi \in \Pi$, dado que el estado inicial es $x_0 = x \in \mathbf{X}$, está dado por

$$V_\alpha(\pi, x) = E_x^\pi \sum_{t=0}^{\infty} \alpha^t C(x_t, a_t); \tag{4.22}$$

mientras que la función de costo óptimo α-descontado es

$$V_\alpha(x) = \inf_{\pi\in\Pi} V_\alpha(\pi, x). \tag{4.23}$$

Ahora definamos para cada $x \in \mathbf{X}$ y $\pi \in \Pi$:

$$V_\alpha^n(\pi,x) := E_x^\pi \sum_{t=0}^{n-1} \alpha^t C(x_t,a_t) \quad \text{y} \quad V_\alpha^n(x) := \inf_{\pi\in\Pi} V_\alpha^n(\pi,x), \quad n\in\mathbf{N}. \tag{4.24}$$

Del *Algoritmo de Programación Dinámica* ([34], Chapter 3, Section 3.4, p. 31), para cada $n\in\mathbf{N}$ y $\alpha\in(0,1)$

$$\begin{aligned} V_\alpha^{n+1}(x) &= TV_\alpha^n(x) \quad \forall x\in\mathbf{X}, \\ &= \inf_{a\in A(x)} \left[C(x,a) + \alpha \int_{\mathbf{X}} V_\alpha^n(y) Q(dy|x,a) \right], \end{aligned} \tag{4.25}$$

donde $V_\alpha^0(\cdot)\equiv 0$ y T es el operador definido en (4.19).

Teorema 4.4.3. Suponga que se satisfacen H4.3.1 y H4.3.2(a). Para cada $\alpha\in(0,1)$ se cumple lo siguiente:

(a) las funciones $V_\alpha^n(\cdot)$ y $V_\alpha(\cdot)$ son medibles y existe una constante K_α tal que

$$||V_\alpha^n - V_\alpha||_V \le K_\alpha \alpha^n \quad \forall n\in\mathbf{N}; \tag{4.26}$$

(b) $V_\alpha(\cdot)$ satisface la *Ecuación de Optimalidad en Costo α-Descontado* ($\alpha-$EOCD)

$$V_\alpha(x) = \min_{a\in A(x)} \left[C(x,a) + \alpha \int_{\mathbf{X}} V_\alpha(y) Q(dy|x,a) \right] \quad \forall x\in\mathbf{X}; \tag{4.27}$$

(c) existe una politica estacionaria $f_\alpha\in\mathbf{F}$ tal que

$$V_\alpha(x) = C(x,f_\alpha) + \alpha \int_{\mathbf{X}} V_\alpha(y) Q(dy|x,f_\alpha) \quad \forall x\in\mathbf{X}; \tag{4.28}$$

además, f_α es *óptima en costo α-descontado*, i.e.,

$$V_\alpha(x) = V_\alpha(f_\alpha,x) \quad \forall x\in\mathbf{X}; \tag{4.29}$$

reciprocamente,

(d) si $f\in\mathbf{F}$ es óptima en costo α-descontado, entonces satisface (4.28).

Demostración del Teorema 4.4.3(a) Note que

$$V_\alpha(\pi,x) = V_\alpha^n(\pi,x) + \alpha^n E_x^\pi \sum_{t=n}^{\infty} C(x_t,a_t),$$

para cada $x \in \mathbf{X}$, $\pi \in \pi$ y $n \in \mathbf{N}_0$. Entonces, de H4.4.3(a) y la Proposición 4.4.2, se tiene que

$$\begin{aligned}
|V_\alpha^n(x) - V_\alpha(x)| &\leq \sup_{\Pi} |V_\alpha^n(\pi,x) - V_\alpha(\pi,x) \\
&\leq \alpha^n \sup_{\Pi} E_x^\pi \sum_{t=n}^{\infty} \alpha^{t-n} |C(x_t,a_t)| \\
&\leq K_\alpha \alpha^n V(x),
\end{aligned}$$

donde b, B son las constantes de la 4.4.1(c) y $K_\alpha := [(1-B)(1-\alpha)]^{-1} b$. La medibilidad de las funciones $V_\alpha^n(\cdot)$ y $V_\alpha(\cdot)$ se sigue de la Proposición 4.4.1(a), (4.25) y (4.26).

(b) De (4.25) y (4.26) y usando el Lema de Fatou, se deduce que

$$V_\alpha(x) \geq \min_{a \in A(x)} \left[C(x,a) + \alpha \int_{\mathbf{X}} V_\alpha(y) Q(dy|x,a) \right] \quad \forall x \in \mathbf{X}.$$

La desigualdad en sentido contrario se obtiene de forma similar, notando que

$$V_\alpha^{n+1}(x) \leq C(x,a) + \int_{\mathbf{X}} V_\alpha^n(y) Q(dy|x,a) \quad \forall (x,a) \in \mathbf{K}.$$

Para demostrar (c) y (d), note que de Proposición 4.4.2(a) se tiene que

$$|V_\alpha(\pi,x)| \leq [(1-B)(1-\alpha)]^{-1} b V(x) \quad \forall x \in \mathbf{X}, \pi \in \Pi, \tag{4.30}$$

lo cual implica

$$|V_\alpha(x)| \leq [(1-B)(1-\alpha)]^{-1} b V(x) \quad \forall x \in \mathbf{X}, \pi \in \Pi. \tag{4.31}$$

Entonces, de la Proposición 4.4.2(c),

$$\lim_{n\to\infty} \alpha^n E_x^\pi |V_\alpha(\pi, x_n)| = \lim_{n\to\infty} \alpha^n E_x^\pi |V_\alpha(x_n)| = 0 \quad \forall x \in \mathbf{X}, \pi \in \Pi. \tag{4.32}$$

Puesto que la demostración de (c) y (d) es estándar ([34], Chapter 4), la omitimos. ■

Observación 4.4. Sean $\alpha \in (0,1)$ y $f \in \mathbf{F}$ arbitrarios. Note que para cad a $x \in \mathbf{X}$ se cumplen las siguientes desigualdades:

$$\begin{aligned}
|(1-\alpha)V_\alpha(f,x) - J(f)| &\leq (1-\alpha)\sum_{t=0}^{+\infty} \alpha^t |E_x^f C(x_t, f) - J(f)| \\
&\leq (1-\alpha)V(x)\sum_{t=0}^{+\infty} \alpha^t \gamma^t \\
&= \frac{(1-\alpha)V(x)}{1-\alpha\gamma}.
\end{aligned}$$

Entonces,

$$\lim_{\alpha\to 1^-} (1-\alpha)V_\alpha(f,x) = J(f) \quad \forall x \in \mathbf{X}; \tag{4.33}$$

por lo tanto,

$$\inf_{\mathbf{F}} J(f) \geq \limsup_{\alpha\to 1^-} (1-\alpha)V_\alpha(x) \quad \forall x \in \mathbf{X}. \tag{4.34}$$

4.5 La ecuación de optimalidad en costo promedio

Definición 4.5.1. Sea ρ una constante, $f \in \mathbf{F}$ y $h : \mathbf{X} \to \mathbb{R}$ una función medible.

(a) Diremos que la terna (ρ, f, h) es una solución de la *Desigualdad de Optimalidad en Costo Promedio* (DOCP) si

$$\rho + h(x) \geq \min_{a\in A(x)} \left[C(x,a) + \int_{\mathbf{X}} h(y)Q(dy|x,a)\right] \quad \forall x \in \mathbf{X} \tag{4.35}$$

$$= C(x,f) + \int_{\mathbf{X}} h(y)Q(dy|x,f). \tag{4.36}$$

(b) Si (ρ, f, h) satisface la relación (4.35)-(4.36) con igualdad diremos que es una solución de la *Ecuación de Optimalidad en Costo Promedio* (EOCP). En este caso, diremos que la política estacionaria f y la función h pertenecen a (o provienen de) una solución de la EOCP.

En el siguiente teorema mostramos que bajo H4.3.1, H4.3.2 y H4.3.8 existe una solución (ρ^*, f^*, h^*) de la EOCP, donde $h^*(\cdot)$ es una función en L_V^∞. También mostraremos que las afirmaciones "f es *óptima en costo promedio*", "f es *fuertemente óptima*", "f es *Flynn-óptima*" son equivalentes.

Teorema 4.5.2. Supongamos que se satisfacen H4.3.1, H4.3.2 y H4.3.8. Entonces:
(a) existe una terna (ρ^*, f^*, h^*), con $h^*(\cdot)$ en L_V^∞, que satisface la EOCP; es decir,

$$\begin{aligned}\rho^* + h^*(x) &= \min_{a\in A(x)}\left[C(x,a) + \int_{\mathbf{X}} h^*(y)Q(dy|x,a)\right] \quad \forall x \in \mathbf{X} \\ &= C(x,f^*) + \int_{\mathbf{X}} h^*(y)Q(dy|x,f^*).\end{aligned}$$

Además, f^* es *óptima en costo promedio* y el valor óptimo en costo promedio es ρ^*, es decir,

$$J^*(x) = J(f^*, x) = \rho^* \quad \forall \in \mathbf{X}.$$

(b) Para cada $f \in \mathbf{F}$, las siguientes afirmaciones son equivalentes: **(i)** f es Flynn-óptima; **(ii)** f es F-óptima; **(iii)** f es óptima en costo promedio. Por lo tanto, la política f^* en (a) es Flynn-óptima;
(c) Si $f \in \mathbf{F}$ es una política óptima en costo promedio, entonces (ρ^*, f, h^*) satisface la EOCP Q_f−casi dondequiera, donde $Q_f(\cdot)$ es la medida de probabilidad invariante correspondientententes a f.

Para demostrar la existencia de una solución de la EOCP, primero demostraremos que existe una política estacionaria "óptima" en costo promedio en la clase $\mathbf{F}$ usando el enfoque de *Aproximaciones por Problemas Descontados;* luego, usaremos el *Algoritmo de Iteración de Políticas* para obtener una solución de la EOCP con la particularidad de que la política de "inicialización" es una política óptima en $\mathbf{F}$.

Teorema 4.5.3. Suponga que se satisface H4.3.1, H4.3.2 y H4.3.8. Entonces existe una terna (ρ^*, f_*, h_*), con $h_*(\cdot)$ en L_V^∞, que satisface la DOCP. Además,

$$\rho^* = \inf_{\mathbf{F}} J(f, x) = J(f_*, x) = \lim_{\alpha \to 1^-} (1-\alpha) V_\alpha(x) \quad \forall x \in \mathbf{X}.$$

Demostración del Teorema 4.5.3. La demostración de este teorema sólo será bosquejada ya que, esencialmente, es la misma que la demostración dada en [28], Theorem 3.5, y los argumentos son similares a los de la demostración del Teorema 3.3.3. Sin embargo, sólo podemos concluir que existe una política estacionaria "óptima" en la clase $\mathbf{F}$, ya que el Teorema Abeliano (ver Teorema 3.2.4 o Teorema A.E.5) puede no cumplirse debido a que en el presente contexto no suponemos que la función de costo este acotada inferiormente por una constante.

Definiendo para cada $\alpha \in (0,1)$, $x \in \mathbf{X}$, las funciones

$$\rho_\alpha := (1-\alpha) V_\alpha(z), \quad \text{y} \quad h_\alpha(x) := V_\alpha(x) - V_\alpha(z), \tag{4.37}$$

donde z es un estado fijo arbitrario, la Ecuación de Optimalidad en Costo α–Descontado (4.27) se puede re-escribir como

$$\rho_\alpha + h_\alpha(x) = \min_{a \in A(x)} \left[C(x,a) + \alpha \int_{\mathbf{X}} h_\alpha(y) Q(dy|x,a) \right] \quad \forall x \in \mathbf{X}. \tag{4.38}$$

Por otra parte,

$$|h_\alpha(x)| \leq (1-\gamma)^{-1} M [V(x) + V(z)] \quad \forall x \in \mathbf{X}, \alpha \in (0,1), \tag{4.39}$$

donde γ y M son las constantes en H4.3.8. Para verificar esto último, del Teorema 4.4.3(c), note que

$$\begin{aligned} |h_\alpha(x)| &= |\inf_{\mathbf{F}} V_\alpha(f,x) - \inf_{\mathbf{F}} V_\alpha(f,z)| \\ &\leq \sup_{\mathbf{F}} |V_\alpha(f,x) - V_\alpha(f,z)|. \end{aligned} \tag{4.40}$$

También note que

$$\begin{aligned}\sup_{\mathbf{F}}|V_\alpha(f,x)-V_\alpha(f,z)| \quad &\leq \sup_{\mathbf{F}}\sum_{t=0}^{\infty}\alpha^t|E_x^f C(x_t,a_t)-\int_{\mathbf{X}}C(y,f)Q_f(dy)| \\ &+\sup_{\mathbf{F}}\sum_{t=0}^{\infty}\alpha^t|E_z^f C(x_t,a_t)-\int_{\mathbf{X}}C(y,f)Q_f(dy)|.\end{aligned} \tag{4.41}$$

Entonces, del Teorema 4.3.4(c) y H4.3.8, se obtiene

$$\sup_{\mathbf{F}}|V_\alpha(f,x)-V_\alpha(f,z)|\leq(1-\gamma)^{-1}M[V(x)+V(z)] \;\; \forall x\in\mathbf{X},\alpha\in(0,1). \tag{4.42}$$

Combinando (4.40), (4.41) y (4.42), se obtiene (4.39).

Por otra lado, de (4.31), se tiene que

$$|\rho_\alpha|\leq(1-B)bV(z)\;\forall\alpha\in(0,1)$$

donde B, b son las constantes en la Proposición 4.4.1(c). Tomando $\alpha_n(n)\uparrow 1$ tal que

$$\rho^*:=\limsup_{\alpha\uparrow 1}\rho_\alpha=\lim_{n\to\infty}\rho_{\alpha(n)}, \tag{4.43}$$

y sustituyendo α con $\alpha(n)$ en (4.38), argumentos similares a los dados en la demostración del Teorema 3.3.3 demuestran que ρ^* y la función

$$h_*(x):=\liminf_{n\to\infty}h_{\alpha(n)}(x),\;\; x\in\mathbf{X},$$

satisfacen la DOCP, es decir,

$$\begin{aligned}\rho^*+h_*(x) \;\; &\geq \;\; Th_*(x) \\ &= \min_{a\in A(x)}\left[C(x,a)+\int_{\mathbf{X}}h_*(y)Q(dy|x,a)\right] \quad \forall x\in\mathbf{X}.\end{aligned}$$

Note que $h_*(\cdot)$ pertenece al espacio L_V^∞; entonces, de la Proposición 4.4.1(a), existe $f_* \in \mathbf{F}$ tal que

$$Th_*(x) = C(x, f_*) + \int_{\mathbf{X}} h_*(y)Q(dy|x, f_*) \quad \forall x \in \mathbf{X};$$

es decir, la terna (ρ^*, f_*, h_*) satisface la *Desigualdad de Optimalidad en Costo Promedio*.

Iterando la desigualdad

$$\rho^* + h_*(x) \geq C(x, f_*) + \int_{\mathbf{X}} h_*(y)Q(dy|x, f_*) \quad \forall x \in \mathbf{X},$$

de (4.13) y la Proposición 4.4.2(d), se concluye que

$$\rho^* \geq J(f_*, x) = J(f_*) \quad \forall x \in \mathbf{X}.$$

Ahora, de (4.33) y (4.34), tenemos que

$$J(f_*) = \lim_{\alpha \to 1} (1-\alpha)V_\alpha(f_*, z) \geq \limsup_{\alpha \to 1} (1-\alpha)V_\alpha(z).$$

Por lo tanto,

$$\rho^* = \inf_{\mathbf{F}} J(f, x) = J(f_*, x) \quad \forall x \in \mathbf{X}.$$

Sea $\beta(n) \uparrow 1$ tal que

$$\rho_* := \liminf_{\alpha \uparrow 1} \rho_\alpha = \lim_{n \to \infty} \rho_{\beta(n)}.$$

Repitiendo los argumentos anteriores, se concluye que

$$\rho^* = \rho_* = \lim_{\alpha \uparrow 1} (1-\alpha)V_\alpha(z) \quad \forall x \in \mathbf{X}. \tag{4.44}$$

Finalmente, observe que $\forall x \in \mathbf{X}, \alpha \in (0,1)$, se tiene

$$
\begin{aligned}
|(1-\alpha)V_\alpha(x) - \rho_\alpha| &= (1-\alpha)|h_\alpha(x)| \\
&\leq \left\{(1-\gamma)^{-1}M[V(x)+V(z)]\right\}(1-\alpha)
\end{aligned}
$$

lo cual combinado con (4.44), implica

$$\rho^* = \lim_{\alpha \uparrow 1}(1-\alpha)V_\alpha(x) \quad \forall x \in \mathbf{X}. \blacksquare$$

Demostración del Teorema 4.5.2.(a) Como se mencionó previamente, la demostración de esta parte se hará por medio del *Algortimo de Iteración de Políticas* con la particularidad de que la política estacionaria en la parte inicial del proceso recursivo es una política óptima en costo promedio en **F**.

Del Teorema 4.5.3, existe una política f_0 tal que

$$J(f_0) = \rho^* = \lim_{\alpha \uparrow 1}(1-\alpha)V_\alpha(x) \quad \forall x \in \mathbf{X}.$$

Sea $h_0 \in L_V^\infty$ una solución de la Ecuación de Poissson correspondiente a f_0, es decir,

$$\rho^* + h_0(x) = C(x, f_0) + \int_{\mathbf{X}} h_0(y)Q(dy|x, f_0) \quad \forall x \in \mathbf{X}. \tag{4.45}$$

De la Proposición 4.4.1(a), existe $f_1 \in \mathbf{F}$ tal que

$$Th_0(x) = C(x, f_1) + \int_{\mathbf{X}} h_0(y)Q(dy|x, f_1) \quad \forall x \in \mathbf{X}, \tag{4.46}$$

de manera que

$$\rho^* + h_0(x) \geq C(x, f_1) + \int_{\mathbf{X}} h_0(y)Q(dy|x, f_1) \quad \forall x \in \mathbf{X}. \tag{4.47}$$

Ahora denote por $h_1 \in L_V^\infty$ a una solución de la Ecuación de Poisson correspondiente a la política f_1, i.e.,

$$\rho^* + h_1(x) = C(x, f_1) + \int_{\mathbf{X}} h_0(y) Q(dy|x, f_1) \quad \forall x \in \mathbf{X}. \tag{4.48}$$

Entonces, de (4.47) y (4.48), se obtiene

$$h_0(x) - h_1(x) \geq \int_{\mathbf{X}} [h_0(y) - h_1(y)] Q(dy|x, f_1) \quad \forall x \in \mathbf{X},$$

lo cual, por (4.8), implica que

$$h_0(x) - h_1(x) \geq \int_{\mathbf{X}} [h_0(y) - h_1(y)] Q_{f_1}(dy) \quad \forall x \in \mathbf{X}. \tag{4.49}$$

Por otra parte, definiendo $k_1 := \inf_y [h_0(y) - h_1(y)]$, se deduce de (4.49) que $h_0(\cdot) - h_1(\cdot) = k_1$ $Q_{f_1}-$casi dondequiera. Es decir, existe $N_1 \in \mathcal{B}(\mathbf{X})$ con $Q_{f_1}(N_1) = 1$ tal que

$$\begin{aligned} h_0(x) - h_1(x) &= k_1 \quad \forall x \in N_1, \\ h_0(x) - h_1(x) &> k_1 \quad \forall x \in N_1^c, \end{aligned}$$

donde $N_1^c := X - N$.

La repetición del proceso anterior genera una sucesión de funciones $\{h_n\} \subset L_V^\infty$, una sucesión de políticas $\{f_n\} \subset \mathbf{F}$ y una sucesión de conjuntos $\{N_n\} \subset \mathcal{B}(\mathbf{X})$ con las siguientes características: Para cada $n \in \mathbf{N}_0$:

(I)

$$\rho^* + h_n(x) = C(x, f_n) + \int_{\mathbf{X}} h_n(y) Q(dy|x, f_n) \quad \forall x \in \mathbf{X}; \tag{4.50}$$

(II)

$$Th_n(x) = C(x, f_{n+1}) + \int_{\mathbf{X}} h_n(y) Q(dy|x, f_{n+1}) \quad \forall x \in \mathbf{X}; \tag{4.51}$$

(III) Además, $Q_{f_{n+1}}(N_{n+1}) = 1$ y

$$h_n(x) - h_{n+1}(x) \quad = \quad k_{n+1} \quad \forall x \in N_{n+1}, \tag{4.52}$$

$$h_n(x) - h_{n+1}(x) \quad > \quad k_{n+1} \quad \forall x \in N^c_{n+1}. \tag{4.53}$$

Ahora defina el conjunto

$$N^* := \bigcap\nolimits_{n=0}^{\infty} N_n,$$

el cual es no vacío, ya que de lo contrario de H4.3.2(b) y la Observación 4.3.10 tendríamos que $\psi(\mathbf{X}) = 0$. Sea $z^* \in N^*$ y defina las funciones

$$h^*_n(x) := h_n(x) - h_n(z^*) \quad \forall x \in \mathbf{X}, n \in \mathbf{N}_0. \tag{4.54}$$

Entonces, de (4.52)-(4.53), se deduce que

$$h^*_n(\cdot) \geq h^*_{n+1}(\cdot) \quad \forall n \in \mathbf{N}_0. \tag{4.55}$$

Note que la función

$$h^*(x) := \lim_{n\to\infty} h^*_n(x), \quad x \in \mathbf{X}, \tag{4.56}$$

pertenece a L^∞_V (ver Observación 4.3.9). Ahora, demostraremos que h^* satisface la EOCP. Para hacer esto, primero observe que

$$Th^*(x) \leq Th^*_n(x) \leq C(x, f_n) + \int_{\mathbf{X}} h^*_n(y) Q(dy|x, f_n) = \rho^* + h^*_n(x) \quad \forall x \in \mathbf{X}, n \in \mathbf{N};$$

por lo tanto,

$$Th^*(x) \leq \rho^* + h^*(x) \quad \forall x \in \mathbf{X}. \tag{4.57}$$

Para obtener la desigualdad en sentido opuesto, note que combinando (4.51) y (4.54), se tiene que $\forall x \in \mathbf{X}, n \in \mathbf{N}$:

$$\begin{aligned} Th_n^*(x) &= C(x, f_{n+1}) + \int_{\mathbf{X}} h_n^*(y) Q(dy|x, f_{n+1}) \\ &\geq C(x, f_{n+1}) + \int_{\mathbf{X}} h_{n+1}^*(y) Q(dy|x, f_{n+1}) \\ &= \rho^* + h_{n+1}(x) \geq \rho^* + h^*(x). \end{aligned} \tag{4.58}$$

En consecuencia,

$$C(x,a) + \int_{\mathbf{X}} h_n^*(y) Q(dy|x,a) \geq \rho^* + h^*(x) \quad \forall x \in \mathbf{X}, a \in A(x),$$

de donde se obtiene, al tomar límite cuando $n \to \infty$, que

$$C(x,a) + \int_{\mathbf{X}} h^*(y) Q(dy|x,a) \geq \rho^* + h^*(x) \quad \forall x \in \mathbf{X}, a \in A(x).$$

Entonces,

$$Th^*(x) \geq \rho^* + h^*(x) \quad \forall x \in \mathbf{X}. \tag{4.59}$$

Por lo tanto, de (4.57) y (4.59), tenemos que h^* satisface la EOCP

$$\rho^* + h^*(x) = \min_{a \in A(x)} \left[C(x,a) + \int_{\mathbf{X}} h^*(y) Q(dy|x,a) \right] \quad \forall x \in \mathbf{X},$$

y de la Proposición 4.4.1(a), existe $f^* \in \mathbf{F}^*$ tal que

$$\rho^* + h^*(x) = C(x, f^*) + \int_{\mathbf{X}} h^*(y) Q(dy|x, f^*) \quad \forall x \in \mathbf{X}.$$

(b) De acuerdo a la Observación 4.2.2, sólo hace falta demostrar que (iii) implica (i). Primero demostraremos que las funciones

$$L^i(x) := \liminf_{n\to\infty} [J_n^*(x) - n\rho^*] \quad \text{and} \quad L^s(x) := \limsup_{n\to\infty} [J_n^*(x) - n\rho^*], \quad x \in \mathbf{X},$$

solamente toman valores en el conjunto de los números reales. Para hacer esto, considere la terna (ρ^*, f^*, h^*) de la parte (a) y note que

$$[J_n(f^*, x) - n\rho^*] + E_x^{f^*} h^*(x) \leq [J_n^*(x) - n\rho^*] + \sup_{\pi} E_x^{\pi} |h^*(x_n)| \quad \forall x \in \mathbf{X}, n \in \mathbf{N}.$$

Entonces,

$$\widehat{h}_{f^*}(x) + Q_{f^*}\widehat{h}_{f^*} \leq L^i(x) + \limsup_{n\to\infty} \sup_{\pi} E_x^{\pi} |h^*(x_n)| \quad \forall x \in \mathbf{X},$$

lo cual implica, por la Proposición 4.4.2(c), que

$$-\infty < L^i(x) \quad \forall x \in \mathbf{X}. \tag{4.60}$$

Por otra parte, puesto que

$$J_n^*(x) \leq J_n(f^*, x) \quad \forall x \in \mathbf{X},$$

tenemos que

$$L^s(x) \leq \widehat{h}_{f^*}(x) \quad \forall x \in \mathbf{X},$$

de manera que

$$L^s(x) < +\infty \quad \forall x \in \mathbf{X}. \tag{4.61}$$

Ahora observe que (4.60) y (4.61) implican que

$$\lim_{n\to\infty} \frac{1}{n} J_n^*(x) = \rho^* \quad \forall x \in \mathbf{X}. \tag{4.62}$$

Finalmente considere una política $f \in \mathbf{F}$ óptima en costo promedio, i.e., $J(f,x) = \rho^* \; \forall x \in \mathbf{X}$. De (4.13), tenemos que

$$J(f,x) = \lim_{n\to\infty} \frac{1}{n} J_n(f,x) = \rho^* \quad \forall x \in \mathbf{X}. \tag{4.63}$$

Combinando (4.62) y (4.63), se concluye que f es Flynn-óptima, es decir,

$$\lim_{n\to\infty} \frac{1}{n}[J_n(f,x) - J_n^*(x)] = 0 \quad \forall x \in \mathbf{X}.$$

(c) Suponga que $f \in \mathbf{F}$ es óptima en costo promedio y $Q_f(\cdot)$ la medida invariante correspondiente. Observe que si el conjunto de puntos donde se satisface la desigualdad

$$\rho^* + h^*(x) < C(x, f^*) + \int_{\mathbf{X}} h^*(y) Q(dy|x, f)$$

tiene medida positiva con respecto a $Q_f(\cdot)$, entonces

$$J(f) = \int_{\mathbf{X}} C(x, f) Q_f(dx) > \rho^*,$$

lo cual contradice que f es óptima en costo promedio. Por lo tanto, la terna (ρ^*, f, h^*) satisface la EOCP $Q_f(\cdot)$−casi dondequiera. ■

En la siguiente proposición mostramos que las soluciones de la EOCP son *únicas excepto por constantes aditivas*.

Proposición 4.5.4. Suponga que se satisfacen H4.3.1, H4.3.2. y H4.3.8 y que (ρ^*, f_i, h_i) es solucion de la EOCP, donde $h_i \in L_V^\infty$, para $i = 1, 2$. Entonces, la función $H(\cdot) := h_2(\cdot) - h_1(\cdot)$ es constante.

Demostración de la Proposición 4.5.4. Para demostrar esto, primero observe que

$$\rho^* + h_1(x) \;\; = \;\; C(x, f_1) + \int_{\mathbf{X}} h_1(y) Q(dy|x, f_1) \quad \forall x \in \mathbf{X}$$

$$\rho^* + h_2(x) \quad \leq \quad C(x, f_1) + \int_{\mathbf{X}} h_2(y) Q(dy|x, f_1) \quad \forall x \in \mathbf{X},$$

de manera que

$$H(x) \leq \int_{\mathbf{X}} H(y) Q(dy|x, f_1) \quad \forall x \in \mathbf{X},$$

lo cual implica que

$$H(x) \leq \int_{\mathbf{X}} H(y) Q_{f_1}(dy) \quad \forall x \in \mathbf{X}. \tag{4.64}$$

Usando argumentos similares se obtiene

$$\int_{\mathbf{X}} H(y) Q_{f_2}(dy) \leq H(x) \quad \forall x \in \mathbf{X}. \tag{4.65}$$

Ahora definamos

$$k_1 := \sup_{x \in \mathbf{X}} H(x) \quad \text{y} \quad k_2 := \inf_{x \in \mathbf{X}} H(x).$$

Las desigualdades (4.64) y (4.65) implican que

$$H(x) \quad = \quad k_1 \quad \forall x \in N_1,$$

$$H(x) \quad = \quad k_2 \quad \forall x \in N_2,$$

donde N_1 y N_2 son tales que $Q_{f_1}(N_1) = Q_{f_2}(N_2) = 1$. Finalmente, note que $N_1 \cap N_2 \neq \emptyset$, ya que de lo contrario tendríamos que $\psi(\mathbf{X}) = 0$, donde ψ es la medida de irreducibilidad en H4.3.2, puesto que dicha medida es absolutamente continua con respecto a las medidas (invariantes) $Q_{f_1}(\cdot)$ y $Q_{f_2}(\cdot)$ [ver Observación 4.3.10]. Entonces, $k_1 = k_2$ y, por lo tanto, $H(\cdot)$ es una función constante. ■

4.6 Aplicación a un problema de inventarios

De nuevo consideraremos el sistema de inventario presentado en la Sección 3.2.4, el cual evoluciona de acuerdo a

$$x_{t+1} = \max(x_t + a_t - w_t, 0), \quad t \in \mathbf{N}_0; \quad x_0 = x, \tag{4.66}$$

en el espacio $\mathbf{X} = [0, +\infty)$, pero ahora suponemos que variables de control a_t toman valores en $\mathbf{A} := [0, \theta]$ independientemente del nivel de inventario, donde θ es una constante positiva. También suponemos que el proceso de demanda $\{w_t\}$ satisface las siguientes condiciones.

H4.5.1(a) El proceso $\{w_t\}$ está formado por variables aleatorias no negativas independientes e identicamente distribuidas. La función de distribución es denotada por $G(\cdot)$ y satisface lo siguiente:

(a.1) $G(\cdot)$ tiene densidad continua y acotada;

(a.2) $w^* := \int_0^{+\infty} yG(dy) < +\infty$.

(b) $\theta < w^*$.

La esperanza con respecto a la distribución conjunta de $\{w_t\}$ se denota por E.

La función de costo por etapa se toma como

$$C(x, a) := ba + h_c(x + a) - sE\min(x + a, w_0), \quad (x, a) \in \mathbf{K}, \tag{4.67}$$

donde b, h_c y s son constantes positivas.

Ahora mostraremos que el sistema de inventario (4.66)-(4.67), bajo las condiciones H4.5.1, satisface H4.3.1, H4.3.2 y H4.3.8.

Verificación de H4.3.1. Es claro que H4.3.1(a)-(b) se cumplen, mientras que H4.3.1(c) se obtiene observando que

$$\int_{\mathbf{X}} u(y)Q(dy|x, a) = Eu[(x + a - w_0)^+],$$

para todo $(x, a) \in \mathbf{K}$ y toda función medible acotada $u \in M_a(\mathbf{X})$, donde $y^+ := \max(0, y)$.

Para obtener una función de "acotamiento" $V(\cdot)$ que satisfaga H4.3.1(d), denotemos por Ψ a la función generadora de momentos de la variable aleatoria $\theta - w_0$, es decir,

$$\Psi(p) := E\exp[p(\theta - w_0)], \quad p \geq 0.$$

Note que $\Psi(0) = 1$ y que su derivada $\Psi'(0) = \theta - w^* < 0$, lo cual implica que existe una constante $r > 0$ tal que

$$\alpha := E\exp[r(\theta - w_0)] < 1. \tag{4.68}$$

Defina

$$V(x) := \beta E\exp[r(x + 2w^*)], \quad x \in \mathbf{X}, \tag{4.69}$$

donde β es una constante positiva. Con cálculos directos se obtiene que

$$\int_{\mathbf{X}} V(y)Q(dy|x,a) = V(x)\int_0^{x+a}\exp[r(a-y)]G(dy) + V(0)[1 - G(x+a)], \tag{4.70}$$

lo cual muestra que H4.3.1(d.1) se satisface.

Por otra parte, se puede verificar fácilmentre que

$$|C(x,a)| \leq s(x + 2w^*), \quad x \in \mathbf{X}.$$

Entonces, puesto que la constante β se puede tomar lo suficientemente grande de manera que

$$V(x) \geq \max\{1, s(x + 2w^*\}, \quad \forall x \in \mathbf{X},$$

tenemos que H4.3.1(d2) se satisface.

Verificación de H4.3.2. Primero mostraremos que H4.3.2(b) se cumple. Defina la medida

$$\psi(E) := \mathbf{I}_E(0), \quad E \in \mathcal{B}(\mathbf{X}), \tag{4.71}$$

y la función

$$S(x) := 1 - G(x + 0), \quad x \in \mathbf{X}. \tag{4.72}$$

Es fácil verificar que

$$Q(\cdot|x, a) \geq \psi(\cdot)S(x) \quad \forall(x, a) \in \mathbf{K}, \tag{4.73}$$

de donde se deduce que, para cada $f \in \mathbf{F}$, el proceso $\{x_t\}$ es $\psi-$ irreducible y aperiódico (ver Definición A.D.5 o [55], Remark 2.1 y Example 2.5); en consecuencia, H4.3.2(b) se cumple.

Para verificar que H4.3.2(a) se satisface, note que de (4.68)-(4.69) se obtiene

$$\int_{\mathbf{X}} V(y)Q(dy|x, a) \leq \alpha V(x) + V(0) \quad \forall(x, a) \in \mathbf{K}.$$

Entonces,

$$\int_{\mathbf{X}} V(y)Q(dy|x, a) \leq BV(x) + V(0)\mathbf{I}_K(x) \quad \forall(x, a) \in \mathbf{K}, \tag{4.74}$$

donde $B := \frac{1}{2}(1 - \alpha) < 1$ y $K := \{y \in \mathbf{X} : V(y) \leq (1 - \alpha)^{-1}V(0)\}$.

Por otro lado, puesto que la función definida en (4.72) es continua, la relación en (4.73) implica que para cada $f \in \mathbf{F}$ el proceso $\{x_t\}$ es un *T-proceso* (ver Definición A.D.14 y Teorema A.D.17). Además, del Teorema A.D.16, tenemos que cada subconjunto compacto de $\mathbf{X}$ es un *conjunto petite*. Por lo tanto, (4.74) muestra que H4.3.2(a) se cumple.

Verificación de H4.3.8. Para verificar esta condición usaremos un resultado de [28], para lo cual introducimos la siguiente notación: sea ν una medida con signo finita en $\mathcal{B}(\mathbf{X})$ y defina

$$||\nu||_V^* := \int_{\mathbf{X}} V(y)|\nu|(dy),$$

donde $V(\cdot)$ es la función definida en (4.69) y $|\nu|$ es la variación total de ν. Además, defina las medidas con signo

$$\nu_f^n(\cdot|x) := Q^n(\cdot|x, f) - Q_f(\cdot), \quad x \in \mathbf{X}, f \in \mathbf{F}, n \in \mathbf{N},$$

y las funciones

$$S_f(x) := 1 - G(x + f(x)), \quad x \in \mathbf{X}, f \in \mathbf{F}.$$

Note que

$$\sup_{f \in \mathbf{F}} ||\nu_f^n(\cdot|x)||_V^* \leq V(x) M \gamma^n \quad \forall x \in \mathbf{X}, n \in \mathbf{N}, \tag{4.75}$$

implica que H4.3.8 se satisface.

Ahora note que para cada $x \in \mathbf{X}$ y $f \in \mathbf{F}$, las siguientes propiedades se satisfacen:

$$Q(\cdot|x, f) \quad \geq \quad S_f(x)\psi(\cdot); \tag{4.76}$$

$$\int_{\mathbf{X}} V(y) Q(dy|x, f) \quad \leq \quad \alpha V(x) + ||\psi||_V^* S_f(x); \tag{4.77}$$

$$\inf_{f \in \mathbf{F}} \int_{\mathbf{X}} S_f(y)\psi(dy) \quad \geq \quad \int_{\mathbf{X}} S(y)\psi(dy) \geq S(0) > 0. \tag{4.78}$$

En el Lemma 3.4 de [28], se muestra que las propiedades (4.76)-(4.78) implican que (4.75) se cumple. Por lo tanto, H4.3.8 se satisface.

4.7 Conclusiones

En este capítulo, suponiendo que se satisfacen condiciones de continuidad y compacidad de uso estándar en la literatura y una condición del tipo *Lyapunov*, se demostró lo siguiente:

(a) existe una solución (ρ^*, f^*, h^*) de la Ecuación de Optimalidad en Costo Promedio (EOCP) y que $J(f^*, x) = \rho^* \ \forall x \in \mathbf{X}$;

(b) las afirmaciones "f es óptima en costo promedio", "f es fuertemente óptima" y "f es Flynn-óptima" son equivalentes, sin importar si la política f proviene de una solución de la EOCP;

(c) las funciones que pertenecen a una solución de la EOCP son únicas excepto por constantes aditivas.

Resultados similares a éstos se presentan en ([3], [30], [36], [56]) para el caso de *espacios discretos* o *espacios de Borel y costos acotados*, y en ([28], [29]) para *espacios de Borel y costos no acotados*.

En las dos últimas referencias, se supone que el modelo de control satisface una condición de Lyapunov similar a H4.3.2 , pero para obtener una solución (ρ^*, f^*, h^*) de la EOCP se imponen condiciones de "contracción" tanto en la función de costo como en la ley de transición del sistema, que en general parecen ser muy restrictivas y difíciles de verificar. Por otra parte, sólo se muestra que la política f^* es Flynn-óptima (por lo tanto, fuertemente óptima y óptima en costo promedio), pero suponiendo que el soporte de la medida invariante Q_{f^*} es igual al espacio de estados $\mathbf{X}$. En relación a la unicidad de las soluciones de la EOCP no se hace ningún señalamiento. Finalmente, es importante mencionar que la condición de Lyapunov que se usa en estos trabajos implica que H4.3.8 se cumple (ver [28], Lemma 3.4), es decir, que las tasas de convergencia de los procesos de Markov (inducidos por las políticas estacionarias) están uniformemente acotadas por una constante estrictamente menor que uno. De hecho, se usó este resultado en la sección anterior [ver, (4.76)–(4.78)] para mostrar que un sistema de inventarios satisface la condición H4.3.8.

Capítulo 5

Criterios Sensibles al Horizonte de Planeación

5.1 Introducción

En este capítulo se estudian criterios sensibles al crecimiento de los costos en horizonte finito cuando este crece arbitrariamente. En la Sección 2, se introducen las nociones de políticas *fuertemente dominantes* (Ramsey [60]), políticas *dominantes* (Gale [26], von Weiszäcker [81]) *costo de oportunidad* ([25]), *optimalidad en sesgo* ([80]), *costo de Dutta* ([21]). En la Sección 3 se presentan algunas relaciones generales entre estos criterios, mientras que en la Sección 4, bajo las condiciones de continuidad/compacidad y de Lyapunov (H4.3.2, H4.3.3, H4.3.8) introducidas en el Capítulo 4, se demuestra que existe una política estacionaria Dutta óptima en la clase de las políticas estacionarias **F**. También se demuestra que todos estos criterios, exceptuando el de políticas fuertemente dominantes, son equivalentes cuando el análisis se restringe la clase **F**.

5.2 Indices de funcionamiento

Recuerde que para cada política $\pi \in \Pi$ y estado inicial $x_0 = x \in \mathbf{X}$, el *costo en n-etapas* está dado por

$$J_n(\pi, x) := E_x^\pi \sum_{t=0}^{n-1} C(x_t, a_t), \quad n \in \mathbf{N} \tag{5.1}$$

y el *costo promedio por etapa* es

$$J(\pi, x) := \limsup_{n \to \infty} \frac{1}{n} J_n(\pi, x). \tag{5.2}$$

Además, las funciones de costo óptimo correspondientes son

$$J_n^*(x) \quad : \quad = \inf_{\pi \in \Pi} J_n(\pi, x), \quad x \in \mathbf{X}, n \in \mathbf{N}, \tag{5.3}$$

$$J^*(x) \quad : \quad = \inf_{\pi \in \Pi} J(\pi, x), \quad x \in \mathbf{X}. \tag{5.4}$$

Usando las funciones (5.1)-(5.4) se pueden definir un buen número de criterios sensibles al horizonte de planeación ([25], [59]). En este capítulo estudiaremos algunos de ellos.

Definición 5.2.1.(a) Una política π^* es *fuertemente dominante* (FD) si para cada estado inicial $x \in \mathbf{X}$ y cada política $\pi \in \Pi$, existe un número natural $N = N(\pi^*, \pi, x)$ tal que

$$J_n(\pi^*, x) \leq J_n(\pi, x) \quad \forall n \geq N. \tag{5.5}$$

Observe que la noción de políticas *fuertemente dominantes* es más fuerte que la noción de optimalidad en costo promedio. Por otro lado, en [11] y [59] proporcionan ejemplos sencillos en los cuales no existen políticas fuertemente dominantes. De hecho, en algunas aplicaciones la noción de políticas fuertemente dominantes es demasiado restrictiva como para que sea de alguna utilidad. Por esta razón, a continuación se introducen otros criterios menos restrictivos, empezando con una versión debilitada de la noción de políticas fuertemente dominantes.

Definición 5.2.2.(*Gale* [26], *von Weizsäcker* [81]). Una política π^* es *dominante* (D)si para cada $x \in \mathbf{X}$, $\pi \in \Pi$ y $\varepsilon > 0$ existe un natural $N = N(\pi^*, \pi, x)$ tal que

$$J_n(\pi^*, x) \leq J_n(\pi, x) + \varepsilon \quad \forall n \geq N; \tag{5.6}$$

equivalentemente, π^* es *dominante* si

$$\limsup_{n \to \infty} [J_n(\pi^*, x) - J_n(\pi, x)] \leq 0 \quad \forall x \in \mathbf{X}, \pi \in \Pi. \tag{5.7}$$

Definición 5.2.3.(*Flynn* [25]). El *costo de oportunidad* (CO) de una política $\pi \in \Pi$, dado que $x_0 = x \in \mathbf{X}$, se define por

$$CO(\pi, x) := \limsup_{n \to \infty} [J_n(\pi, x) - J_n^*(x)], \tag{5.8}$$

y el *costo de oportunidad* (CO) *óptimo* como

$$CO^*(x) := \inf_{\pi \in \Pi} CO(\pi, x), \quad x \in \mathbf{X}. \tag{5.9}$$

Una política π^* es *óptima en costo de oportunidad* (O.C.O.) si

$$CO(\pi^*, x) = CO^*(x) \quad \forall x \in \mathbf{X}. \tag{5.10}$$

Definición 5.2.4.(*Dutta* [21]). Para una política $\pi \in \Pi$, dado $x_0 = x \in \mathbf{X}$, se define el *costo de Dutta* por

$$D(\pi, x) := \limsup_{n \to \infty} [J_n(\pi, x) - nJ^*(x)], \tag{5.11}$$

y el *costo óptimo de Dutta*

$$D^*(x) := \inf_{\pi \in \Pi} D(\pi, x), \quad x \in \mathbf{X}. \tag{5.12}$$

Una política π^* es *Dutta óptima* si

$$D(\pi^*, x) = D^*(x) \quad \forall x \in \mathbf{X}. \tag{5.13}$$

En lo que sigue se discuten algunas relaciones básicas entre los criterios de optimalidad introducidos en los parrafos previos, para lo cual requerimos la siguiente notación. Para cada $\pi \in \Pi$ y $x \in \mathbf{X}$, se definen las funciones:

$$CO^i(\pi, x) := \liminf_{n\to\infty} [J_n(\pi, x) - J_n^*(x)], \tag{5.14}$$

$$D^i(\pi, x) := \liminf_{n\to\infty} [J_n(\pi, x) - nJ^*(x)], \tag{5.15}$$

$$L^s(x) :=:= \limsup_{n\to\infty} [J_n^*(x) - nJ^*(x)], \tag{5.16}$$

$$L^i(x) := \liminf_{n\to\infty} [J_n^*(x) - nJ^*(x)]. \tag{5.17}$$

Además, si $L^s(\cdot) = L^i(\cdot)$, se define

$$L(x) := L^s(x) = L^i(x) = \lim_{n\to\infty} [J_n^*(x) - nJ^*(x)], \quad x \in \mathbf{X}. \tag{5.18}$$

Observación 5.2.5.(a) Note que si $D(\pi, \cdot) < +\infty$, entonces para cada $x \in \mathbf{X}$ y $\varepsilon > 0$ existe $N = N(x, \varepsilon)$ tal que

$$J_n(\pi, x) - nJ^*(x) \leq D(\pi, x) + \varepsilon \quad \forall n \geq N,$$

lo cual refuerza la interpretación de $J^*(\cdot)$ como la "tasa" mínima de crecimiento de los costos en horizontes finitos. De forma análoga, si $CO(\pi, x) < +\infty$, el crecimiento de la sucesión de costos $\{J_n(\pi, x)\}$ es mínimo en el sentido de ser similar a la de los costos óptimos $\{J_n^*(x)\}$.

(b) Por otro lado, sin condiciones adecuadas, no es posible establecer una relacion general entre los costos de Dutta y de oportunidad, debido principalmete a que las funciones $L^s(\cdot)$ y $L^i(\cdot)$ pueden no coincidir o tomar los valores $\pm\infty$. Sin embargo, si $L^s(\cdot) = L^i(\cdot)$, entonces

$$CO(\pi, x) = D(\pi, x) - L(x) \quad \forall x \in \mathbf{X}. \tag{5.19}$$

Además, si $L(\cdot)$ únicamente toma valores reales, entonces

$$CO^*(x) = D^*(x) - L(x) \quad \forall x \in \mathbf{X}. \tag{5.20}$$

En este caso, π^* es Dutta óptima si y sólo si es óptima en costo de oportunidad.

En la siguiente proposición se presentan otras propiedades, cuyas demostraciones se omiten ya que se obtienen directamente de las definiciones.

Proposición 5.2.6.(a) Si π^* es dominante, entonces es Dutta óptima y óptima en costo de oportunidad;

(b) si π^* es Dutta óptima y

$$D^*(x) = \inf_{\pi \in \Pi} D^i(\pi, x) < +\infty \quad \forall x \in \mathbf{X}, \tag{5.21}$$

entonces, π^* es dominante; por lo tanto, de (a), también es óptima en costo de oportunidad. Similarmente,

(c) si π^* es óptima en costo de oportunidad y

$$CO^*(x) = \inf_{\pi \in \Pi} CO^i(\pi, x) < +\infty \quad \forall x \in \mathbf{X}, \tag{5.22}$$

entonces, π^* es dominante; por lo tanto, de (a), también es Dutta óptima.

(d) Si $CO(\pi, \cdot) < +\infty$, entonces π es Flynn-óptima, es decir,

$$\lim_{n \to \infty} \frac{1}{n} [J_n(\pi, x) - J_n^*(x)] = 0 \quad \forall x \in \mathbf{X}.$$

Análogamente,

(e) si $D(\pi, x) < +\infty$, entonces π es óptima en costo promedio, es decir, $J(\pi, x) = J^*(x)$ $\forall x \in \mathbf{X}$. Si además $L^i(\cdot) < +\infty$, entonces $CO(\pi, \cdot) < +\infty$; por lo tanto de (d), π es fuertemente óptima.

(f) Si π no es óptima en costo promedio [es decir, $J(\pi, \cdot) > J^*(\cdot)$], entonces

$$D(\pi, \cdot) = CO(\pi, \cdot) = +\infty.$$

Observación 5.2.7.(a). Flynn ([25], Example 1) proporciona un ejemplo en el cual existe una política dominante π^*, pero que no es "buena" respecto al costo de oportunidad, es decir, $CO(\pi^*, x) = +\infty$. [Es fácil verificar que lo anterior implica que $CO^*(\cdot) = +\infty$.] Esto muestra, por una parte, que la relación entre la noción de políticas dominantes y el costo de oportunidad no es tan "limpia" como lo sugiere la Proposición 5.2.6(a) y, además, que existe una diferencia importante entre los problemas de control con horizontes "grandes" de planeación y los problemas en horizonte infinito.

(b) La condición (5.22) se cumple si, por ejemplo, $CO^*(\cdot) = 0$, ya que $CO(\pi, \cdot) \geq CO^i(\pi, \cdot) \geq 0$ para cada $\pi \in \Pi$.

(c) Note que si (5.21) y (5.22) se cumplen, entonces de la Proposición 5.2.6(a)-(c) tenemos que las nociones de optimalidad en costo de oportunidad, optimalidad en el sentido de Dutta y la noción de políticas fuertemente dominantes son *equivalentes*.

(d) De la Proposición 5.2.6(a), (d), se deduce que el análisis de existencia de políticas dominantes, óptimas en costo de oportunidad o Dutta óptimas puede ser restringido a la clase de las políticas óptimas en costo promedio.

5.3 Existencia de políticas sensibles al horizonte de planeación

En esta sección suponemos que se satisfacen las hipótesis H4.3.1, H4.3.2 y H4.3.8 (introducidas en el Capítulo 4) y mostraremos que existe una política estacionaria f^* Dutta óptima en $\mathbf{F}$ y que las afirmaciones "$f \in \mathbf{F}$ es Dutta óptima en $\mathbf{F}$", "$f \in \mathbf{F}$ es dominante en $\mathbf{F}$" y "$f \in \mathbf{F}$ es óptima en costo de oportunidad en $\mathbf{F}$" son equivalentes. También mostraremos que cualquiera de estas afirmaciones es equivalente a la "optimalidad en sesgo" que se introduce a continuación [ver Observación 5.3.2(c) y Teorema 5.3.7].

Recuerde que bajo H4.3.1, H4.3.2 y H4.3.8, para cada política estacionaria $f \in \mathbf{F}$, existe una solución $(J(f), h_f)$, con h_f en L_V^∞, de la Ecuación de Poisson [ver Teorema 4.3.2 y Observación 4.3.7]

$$J(f) + h_f(x) = C(x, f) + \int_{\mathbf{X}} h_f(y) Q(dy|x, a) \quad \forall x \in \mathbf{X}, \tag{5.23}$$

y que

$$\widehat{h}_f(x) := h_f(x) - \int_{\mathbf{X}} h_f(y) Q_f(dy), \quad x \in \mathbf{X}, \tag{5.24}$$

es la única solución que satisface $\int_{\mathbf{X}} \widehat{h}_f(y) Q_f(dy) = 0$.

También recuerde que existe una solución (ρ^*, f^*, h^*), con h^* en L_V^∞, de la EOCP [ver Teorema 4.5.2]

$$\rho^* + h^*(x) = \min_{a \in A(x)} \left[C(x,a) + \int_{\mathbf{X}} h^*(y) Q(dy|x,a) \right] \quad \forall x \in \mathbf{X}, \tag{5.25}$$

$$= C(x, f^*) + \int_{\mathbf{X}} h^*(y) Q(dy|x, f) \quad \forall x \in \mathbf{X}, \tag{5.26}$$

lo cual implica que $J^*(\cdot) = J(f^*, \cdot) = \rho^*$.

Denotamos por $\mathbf{F}^*$ la clase de las políticas estacionarias óptimas en costo promedio y definimos la función

$$\widehat{h}(x) := \inf\{\widehat{h}_f(x) : f \in \mathbf{F}^*\}. \tag{5.27}$$

Definición 5.3.1.[*Veinott* ([80])] Una política $f \in \mathbf{F}^*$ es *óptima en sesgo* (O.S.) si

$$\widehat{h}(x) = \widehat{h}_f(x) \quad \forall x \in \mathbf{X}. \tag{5.28}$$

Observación 5.3.2.(a) De la Observación 4.3.7(c), tenemos que si $f \in \mathbf{F}^*$, entonces

$$\widehat{h}_f(x) = D(f,x) = D^i(f,x) = \lim_{n\to\infty} [J_n(f,x) - n\rho^*] \quad \forall x \in \mathbf{X}, \tag{5.29}$$

lo cual combinado con la Proposición 5.2.6(d), implica que

$$\widehat{h}(x) = \inf_{\mathbf{F}} D(f,x) \quad \forall x \in \mathbf{X}. \tag{5.30}$$

(b) Bajo las hipótesis H4.3.1, H4.3.2 y H4.3.8, combinando el Teorema 4.5.2 y la Proposición

5.2.6, se deduce que las implicaciones en el siguiente diagrama se cumplen.

Diagrama 1.

$$\begin{array}{ccccccc}
\pi \text{ es FD} & \Rightarrow & \pi \text{ es D} & \Rightarrow & \pi \text{ es O.C.O.} & \Rightarrow & \pi \text{ es Flynn-op.} \\
 & & \Downarrow & & & & \Updownarrow \\
 & & \pi \text{ es Dutta op.} & & \pi \text{ es CP op.} & \Longleftrightarrow & \pi \text{ es F-op.}
\end{array}$$

(c) Además, en el Teorema 5.3.2, demostramos que restringiendo el análisis a la clase de las políticas estacionarias, las equivalencias en el siguiente diagrama se verifican.

Diagrama 2. En $\mathbf{F}$:

$$\begin{array}{ccc}
f \text{ es D} & \Leftrightarrow & f \text{ es Dutta op.} \\
\Updownarrow & & \Updownarrow \\
f \text{ es O.S.} & \Leftrightarrow & f \text{ es CO op.}
\end{array}$$

Teorema 5.3.3. Suponga que se satisfacen H4.3.1, H4.3.2 y H4.3.8. Entonces, existe una política estacionaria f^* Dutta óptima en $\mathbf{F}$, es decir,

$$D(f^*, x) = \inf_{\mathbf{F}} D(f, x), \quad \forall x \in \mathbf{X}.$$

Es importante e interesante señalar que la demostración del Teorema 5.3.3 *es exactamente la misma demostración que Nowak* ([56]) *da para la existencia de una política estacionaria dominante en* $\mathbf{F}$, lo cual combinado con (5.30) sugiere las equivalencias en el Diagrama 2.

Antes de probar el Teorema 5.3.3, demostraremos que la búsqueda de políticas Dutta óptimas en $\mathbf{F}$ debe realizarse en el conjunto de las políticas estacionarias que provienen de una solución de la EOCP. Para precisar ésto, denotemos por $\mathbf{F}_*$ a la clase formada por tales políticas y note que por la Proposición 4.5.4,

$$\mathbf{F}_* = \left\{ f \in \mathbf{F} : \rho^* + h^*(x) = C(x, f) + \int_{\mathbf{X}} h^*(y) Q(dy|x, f) \; \forall x \in \mathbf{X} \right\},$$

donde h^* es como en (5.25).

Proposición 5.3.4. Bajo las hipótesis del Teorema 5.3.3 se cumple que

$$\inf_{\mathbf{F}} D(f,x) = \inf_{\mathbf{F}^*} D(f,x) = \inf_{\mathbf{F}_*} D(f,x) \quad \forall x \in \mathbf{X}.$$

Demostración de la Proposición 5.3.4. La primera igualdad se sigue de la Proposición 5.2.6(f). A continuación demostraremos que la segunda desigualdad también se cumple. Para hacer esto, sea f una política óptima en costo promedio y (ρ^*, g^*, h^*) una solución de la EOCP. Del Teorema 4.5.2(c) existe un conjunto $N \in \mathcal{B}(\mathbf{X})$, con $Q_f(N) = 1$, tal que

$$\begin{aligned} \rho^* + h^*(x) &= C(x,f) + \int_{\mathbf{X}} h^*(y) Q(dy|x,f) \quad \forall x \in N, \\ \rho^* + h^*(x) &< C(x,f) + \int_{\mathbf{X}} h^*(y) Q(dy|x,f) \quad \forall x \in N^c. \end{aligned} \tag{5.31}$$

Ahora, consideremos la politica estacionaria definida como sigue: $f' = f$ en N y $f' = g^*$ en N^c. Note que (ρ^*, f', h^*) es una solución de la EOCP y que $Q_{f'}(\cdot) = Q_f(\cdot)$. En consecuencia,

$$\int_{\mathbf{X}} h^*(y) Q_{f'}(dy) = \int_{\mathbf{X}} h^*(y) Q_f(dy). \tag{5.32}$$

Por otro lado, iterando (5.31) se obtiene que

$$h^*(x) - E_x^f h^*(x_n) \leq J_n(f,x) - n\rho^* \quad \forall x \in \mathbf{X}, n \in \mathbf{N}.$$

Entonces, tomando límiote cuando $n \to \infty$ y usando (5.32), se obtiene para cada $x \in \mathbf{X}$:

$$\begin{aligned} \widehat{h}_{f'}(x) &= h^*(x) - \int_{\mathbf{X}} h^*(y) Q_{f'}(dy) \\ &= h^*(x) - \int_{\mathbf{X}} h^*(y) Q_{f'}(dy) \\ &\leq \widehat{h}_f(x). \end{aligned}$$

Por lo tanto, puesto que $f' \in \mathbf{F}_*$, tenemos que

$$\inf_{\mathbf{F}^*} D(f,x) = \inf_{\mathbf{F}_*} D(f,x) \quad \forall x \in \mathbf{X}. \blacksquare$$

Demostración del Teorema 5.3.3. De la Proposición 5.3.4, tenemos que para minimizar $D(\cdot,x)$ (en $\mathbf{F}$) podemos restringirnos a $\mathbf{F}_*$, la clase de las políticas estacionarias que provienen de una solución de la EOCP. Es decir, el problema se reduce a encontrar una política $f^* \in \mathbf{F}^*$ tal que

$$D(f^*,x) = \inf_{\mathbf{F}} D(f,x) = \inf_{\mathbf{F}^*} D(f,x) = h^*(x) + \inf_{\mathbf{F}^*}\left[-\int_{\mathbf{X}} h^*(y)Q_f(dy)\right] \quad \forall x \in \mathbf{X}.$$

Por otra parte—como en Nowak ([56])—la minimización $D(\cdot,x)$ en $\mathbf{F}^*$, para todo $x \in \mathbf{X}$, es equivalente a buscar una política estacionaria óptima en costo promedio en un *nuevo modelo de control* $(\mathbf{X}, \mathbf{A}, \{A^*(x) : x \in \mathbf{X}\}, Q, \widehat{C})$, donde

$$A^*(x) \quad : \quad = \left\{a \in A(x) : C(x,a) + \int_{\mathbf{X}} h^*(y)Q(dy|x,a) = \rho^* + h^*(x)\right\}, \; x \in \mathbf{X},$$

$$\widehat{C}(x,a) \quad : \quad = -h^*(x), \quad a \in A^*(x), x \in \mathbf{X},$$

y $\mathbf{X}$, $\mathbf{A}$ y Q son los mismos elementos del modelo de control original.

Es fácil verificar que el nuevo modelo de control satisface las hipótesis H4.3.1, H4.3.2 y H4.3.8; por lo tanto, existe una política estacionaria $f^* \in \mathbf{F}_*$ óptima en costo promedio en el nuevo modelo de control, es decir,

$$D(f^*,x) = h^*(x) - \int_{\mathbf{X}} h^*(dy)Q_{f^*}(dy) = \inf_{\mathbf{F}_*} D(f,x) \quad \forall x \in \mathbf{X}. \blacksquare$$

Los resultados en el siguiente lema nos serán de utilidad para demostrar las equivalencias anunciadas en el Diagrama 2.

Lema 5.3.5. Bajo las condiciones H4.3.1, H4.3.2 y H4.3.8, para cada política estacionaria $f \in \mathbf{F}_*$ se cumple lo siguiente:

(a) $CO(f,\cdot) \leq ||D(f,\cdot)||_V b(1-B)^{-1}$, donde b y B son las constantes en el Proposición 4.4.1(c);

(b) $D(f,\cdot) = L^s(\cdot) + CO^i(f,\cdot) = L^i(\cdot) + CO(f,\cdot)$;

(c) $CO^i(f,\cdot) = k_f = \liminf_{n\to\infty} [J_n(f,\cdot) - J_n^*(\cdot)]$, Q_f–casi dondequiera, donde k_f es una constante no-negativa.

Observación 5.3.6. Sea (ρ^*, f, h) una solución de la EOCP, donde $\int_{\mathbf{X}} h(y)Q_f(dy) = 0$. Un problema importante relacionado con el Lema 5.3.5 es el de la convergencia del *Algoritmo de Iteración de Valores*. Usualmente, esta convergencia se obtiene por medio de la convergencia a una constante de las llamadas *funciones de error*

$$e_n(\cdot) := n\rho^* + h(\cdot) - J_n^*(\cdot), \quad n \in \mathbf{N}_0, \tag{5.33}$$

las cuales, por (5.25)-(5.26), se pueden re-escribir como

$$e_n(x) = J_n(f,x) - J_n^*(x) + E_x^f h(x_n). \tag{5.34}$$

La relación con el Lema 5.3.5 se describe a continuación. Puesto que

$$\lim_{n\to\infty} E_x^f h(x_n) = \int_{\mathbf{X}} h(y)Q_f(dy) = 0,$$

tenemos que

$$\lim_{n\to\infty} e_n(x) = k_f \quad \forall x \in \mathbf{X},$$

si y sólo si,

$$OC(f,x) = OC^i(f,x) = \lim_{n\to\infty} [J_n(f,x) - J_n^*(x)] = k_f \quad \forall x \in \mathbf{X}.$$

Además, cualquiera de estas condiciones, combinada con la Proposición 5.3.5(b), implica a las siguientes propiedades: (i) $L(\cdot) = L^s(\cdot) = L^i(\cdot)$; (ii) $D(f,\cdot) = L(x) + k_f$.

Demostración del Lema 5.3.5. Sea f una política estacionaria óptima en costo promedio. Del Teorema 4.5.2(b), existe $h \in L_V^\infty$ tal que (ρ^*, f, h) es una solución de la EOCP. Tomemos h

de manera que $\int_{\mathbf{X}} h(y)Q_f(dy) = 0$, es decir, $h(\cdot) = \widehat{h}_f(\cdot) = D(f, \cdot)$.

(a) De (5.25)-(5.26) tenemos que

$$0 \leq J_n(f,x) - J_n^*(x) \leq \sup_{\Pi} E_x^{\pi} h(x_n) - E_x^f h(x_n),$$

lo cual implica, por la Proposición 4.4.2(c), que

$$CO(f,x) \leq ||h||_V b(1-B)^{-1} \quad \forall x \in \mathbf{X}.$$

(b) De (5.33)-(5.34), se obtiene que $\forall x \in \mathbf{X}$

$$\begin{aligned} \limsup_{n\to\infty} e_n(x) &= h(x) - L^i(x) \\ \liminf_{n\to\infty} e_n(x) &= h(x) - L^s(x), \end{aligned}$$

lo cual, puesto que $h(\cdot) = D(f, \cdot)$, implica que (b) se cumple.

(c) Se puede demostrar ([34], Lemma 5.6.4; [52], Lemma 5.4) que

$$e_{n+k}(x) \geq \int_{\mathbf{X}} e_n(y) Q^k(dy|x, f) \quad \forall n,\ k \in \mathbf{N}_0,\ x \in \mathbf{X}.$$

Tomando lim inf cuando $k \to \infty$, se obtiene

$$CO^i(f,x) \geq \int_{\mathbf{X}} e_n(y) Q_f(dy) \quad \forall n \in \mathbf{N}_0, x \in \mathbf{X},$$

y, aplicando el Lema de Fatou,

$$CO^i(f,x) \geq \int_{\mathbf{X}} CO^i(y) Q_f(dy) \quad \forall x \in \mathbf{X}.$$

Por lo tanto,

$$CO^i(f,x) = \inf_{y \in \mathbf{X}} CO^i(y) =: k_f, \quad Q_f - \text{c.d.} \blacksquare$$

Finalmente, tenemos el teorema anunciado en la Observación 5.3.2(c).

Teorema 5.3.7. Bajo las hipótesis H4.3.1, H4.3.2 y H4.3.8, las siguientes afirmaciones son equivalentes: **(i)** f es Dutta óptima en $\mathbf{F}$; **(ii)** f es óptima en sesgo, **(iii)** f es dominante en $\mathbf{F}$; **(iv)** f es óptima en costo de oportunidad en $\mathbf{F}$. Por lo tanto del Teorema 5.3.3., existe una política estacionaria f que satisface las condiciones (i)-(iv).

Demostración del Teorema 5.3.7. Note primero que por la Proposición 5.2.6(f) el análisis puede restringirse a la clase $\mathbf{F}^*$ de las políticas estacionarias óptimas en costo promedio.
(i)⇔(ii). Esta equivalencia es una consecuencia inmediata de la Observación 5.3.2(a).

(i)⇔(iii). Observe que para cada f y g en $\mathbf{F}^*$, se cumple lo siguiente:

$$D(f,x) - D(g,x) = \lim_n [J_n(f,x) - J_n(g,x)] \quad \forall x \in \mathbf{X},$$

lo cual implica que f es Dutta óptima en $\mathbf{F}^*$ si y sólo si es dominante en $\mathbf{F}^*$.
(iii)⇔(iv). Suponga que $f \in \mathbf{F}^*$ es dominante. Entonces para cada $\varepsilon > 0, g \in \mathbf{F}^*$ y $x \in \mathbf{X}$, existe $N = N(f,g,x,\varepsilon)$ tal que

$$J_n(f,x) - J_n^*(x) \leq J_n(g,x) - J_n^*(x) + \varepsilon \quad \forall n \geq N,$$

lo cual implica que

$$CO(f,x) \leq CO(g,x) + \varepsilon.$$

Por lo tanto,

$$CO(f,x) \leq CO(g,x) \quad \forall x \in \mathbf{X}, g \in \mathbf{F}^*.$$

Ahora supongamos que $f \in \mathbf{F}^*$ es óptima en costo de oportunidad en $\mathbf{F}^*$. Por argumentos similares a los dados en la demostración de la Proposición 5.3.4, existe $f' \in \mathbf{F}_*$ tal que

$$CO(f, x) = CO(f', x) \quad \forall x \in \mathbf{X}. \tag{5.35}$$

Del Lema 5.3.5(b) tenemos que

$$\inf_{\mathbf{F}_*} D(f, x) = \inf_{\mathbf{F}_*} CO(f, x) + L^i(x) \quad \forall x \in \mathbf{X}, \tag{5.36}$$

lo cual combinado con la Proposición 5.3.4, implica que f' es Dutta óptima en $\mathbf{F}^*$. Esto último implica que f también es Dutta óptima ya que de lo contrario, es decir, si

$$D(f', x') < D(f, x')$$

para algún $x' \in \mathbf{X}$, tendríamos que para $\varepsilon > 0$ suficientemente pequeño existe $N = N(f, f', x', \varepsilon)$ tal que

$$J_n(f', x) - J_n^*(x) + \varepsilon \leq J_n(f', x) - J_n^*(x) \quad \forall n \geq N.$$

Entonces

$$CO(f', x') < CO(f, x'),$$

lo cual contradice a (5.35). Por lo tanto f es Dutta óptima, lo cual a su vez implica que dicha política es dominante en $\mathbf{F}^*$.

Observación 5.3.8. Brown [11] proporciona un ejemplo que muestra que *no es posible* extender los resultados de del Teorema 5.3.7 a la clase de todas las políticas sin condiciones adicionales a H4.3.1, H4.3.2 y H4.3.8. De hecho, en este ejemplo existe una política estacionaria óptima en el sentido de *Blackwell* (ver [25]) pero no es dominante en el conjunto de todas las políticas. Sin embargo, en este ejemplo *sí existe una política estacionaria fuertemente dominante en* $\mathbf{F}$. En el siguiente ejemplo, el cual se tomó de [57], mostramos que esto último puede no ser cierto aun cuando se satisfagan las condiciones H4.3.1, H4.3.2 y H4.3.8.

Ejemplo 5.3.9. Suponga que el espacio de estados y controles están dados por $\mathbf{X} = \mathbf{A} =$

$\{1, 2, 3\}$, mientras que los costos y conjuntos de controles admisibles son los siguientes:

$$\begin{aligned} A(1) &= \{1, 2\} & A(2) &= A(3) &= \{1\} \\ C(1,1) &= 1.4 & C(1,2) &= 0.2 \\ C(2,1) &= -9 \\ C(3,1) &= 6. \end{aligned}$$

En lo que sigue usaremos notación matricial. Note que sólo existen dos políticas estacionarias: $f(1) = 1$ y $g(1) = 2$, es decir, $\mathbf{F} = \{f, g\}$. Las matrices de transición en un paso y los costos están dados como sigue:

$$P_f = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \end{bmatrix} \quad C_f = \begin{bmatrix} 1.4 \\ -9 \\ 6 \end{bmatrix}$$

$$P_g = \begin{bmatrix} 0.1 & 0.2 & 0.7 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \end{bmatrix} \quad C_g = \begin{bmatrix} 0.2 \\ -9 \\ 6 \end{bmatrix}.$$

Es claro que este modelo de control satisface las condiciones en el Teorema 5.3.7. Por otra parte, con cálculos directos se puede verificar que

$$Q_f = [15/24, 5/24, 4/24] \quad \text{y} \quad Q_g = [15/42, 11/42, 16/42]$$

son las medidas invariantes a P_f y P_g, respectivamente. Además, el vector

$$h^* = [8, 0, 10]^T,$$

donde T denota la operación de transposición, satisface

$$
\begin{aligned}
h^* &= C_f + P_f h^* \\
&= C_g + P_g h^*.
\end{aligned} \tag{5.37}
$$

Entonces, la constante $\rho^* = 0$, el vector h^* y las políticas f y g satisfacen la EOCP. Por lo tanto, f y g son óptimas en costo promedio.

Ahora observe que

$$Q_f h^* = Q_g h^*;$$

entonces, tanto f como g son *dominantes* en $\mathbf{F}$. A continuación mostraremos que ninguna de ellas es fuertemente dominante en $\mathbf{F}$, es decir, f no domina fuertemente a g ni ésta última a f. Para hacer esto, denotemos por $J_n(f)$ y $J_n(g)$ los costos en n-etapas correspondientes a f y g, respectivamente. Entonces, iterando (5.37), se obtiene

$$A_n := J_n(g) - J_n(f) = P_f^n h^* - P_g^n h^* \quad \forall n \in \mathbf{N}.$$

Por otra parte, usando argumentos de inducción se puede mostrar que

$$P_f^n h^* = \frac{1}{7.056} \begin{bmatrix} 47.04 + 3.36(-0.5)^n + 6.048(0.2)^n \\ 47.04 - 36.96(-0.5)^n - 10.08(0.2)^n \\ 47.04 + 33.6(-0.5)^n - 10.08(0.2)^n \end{bmatrix}$$

$$P_g^n h^* = \frac{1}{7.056} \begin{bmatrix} 47.04 - 117.6(-0.5)^n + 127.008(-0.4)^n \\ 47.04 + 23.52(-0.5)^n - 70.56(-0.4)^n \\ 47.04 + 94.08(-0.5)^n - 70.56(-0.4)^n \end{bmatrix}$$

se cumple para todo $n \in \mathbf{N}$. Entonces,

$$A_n = \begin{bmatrix} a_n \\ b_n \\ c_n \end{bmatrix} := \frac{1}{7.056} \begin{bmatrix} 120.96(-0.5)^n + 6.048(0.2)^n - 127.008(-0.4)^n \\ -60.48(-0.5)^n - 10.08(0.2)^n + 70.56(-0.4)^n \\ -60.48(-0.5)^n - 10.08(0.2)^n + 70.56(-0.4)^n \end{bmatrix}.$$

Suponga que g es fuertemente dominante en $\mathbf{F}$. Entonces existe un número natural N tal que $a_n + b_n + c_n \leq 0\ \forall n \geq N$, lo cual lleva a la contradicción $(-0.4)^n \leq (0.2)^n\ \forall n \geq N$. Por lo tanto, g *no es fuertemente dominante en* $\mathbf{F}$. Argumentos similares muestran que f *no es fuertemente dominante en* $\mathbf{F}$.

En la Sección 4.5 se presentó un ejemplo de inventarios que satisface las condiciones H4.3.1, H4.3.2 y H4.3.8, y por lo tanto, los resultados en los Teoremas 5.3.3 y 5.3.7 se cumplen. Concluimos esta sección discutiendo un ejemplo de Nowak [56], modificado ligeramente, que muestra que algunas de las implicaciones en el Diagrama 1 no se cumplen si no tienen condiciones adecuadas sobre el modelos de control. Concretamente, se muestra que existe una política f^* que proviene de una solución de la EOCP, la cual es Flynn óptima (en consecuencia, también es óptima en costo promedio), pero no es dominante ni óptima en costo de oportunidad. Por otra parte, existe una política f que domina a f^* (en consecuencia, es Flynn óptima) pero no proviene de una solución de la EOCP.

Ejemplo 5.3.10. Considere un modelo de control con espacio de estados $\mathbf{X}=\mathbf{N}_0$ y espacio de controles $\mathbf{A} = A(\cdot) = \{1, 2\}$. El estado $x = 0$ es absorbente y tiene costos iguales a cero, es decir,

$$C(0, \cdot) = 0 \quad \text{y} \quad Q(0|0, \cdot) = 1.$$

Por otra parte, si $x \geq 1$

$$C(x, 1) = x^{-1} - 1, \qquad Q(0|x, 1) = 1,$$

y

$$C(x, 2) = 0, \qquad Q(x+1|x, 2) = 1.$$

Se puede verificar facilmente que $J_n(\pi, 0) = 0$ para cada $\pi \in \Pi$ y $n \in \mathbf{N_0}$, lo cual implica que $J_n^*(0) = 0\ \forall n \in \mathbf{N_0}$. Además, con cálculos directos se obtiene que

$$J_n^*(x) = (x+n)^{-1} - 1 \quad \forall n,\, x \geq 1.$$

Considere las políticas $f^*(\cdot) \equiv 2$ y $f(\cdot) \equiv 1$. Por un parte, note que $(0, f^*, h)$ es una solución de la EOCP, donde $h(0) = 0$ y $h(x) = -1$, si $x \geq 1$; entonces, puesto que $h(\cdot)$ es acotada, f^* es Flynn óptima. Ademas,

$$J_n(f^*, x) = 0 \quad \text{y} \quad J_n(f, x) = x^{-1} - 1 \quad \forall x, n \geq 1.$$

lo cual implica que para cada $x > 1$,

$$\lim_{n\to\infty} [J_n(f, x) - J_n(f^*, x)] < 0 \quad \forall x > 1,$$

y que

$$CO(f, x) < CO(f^*, x) \quad \forall x > 1.$$

Resumiendo: la política f^* pertenece a una solución de la EOCP, pero *no es dominante ni óptima en costo de oportunidad.* Por otra parte, tenemos que f es Flynn-óptima y domina a f^*, pero *no proviene de una solución de la EOCP.*

5.4 Conclusiones

En este capítulo, usando una condición de Lyapunov y condiciones de continuidad/compacidad estándar, se demostró que existe una política Dutta óptima en la clase de las políticas estacionarias y que dicha política también es dominante, óptima en costo de oportunidad y óptima en sesgo.(De hecho se mostró que todos estos conceptos son equivalentes cuando el análisis se restringe a la clase de las políticas estacionarias).

Sin duda alguna, los criterios sensibles al horizonte de planeación son poco comprendidos y lo que se conoce de ellos corresponde más bien a su comportamiento extremadamente patológico

([11], [24], [25], [59]), que aun condiciones tan fuertes como la ergodicidad geométrica uniforme no los elimina por completo ([11]). Por otra parte, hasta donde conocemos, la existencia de políticas estacionarias dominantes en toda la clase de las políticas sólo se ha demostrado para modelos con *espacios finitos* y aun en este caso bajo condiciones muy restricitivas ([19], [43], [44]). De hecho, la hipótesis en [19] implican que si el modelo es ergódico, entonces sólo existe una política óptima en costo promedio, mientras que en [44] esto último se supone directamente.

Capítulo 6

Indice en Costo Promedio: Análisis por Trayectorias

6.1 Introducción

Existe una gran cantidad de literatura que trata con el problema de control en costo promedio bajo una variedad amplia de hipótesis ([3], [6], [34], [59]), pero en la mayoria de los casos los criterios de optimalidad se formulan en términos de costos "esperados" y, típicamente, se demuestra la existencia de una política estacionaria que genera costos esperados en horizonte finito con tasa mínima de crecimiento, y que dicha tasa no depende del estado inicial del sistema controlado. Resultados de este tipo son bastantes aceptables en muchas aplicaciones pero es más atractivo obtener políticas que tengan una tasa de crecimiento mínimo pero con respecto a los costos "observados" (es decir, por trayectorias). Sin embargo, los trabajos en los cuales se realiza una análisis por trayectorias son escasos y, cuando dicho análisis se efectúa, se restringe a espacios de estados discretos, o bien, a espacios de Borel pero con costos acotados ([3], [10], [15], [47]), en ambos casos bajo condiciones de recurrencia muy restrictivas.

En este capítulo se realizará un análisis por trayectorias del índice en costo promedio para modelos de Markov con espacios de Borel y costos estrictamente no acotados (es decir, costos que crecen arbitrariamente en los complementos de conjuntos compactos) bajo condiciones de continuidad y recurrencia bastantes generales. La idea básica al considerar costos estrictamente no acotados es penalizar fuertemente los procesos controlados que "escapan al infinito", de manera que la busqueda de políticas óptimas se puede restringir a las políticas que inducen comportamientos "estables".

6.2 Criterios de optimalidad

En lo que resta de este capítulo, suponemos que el modelo de control $(\mathbf{X}, \mathbf{A}, \{A(x) : x \in \mathbf{X}\}, Q, C)$ satisface las siguientes condiciones:

H6.2.1(a) $C(x, a)$ es no-negativa y semicontinua inferiormente en $(x, a) \in \mathbf{K}$;

(b) C es *estrictamente no acotada* en $\mathbf{K}$; es decir, existe una sucesión de subconjuntos compactos $\{\mathbf{K}_n\}$ que convergen crecientemente a $\mathbf{K}$ tales que

$$\lim_{n\to\infty} \inf\{C(x, a) : (x, a) \notin \mathbf{K}_n\} = +\infty;$$

(c) $Q(\cdot|x, a)$ es *débilmente continua* en $(x, a) \in \mathbf{K}$, es decir, la función

$$(x, a) \mapsto \int_{\mathbf{X}} u(y) Q(dy|x, a)$$

es continua para cada función continua y acotada $u \in C_a(\mathbf{X})$.

Para cada política $\pi \in \Pi$ y distribución inicial $\nu \in \mathbf{P}(\mathbf{X})$, el *costo promedio por trayectoria* (CPT) se define como

$$J_0(\pi, \nu) := \limsup_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} C(x_t, a_t), \tag{6.1}$$

y el *costo promedio esperado* (CPE) por

$$J(\pi,\nu) := \limsup_{n\to\infty} \frac{1}{n} E_\nu^\pi \sum_{t=0}^{n-1} C(x_t, a_t). \tag{6.2}$$

El *costo promedio óptimo* está dado por

$$j^* := \inf_\nu \inf_\pi J(\pi,\nu). \tag{6.3}$$

Para evitar casos triviales, suponemos que se satisface las siguiente condición.

H6.2.2. Existe una política π^* y una distribución inicial $\nu^* \in \mathbf{P}(\mathbf{X})$ tal que $J(\pi^*,\nu^*) < +\infty$.

Los criterios de optimalidad que se analizarán en las secciones subsecuentes se introducen a continuación.

Definición H6.2.3. Sea $\pi^* \in \Pi$ y $\nu^* \in \mathbf{P}(\mathbf{X})$.

(a) (π^*,ν^*) es un *par mínimo* (PM) si $J(\pi^*,\nu^*) = j^*$;

(b) π^* es *óptima en costo promedio esperado* (OCPE) si (π^*,ν) es un par mínimo para cada $\nu \in \mathbf{P}(\mathbf{X})$;

(c) π^* es *fuertemente óptima* (FO) si es óptima en costo promedio esperado y, además, se satisface que

$$\liminf_{n\to\infty} \frac{1}{n} E_\nu^\pi \sum_{t=0}^{n-1} C(x_t, a_t) \geq J(\pi^*,\nu) \quad \forall \pi \in \Pi, \nu \in \mathbf{P}(\mathbf{X}); \tag{6.4}$$

(d) π^* es *óptima en costo promedio por trayectorias* (OCPT) si para cada $\pi \in \Pi$ y $\nu \in \mathbf{P}(\mathbf{X})$, se cumplen las siguientes condiciones:

(d.1) $J_0(\pi^*,\nu) = j^*$ $\;P_\nu^{\pi^*}$–casi seguramente;

(d.2) $J_0(\pi,\nu) \geq j^*$ $\;P_\nu^{\pi}$–casi seguramente.

Observación 6.2.4. En la literatura también se hace referencia a la propiedad en H6.2.1(b) diciendo que C es un *momento* en $\mathbf{K}$. Esta propiedad tiene consecuencias muy interesantes y ha sido explotada en distintos contextos ([32], [35], [48], [49]). De hecho, en [32] se demuestra, bajo las hipótesis H6.2.1 y H6.2.2, que existe una política "estable" φ^* [Definición 6.2.5(a)] y una distribución inicial μ_{φ^*} que forman un par mínimo. En el Teorema 6.3.1 damos una prueba distinta de este resultado y, adicionalmente, mostramos que el costo promedio óptimo se puede "aproximar" por problemas descontados. En el Teorema 6.4.1 demostramos que si la política φ^* adicionalmente es Harris recurrente, entonces es óptima en costo promedio por trayectorias.

A continuación se introducen las clases de las políticas estables y Harris recurrentes, las cuales juegan un papel muy importante en el análisis de existencia de pares mínimos y políticas óptimas por trayectorias.

Definición 6.2.5. Sea φ una política relajada [Definición 2.3.3].
(a) Diremos que φ es una política *estable* si existe una *medida de probabilidad invariante* $\mu_\varphi \in \mathbf{P}(\mathbf{X})$ para la probabilidad de transición $Q(\cdot|x,\varphi)$, i.e.,

$$\mu_\varphi(\cdot) = \int_{\mathbf{X}} Q(\cdot|y,\varphi)\mu_\varphi(dy),$$

y, además,

$$J(\varphi,\mu_\varphi) = \int_{\mathbf{X}} C(y,\varphi)\mu_\varphi(dy) < +\infty.$$

Nos referiremos a μ_φ diciendo que es una medida de probabilidad invariante asociada a φ.
(b) Diremos que φ es *Harris recurrente* si existe una medida σ−finita λ_φ definida en $\mathbf{X}$ tal que el proceso inducido por φ es Harris recurrente [Definición A.D.11], es decir, si $\lambda_\varphi(B) > 0$, entonces

$$P_x^\varphi\left[\bigcap_{N=1}^{\infty}\bigcup_{k=N}^{\infty}\{x_t \in B\}\right] = 1 \quad \forall x \in \mathbf{X}.$$

La clase de las políticas estables y la clase de las políticas Harris recurrentes se denotarán por Φ_E y Φ_H, respectivamente.

Observación 6.2.6 Note que si una política es *estable* y *Harris recurrente*, entonces el proceso de Markov asociado a dicha política es *Harris recurrente positivo* [Definición A.D.11].

Observación 6.2.7. Sea $\varphi \in \Phi_E$ con medidad invariante asociada μ. Con cálculos directos se muestra que $J(\varphi,\mu) = j^*$ si sólo si $J(\varphi,x) = j^*$ μ−casi dondequiera.

6.3 Existencia de un par mínimo

En esta sección mostraremos que bajo las condiciones H6.2.1 y H6.2.2 existe un par mínimo (φ^*, μ^*), donde $\varphi^* \in \Phi_E$ y $\mu^* \in \mathbf{P(X)}$ es la medida invariante asociada a φ^* y que el costo promedio óptimo j^* es "límite" de problemas descontados. Esto último es importante, pues muestra otro aspecto del "buen comportamiento" del PCO en costo promedio para costos por etapa estrictamente no acotados y la plausibilidad de estudiar dicho problema por medio del enfoque de Aproximaciones por Problemas Descontados (ver Teorema 6.3.5). La existencia de un par mínimo (φ^*, μ^*) ya fue demostrada en [32], pero el análisis en este trabajo se dirige hacia los costos promedios esperados. Explícitamente, en esta referencia se demuestra lo siguiente:

(I) Para cada $\pi \in \Pi$ y $\nu \in \mathbf{P(X)}$, existe $\varphi \in \Phi_E$ tal que

$$J(\pi, \nu) \geq J(\varphi, \mu_\varphi),$$

donde μ_φ es una medida de probabilidad invariante asociada a φ. Por lo tanto,

$$j^* = \inf\{J(\varphi, \mu_\varphi) : \varphi \in \Phi_E\}.$$

(II) Existe $\varphi^* \in \Phi_E$, con medida invariante μ^*, tal que

$$J(\varphi^*, \mu^*) = j^*.$$

Los resultados principales de esta sección se presenta en el Teoremas 6.3.1 y 6.3.5, cuyos enunciados requerieren de la siguiente notación. Para cada $\alpha \in (0,1)$, considere las funciones

$$V_\alpha(\pi, \nu) \;:\; = E_\nu^\pi \sum_{t=0}^{\infty} \alpha^t C(x_t, a_t), \quad \pi \in \Pi, \nu \in \mathbf{P(X)}, \tag{6.5}$$

$$m_\alpha \;:\; = \inf_\nu \inf_\pi V_\alpha(\pi, \nu). \tag{6.6}$$

Teorema 6.3.1. Suponga que se satisfacen H6.2.1 y H6.2.2. Entonces, existe un par mínimo

(φ^*, μ^*), donde $\varphi^* \in \Phi_E$ y $\mu^* \in \mathbf{P}(\mathbf{X})$ es una medida de probabilidad invariante asociada a φ^*. Además,

$$J(\varphi^*, \mu^*) = j^* = \lim_{\alpha \to 1^-} (1-\alpha) m_\alpha.$$

La demostración del Teorema 6.3.1 se presentará después de algunos resultados preliminares que se reunen en las Observaciones 6.3.2, 6.3.3 y 6.3.4.

Observación 6.3.2.(a) Puesto que la función de costo por etapa C es no-negativa, del Teorema Abeliano A.E.5 (cf. Teorema 3.2.4) tenemos que

$$\liminf_{\alpha \to 1^-} (1-\alpha) m_\alpha \leq \limsup_{\alpha \to 1^-} (1-\alpha) m_\alpha \leq j^*; \tag{6.7}$$

(b) Por otra parte, note que dado $\varepsilon > 0$ arbitrario, para cada $\alpha \in (0,1)$ existen $\pi_\alpha \in \Pi$ y $\nu_\alpha \in \mathbf{P}(\mathbf{X})$ tales que

$$V_\alpha(\pi_\alpha, \nu_\alpha) \leq m_\alpha + \varepsilon;$$

en consecuencia,

$$\liminf_{\alpha \to 1^-} (1-\alpha) V_\alpha(\pi_\alpha, \nu_\alpha) = \liminf_{\alpha \to 1^-} (1-\alpha) m_\alpha \tag{6.8}$$

Observación 6.3.3.(a) Fijemos $\pi \in \Pi, \nu \in \mathbf{P}(\mathbf{X})$, $\alpha \in (0,1)$ y definamos la medida de probabilidad en $\mathbf{X} \times \mathbf{A}$

$$\gamma(\Gamma) := (1-\alpha) \sum_{t=0}^{+\infty} \alpha^t P_\nu^\pi [(x_t, a_t) \in \Gamma], \quad \Gamma \in \mathcal{B}(\mathbf{X} \times \mathbf{A}).$$

Note que $\gamma(\cdot)$ está concentrada en $\mathbf{K}$ y que para cada función medible no-negativa $v \in M_+(\mathbf{X} \times \mathbf{A})$,

$$(1-\alpha) E_\nu^\pi \sum_{t=0}^{+\infty} \alpha^t v(x_t, a_t) = \int_{\mathbf{K}} v(x,a) \gamma(d(x,a)). \tag{6.9}$$

En particular, se tiene que

$$(1-\alpha)V_\alpha(\pi,\mu) = (1-\alpha)E_\nu^\pi \sum_{t=0}^{+\infty} \alpha^t C(x_t,a_t) = \int_{\mathbf{K}} C(x,a)\gamma(d(x,a)). \tag{6.10}$$

(b) Del Teorema A.E.4, existe una política relajada φ y una distribución inicial $\mu \in \mathbf{P}(\mathbf{X})$ tales que

$$\gamma(B \times C) = \int_B \varphi(C|y)\mu(dy) \quad \forall B \times C \in \mathcal{B}(\mathbf{X} \times \mathbf{A}).$$

Usualmente, a la medida $\mu(\cdot)$ se le llama la *distribución marginal* o *proyección* de $\gamma(\cdot)$ sobre el espacio $\mathbf{X}$.

(c) Observe que para cada $B \in \mathcal{B}(\mathbf{X})$, se cumple que

$$\begin{aligned}
\mu(B) &= \gamma(B \times \mathbf{A}) \\
&= (1-\alpha)\sum_{t=0}^{+\infty} \alpha^t P_\nu^\pi[x_t \in B] \\
&= (1-\alpha)\left\{\nu(B) + \alpha \sum_{t=1}^{+\infty} \alpha^{t-1} P_\nu^\pi[x_t \in B]\right\}.
\end{aligned}$$

De la Observación 2.3.4(c) y propiedades de la esperanza condicional, se obtiene que

$$\mu(B) = (1-\alpha)\left\{\nu(B) + \alpha E_\nu^\pi \sum_{t=1}^{+\infty} \alpha^{t-1} Q(B|x_{t-1},a_{t-1})\right\}.$$

Entonces, de (6.9), se concluye que

$$\mu(B) = (1-\alpha)\nu(B) + \alpha \int_{\mathbf{K}} Q(B|x,a)\gamma(d(x,a)) \quad \forall B \in \mathcal{B}(\mathbf{X}). \tag{6.11}$$

La ecuación (6.11) juega un papel muy importante en la formulación del PCO α-Descontado como un problema de programación lineal (ver [34], Section 6.3). En nuestro caso, esta propiedad es clave para la demostración del Teorema 6.3.1.

Observación 6.3.4. Sean $\gamma(\cdot)$ y $\gamma_n(\cdot)$ medidas de probabilidad en $\mathbf{P}(\mathbf{X} \times \mathbf{A})$ y denotemos por $\mu(\cdot)$ y $\mu_n(\cdot)$ sus distribuciones marginales, respectivamente. Es decir,

$$\mu(B) := \gamma(B \times \mathbf{A}) \quad \text{y} \quad \mu_n(B) := \gamma_n(B \times \mathbf{A}), \quad B \in \mathcal{B}(\mathbf{X}).$$

Suponga que $\gamma_n(\cdot) \xrightarrow{w} \gamma(\cdot)$, es decir,

$$\int_{\mathbf{X}\times\mathbf{A}} v(x,a)\gamma_n(d(x,a) \to \int_{\mathbf{X}\times\mathbf{A}} v(x,a)\gamma(d(x,a),$$

para cada función continua y acotada $v \in C_a(\mathbf{X} \times \mathbf{A})$. Entonces,

$$\mu_n(\cdot) \xrightarrow{w} \mu(\cdot).$$

Demostración del Teorema 6.3.1. Para cada $\alpha \in (0,1)$, sean $\pi_\alpha \in \Pi$ y $\nu_\alpha \in \mathbf{P}(\mathbf{X})$ como en la Observación 6.3.2(b) y una sucesión $\{\alpha(n)\} \subset (0,1)$ convergente a 1 tales que

$$\begin{aligned} \rho^* &:= \liminf_{\alpha\to 1^-} (1-\alpha)m_\alpha \\ &= \liminf_{\alpha\to 1^-} (1-\alpha)V_\alpha(\pi_\alpha, \nu_\alpha) \\ &= \lim_{n\to\infty} (1-\alpha(n))V_{\alpha(n)}(\pi_{\alpha(n)}, \nu_{\alpha(n)}). \end{aligned}$$

Para cada $n \in \mathbf{N}$, defina las medida de probabilidad

$$\gamma_n(\Gamma) := (1-\alpha(n)) \sum_{t=0}^{+\infty} \alpha_n^t P_{\nu_{\alpha(n)}}^{\pi_{\alpha(n)}}[(x_t, a_t) \in \Gamma], \quad \Gamma \in \mathcal{B}(\mathbf{X} \times \mathbf{A}),$$

y note que está concentrada en $\mathbf{K}$. Entonces, de (6.7), (6.8) y (6.10) tenemos que

$$\rho^* = \lim_{n\to\infty} \int_{\mathbf{K}} C(x,a)\gamma_n(d(x,a)) \le j^*,$$

lo cual implica que

$$\sup_n \int_{\mathbf{K}} C(x,a)\gamma_n(d(x,a)) < +\infty.$$

Por los Teoremas A.C.6 y A.C.4 existe una medida de probabilidad $\gamma^*(\cdot) \in \mathbf{P}(\mathbf{K})$ y una subsucesión de $\{\gamma_n(\cdot)\}$, que denotaremos de nuevo como $\{\gamma_n(\cdot)\}$ para no complicar la notación, tales que

$$\gamma_n(\cdot) \xrightarrow{w} \gamma^*(\cdot). \tag{6.12}$$

Entonces, de la semicontinuad inferior de C y de (6.12), se obtiene

$$\rho^* = \lim_{n\to\infty} \int_{\mathbf{K}} C(x,a)\gamma_n(d(x,a)) \geq \int_{\mathbf{K}} C(x,a)\gamma^*(d(x,a)). \tag{6.13}$$

La parte restante de la demostración consiste en mostrar que

$$\int_{\mathbf{K}} C(x,a)\gamma^*(d(x,a)) = J(\varphi^*,\mu^*) \geq j^*,$$

donde $\varphi^* \in \Phi_E$ y μ^* es una medida invariante asociada. Para hacer esto, observe por el Teorema A.E.4 existen políticas relajadas $\varphi(\cdot|\cdot), \varphi_n(\cdot|\cdot)$ y distribuciones iniciales $\mu^*(\cdot), \mu_n(\cdot)$ tales que para todo $B \times C \in \mathcal{B}(\mathbf{X} \times \mathbf{A})$:

$$\gamma^*(B \times C) = \int_B \varphi^*(C|y)\mu^*(dy) \quad \text{y} \quad \gamma_n(B \times C) = \int_B \varphi_n(C|y)\mu_n(dy) \quad \forall n \in \mathbf{N}. \tag{6.14}$$

Ahora, por la Observación 6.3.4 y (6.12), note que las distribuciones $\mu_n(\cdot)$ convergen débilmente a $\mu^*(\cdot)$. Es decir,

$$\int_{\mathbf{X}} v(dy)\mu_n(dy) \to \int_{\mathbf{X}} v(dy)\mu^*(dy) \quad \forall v \in C_a(\mathbf{X}). \tag{6.15}$$

Por otra parte, de la Observación 6.3.3 y (6.11), para cada $n \in \mathbf{N}$ y $B \in \mathcal{B}(\mathbf{X})$:

$$\mu_n(B) = (1-\alpha(n))\nu_{\alpha(n)}(B) + \alpha(n)\int_{\mathbf{K}} Q(B|x,a)\gamma_n(d(x,a)),$$

lo cual implica que

$$\int_{\mathbf{X}} v(y)\mu_n(dy) = (1-\alpha(n))\int_{\mathbf{X}} v(y)\nu_{\alpha(n)}(dy) + \alpha(n)\int_{\mathbf{K}}\int_{\mathbf{X}} v(y)Q(dy|x,a)\gamma_n(d(x,a)) \quad (6.16)$$

para toda función $v \in C_a(\mathbf{X})$.

Ahora note que para cada $v \in C_a(\mathbf{X})$, la sucesión

$$\int_{\mathbf{X}} v(y)\nu_{\alpha(n)}(dy)$$

es acotada y que, por H6.2.1(c), la función

$$\int_{\mathbf{X}} v(y)Q(dy|\cdot,\cdot) \in C_a(\mathbf{K}).$$

Entonces, tomando límite cuando $n \to \infty$ en (6.16), se deduce de (6.12) y (6.15) que

$$\int_{\mathbf{X}} v(y)\mu^*(dy) = \int_{\mathbf{K}}\int_{\mathbf{X}} v(y)Q(dy|x,a)\gamma^*(d(x,a)) \quad \forall v \in C_a(\mathbf{X}),$$

lo cual implica que

$$\int_{\mathbf{X}} v(y)\mu^*(dy) = \int_{\mathbf{X}} v(y)Q(dy|x,\varphi^*)\mu^*(dy) \quad \forall v \in C_a(\mathbf{X}).$$

Es decir, $\mu^*(\cdot)$ es una medida de probabilidad invariante para la probabilidad de transición

$$Q(\cdot|x,\varphi^*) := \int_{\mathbf{A}} Q(\cdot|x,a)\varphi^*(da|x), \quad x \in \mathbf{X},$$

lo cual implica que

$$\int_{\mathbf{K}} C(x,a)\gamma^*(d(x,a)) = \int_{\mathbf{X}} C(x,\varphi^*)\mu^*(dx) = J(\varphi^*,\mu^*) \geq j^*,$$

donde

$$C(x,\varphi^*) := \int_{\mathbf{A}} C(x,a)\varphi^*(da|x), \quad x \in \mathbf{X}.$$

Por lo tanto, de la Observación 6.3.2(a), tenemos que

$$j^* = J(\varphi^*, \mu^*) = \lim_{\alpha \to 1^-} (1-\alpha) m_\alpha. \blacksquare$$

Concluimos esta sección mostrando que bajo condiciones adecuadas se puede obtener por medio del enfoque de Aproximaciones por Problemas Descontados una solución de la Desigualdad de Optimalidad en Costo Promedio y una política estacionaria determinista $f^* \in \mathbf{F}$ tal que $J(f^*, x) = j^*$ $\forall x \in \mathbf{X}$.

Teorema 6.3.5. Suponga que se satisfacen H6.2.1 y H6.2.2. Sea $\{\alpha(n)\} \subset (0,1)$ una sucesión convergente a 1 y (φ, μ) una par mínimo, donde $\varphi \in \Phi_E$ y $\mu \in \mathbf{P}(\mathbf{X})$ una medida invariante asociada. Si la función

$$h(x) := \liminf_{n \to \infty} [V_{\alpha(n)}(\varphi, x) - m_{\alpha(n)}], \quad x \in \mathbf{X}, \tag{6.17}$$

es finita para cada $x \in \mathbf{X}$, entonces existe $f^* \in \mathbf{F}$ tal que

$$j^* + h(x) \geq C(x, f^*) + \int_{\mathbf{X}} h(y) Q(dy|x, f^*) \quad \forall x \in \mathbf{X}, \tag{6.18}$$

y

$$J(f^*, x) = j^* \quad \forall x \in \mathbf{X}. \tag{6.19}$$

Demostración del Teorema 6.3.5. Para cada $n \in \mathbf{N}$, defina las funciones

$$h_n(x) := V_{\alpha(n)}(\varphi, x) - m_{\alpha(n)}, \quad x \in \mathbf{X},$$

y note que son no-negativas. Además observe que se satisfacen las ecuaciones

$$V_{\alpha(n)}(\varphi, x) = C(x, \varphi) + \alpha(n) \int_{\mathbf{X}} V_{\alpha(n)}(\varphi, y) Q(dy|x, \varphi) \quad \forall x \in \mathbf{X}, n \in \mathbf{N}$$

las cuales pueden re-escribirse equivalentemente como

$$(1-\alpha(n)) + h_n(x) = C(x,\varphi) + \alpha(n)\int_{\mathbf{X}} h_n(y)Q(dy|x,\varphi) \quad \forall x \in \mathbf{X}, n \in \mathbf{N}. \tag{6.20}$$

Del Teorema 6.3.1, de la no-negatividad de las funciones $h_n(\cdot)$ y el Lema de Fatou, al tomar límite inferior cuando $n \to \infty$ en (6.20), se obtiene la desigualdad

$$j^* + h(x) \geq C(x,\varphi) + \int_{\mathbf{X}} h(y)Q(dy|x,\varphi) \quad \forall x \in \mathbf{X}. \tag{6.21}$$

Por (6.21) y el Lemma 5.5 en [32], existe una política estacionaria determinista $f^* \in \mathbb{F}$ que satisface (6.18). Finalmente, (6.19) se obtiene de (6.18) y la no-negatividad de $h(\cdot)$ siguiendo argumentos similares a los de la demostración del Teorema 3.3.3. ∎

6.4 Existencia de políticas óptimas por trayectorias

Esta sección contiene los resultados principales del capítulo, los cuales se ilustran en la Sección 6.5 con algunos ejemplos de inventarios.

Teorema 6.4.1. Suponga que se satisfacen H6.2.1 y H6.2.2. Entonces

(a) para cada política $\pi \in \Pi$ y distribución inicial $\nu \in \mathbf{P}(\mathbf{X})$, se cumple que

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{t=0}^{n-1} C(x_t, a_t) \geq j^* \quad P_\nu^\pi - \text{casi seguramente}; \tag{6.22}$$

(b) Si la política $\varphi^* \in \Phi_E$ en el Teorema 6.3.1 es Harris recurrente, entonces es óptima en costo promedio por trayectorias.

Demostración del Teorema 6.4.1. Primero demostraremos que (b) es una consecuencia de (a). Por la Observación 6.3.2, el proceso $\{x_t\}$ inducido por φ^* es Harris recurrente positivo; entonces, por la Ley de Grandes Números para Procesos de Markov (Teorema A.D.12), para cada $\nu \in \mathbf{P}(\mathbf{X})$:

$$J_0(\varphi^*, \nu) = \lim_{n\to\infty} \frac{1}{n}\sum_{t=0}^{n-1} C(x_t, a_t) = j^* \quad P_\nu^{\varphi^*} - \text{casi seguramente}.$$

Combinando lo anterior con Teorema 6.4.1(a), se concluye que φ^* es óptima por trayectorias.

Ahora, demostraremos que (6.22) se cumple. Para hacer esto, fije $\pi \in \Pi$, $\nu \in \mathbf{P}(\mathbf{X})$ y considere la variable aleatoria

$$\widehat{J} := \liminf_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} C(x_t, a_t). \tag{6.23}$$

definida en $(\Omega, \mathcal{F})$, donde $\Omega = (\mathbf{X} \times \mathbf{A})^\infty$ y $\mathcal{F}$ es la $\sigma-$álgebra producto correspondiente. Observe que podemos suponer sin pérdida de generalidad que $\widehat{J}$ es una variable aleatoria finita.

Defina las *medidas (de ocupación) empíricas* por

$$\gamma_n(\Gamma) := \frac{1}{n} \sum_{t=0}^{n-1} \mathbf{I}_\Gamma[(x_t, a_t)], \quad n \in \mathbf{N}, \Gamma \in \mathcal{B}(\mathbf{X} \times \mathbf{A}),$$

donde $\mathbf{I}_\Gamma(\cdot)$ denota la función indicadora del conjunto Γ. Note que las medidas $\{\gamma_n(\cdot)\}$ están concentradas en $\mathbf{K}$ y que

$$\widehat{J} = \liminf_{n\to\infty} \int_{\mathbf{K}} C(x,a)\gamma_n(d(x,a)).$$

Sea $\omega \in \Omega$ fijo y tomemos una subsucesión $\{n_k\} = \{n_k(\omega)\}$ tal que

$$+\infty > \widehat{J} = \lim_{k\to\infty} \int_{\mathbf{K}} C(x,a)\gamma^\omega_{n_k}(d(x,a)). \tag{6.24}$$

De H6.2.1(b) y la Proposición A.C.6, tenemos que la sucesión de medidas $\{\gamma^\omega_{n_k}(\cdot)\}$ es *tensa*; luego, por el Teorema de Prohorov (Teorema A.C.4), existe una medida de probabilidad $\gamma^\omega(\cdot) \in \mathbf{P}(\mathbf{K})$ y una subsucesión $\{m_k\}$ de $\{n_k\}$ tal que

$$\gamma^\omega_{m_k}(\cdot) \text{ converge débilmente a } \gamma^\omega(\cdot). \tag{6.25}$$

Entonces, de H6.2.1(a) y (6.25), se deduce que

$$\widehat{J} = \liminf_{n\to\infty} \int_{\mathbf{K}} C(x,a)\gamma^\omega_n(d(x,a)) \geq \int_{\mathbf{K}} C(x,a)\gamma^\omega(d(x,a)).$$

Por otra parte, del Teorema A.E.4, se tiene que la medida $\gamma^\omega(\cdot)$ se puede descomponer como

$$\gamma^{\omega}(B \times D) = \int_B \varphi^{\omega}(D|x)\mu^{\omega}(dx), \quad B \times D \in \mathcal{B}(\mathbf{X} \times \mathbf{A}), \tag{6.26}$$

donde $\varphi^{\omega} \in \Phi$, es decir, φ^{ω} es una política relajada y $\mu^{\omega}(\cdot) \in \mathbf{P}(\mathbf{X})$. De ésto se deduce que

$$\widehat{J} \geq \int_{\mathbf{X}} C(x, \varphi^{\omega})\mu^{\omega}(dx)$$

donde

$$C(x, \varphi^{\omega}) := \int_{\mathbf{A}} C(x, a)\varphi^{\omega}(da|x), \quad x \in \mathbf{X}.$$

Ahora demostraremos que (P_{ν}^{π}−casi seguramente) $\mu^{\omega}(\cdot)$ es una *medida de probabilidad invariante* para la probabilidad de transición

$$Q(\cdot|x, \varphi^{\omega}) := \int_{\mathbf{A}} Q(\cdot|x, a)\varphi^{\omega}(da|x), \quad x \in \mathbf{X};$$

en consecuencia, φ^{ω} es una política *estable*. Con ésto, habremos completado la demostración, ya que en este caso tendríamos que

$$\widehat{J} \geq \int_{\mathbf{X}} C(x, \varphi^{\omega})\mu^{\omega}(dx) = J(\varphi^{\omega}, \mu^{\omega}) \geq j^{*}. \tag{6.27}$$

Del Teorema A.E.1, existe una métrica d^{*} en $\mathbf{X}$ tal que el subespacio $\mathcal{U}_{d^{*}}(\mathbf{X})$ $[\subset \mathcal{C}_u(\mathbf{X})]$ de las funciones uniformemente continuas con respecto a d^{*} es *separable*; es decir, existe una subclase $\mathcal{U}$ de $\mathcal{U}_{d^{*}}(\mathbf{X})$, la cual es *numerable* y *densa* en $\mathcal{U}_{d^{*}}(\mathbf{X})$. Para cada $u \in \mathcal{U}$, defina la función

$$Lu(x, a) := \int_{\mathbf{X}} u(y)Q(dy|x, a) - u(x), \quad (x, a) \in \mathbf{K},$$

y el proceso estocástico

$$M_0(u) \quad : \quad = u(x_0)$$

$$M_n(u) \quad : \quad = u(x_n) - \sum_{t=0}^{n-1} Lu(x_t, a_t), \quad n \geq 1.$$

Note que para cada $u \in \mathcal{U}$, la función $Lu \in \mathcal{C}_a(\mathbf{K})$ y que el proceso $\{M_n(u)\}$ es una martingala con respecto a la filtración $\{\sigma(h_n, a_n)\}$. Entonces, de la Ley de Grandes Números para Martingalas (Teorema A.E.3), se tiene que para cada $u \in \mathcal{U}$, existe $U_u \in \mathcal{F}$ tal que $P_\nu^\pi(U_u) = 1$ y

$$\lim_{n\to\infty} \frac{1}{n} M_n^\omega(u) = 0 \quad \forall \omega \in U_u.$$

Observe también que

$$\frac{1}{n} M_n^\omega(u) = \frac{1}{n} u(x_n) - \int_{\mathbf{K}} Lu(x,a) \gamma_n^\omega(d(x,a)).$$

Entonces, puesto que u es una función acotada, se tiene que

$$\lim_{n\to\infty} \int_{\mathbf{K}} Lu(x,a) \gamma_n^\omega(d(x,a)) = 0 \quad \forall \omega \in U_u,$$

lo cual implica que

$$\lim_{n\to\infty} \int_{\mathbf{K}} Lu(x,a) \gamma_n^\omega(d(x,a)) = 0 \quad \forall \omega \in U = \bigcap_{u\in\mathcal{U}} U_u, \text{ y } u \in \mathcal{U}. \tag{6.28}$$

Para cada $\omega \in U$, considere la sucesión $\{m_k\} = \{m_k(\omega)\}$ y la medida $\gamma^\omega(\cdot)$ como en (6.25). Entonces,

$$\int_{\mathbf{K}} Lu(x,a) \gamma^\omega(d(x,a)) = 0 \quad \forall u \in \mathcal{U},$$

lo cual implica, por el Teorema A.E.1, que

$$\int_{\mathbf{K}} Lu(x,a) \gamma^\omega(d(x,a)) = 0 \quad \forall u \in \mathcal{C}_a(\mathbf{X}). \tag{6.29}$$

A su vez, usando la descomposición (6.26), la relación en (6.29) implica que

$$\int_{\mathbf{X}} u(x)\mu^{\omega}(dx) = \int_{\mathbf{X}}\int_{\mathbf{X}} u(y)Q(dy|x,\varphi^{\omega})\mu^{\omega}(dx) \quad \forall u \in \mathcal{C}_a(\mathbf{X}).$$

Por lo tanto, para cada $\omega \in U$, $\mu^{\omega}(\cdot)$ es una medida de probabilidad invariante para $Q(\cdot|\cdot,\varphi^{\omega})$ y, en consecuencia, (6.27) se cumple. Finalmente, para completar la demostración note que $P_{\nu}^{\pi}(U) = 1$; entonces,

$$\widehat{J} = \liminf_{n\to\infty} \int_{\mathbf{K}} C(x,a)\gamma_n^{\omega}(d(x,a)) \geq j^* \quad P_{\nu}^{\pi} - \text{casi seguramente.} \ \blacksquare$$

Finalizamos esta sección presentando algunas consecuencias inmediatas, pero interesantes, del Teorema 6.4.2

Teorema 6.4.2. Suponga que se cumplen H6.2.1 y H6.2.2. Entonces:

(a) una política $\pi^* \in \Pi$ es óptima en costo promedio si y sólo si es fuertemente óptima;

(b) si $\pi^* \in \Pi$ es óptima en costo promedio, entonces

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{t=0}^{n-1} C(x_t,a_t) = j^* \quad P_{\nu}^{\pi^*} - \text{casi seguramente.} \tag{6.30}$$

(c) $j^* = \inf_{\nu}\inf_{\pi}\left\{\liminf_{n\to\in} \frac{1}{n} J_n(\pi,\nu)\right\}.$

Demostración del Teorema 6.4.2(a). Observe que la parte "sólo si" se obtiene directamente de las definiciones, de manera que sólo falta demostrar que si $\pi^* \in \Pi$ es una política óptima en costo promedio esperado, entonces es fuertemente óptima. Lo anterior se deduce de (6.22) y del Lema de Fatou; de hecho, se obtiene que

$$J(\pi,\nu) \geq \liminf_{n\to\infty} \frac{1}{n} E_{\nu}^{\pi}\sum_{t=0}^{n-1} C(x_t,a_t) \geq j^* = J(\pi^*,\nu) \quad \pi \in \Pi,\ \nu \in \mathbf{P}(\mathbf{X}). \tag{6.31}$$

(b) De la optimalidad de π^*, del Teorema 6.4.1(a) y el Lema de Fatou, se obtiene que

$$j^* \geq E_{\nu}^{\pi^*}\left[\liminf_{n\to\infty} \frac{1}{n}\sum_{t=0}^{n-1} C(x_t,a_t)\right] \geq j^*,$$

lo cual implica, por el Teorema 6.4.1(a) de nuevo, que (6.30) se cumple para cada distribución inicial $\nu \in \mathbf{P}(\mathbf{X})$.

(c) Esta parte se sigue de las dos primeras desigualdades en (6.31). ■

6.5 Ejemplos

Ahora, discutiremos algunos ejemplos de inventarios con la idea de ilustrar las posibilidades de aplicación del enfoque desarrollado en las secciones precedentes. De hecho, en el Ejemplo B, se calcula explícitamente una política estacionaria Ha Harris recurrente positiva la cual es óptima en costo promedio por trayectorias. En [32], [35] y [49] se estudian otros ejemplos interesantes que satisfacen las condiciones de los Teoremas 6.4.1 y 6.4.2, entre los cuales se incluye el problema de control para sistemas lineales con costos cuadráticos y perturbaciones gaussianas.

En lo que sigue, estudiaremos un sistema de inventario con capacidades de producción y almacenamiento infinitas, en el cual el "exceso" de demanda se pierde. Denotaremos por x_t y a_t el nivel de inventario y la cantidad producto ordenada a la unidad de producción, respectivamente, al iniciar el periodo t. La demanda del producto durante el periodo t es una variable aleatoria no negativa que representaremos por w_t. El sistema de inventario evoluciona de acuerdo a la dinámica

$$x_{t+1} = \max(x_t + a_t - w_t, 0), \quad t \in \mathbf{N}; \quad x_0 = x, \tag{6.32}$$

tomando valores en el espacio $\mathbf{X} := [0, +\infty)$, mientras que las variables de control $\{a_t\}$ toman valores en $\mathbf{A} := [0, +\infty)$ independientemente del nivel de inventario, es decir, $\mathbf{A} = A(x)\ \forall x \in \mathbf{X}$.

También suponemos que el proceso de demanda $\{w_t\}$ satisface las siguientes condiciones.

H6.5.1(a) Las variables aleatorias $\{w_t\}$ son independientes e idénticamente distribuidas; la función de distribución será denotada por $G(\cdot)$;

(b) $G(y) < 1\ \forall y \geq 0$.

La esperanza con respecto a la distribución conjunta de las variables $\{w_t\}$ se denotará como E.

Observación 6.5.2. La ley de evolución (6.32) se puede expresar equivalentemente por medio de la probabilidad de transición

$$Q(B|x,a) := \int_0^{+\infty} \mathbf{I}_B[(x+a-w)^+]G(dw), \quad B \in \mathcal{B}(\mathbf{X}), (x,a) \in \mathbf{K}.$$

Ademas, es fácil verificar que

$$\int_{\mathbf{X}} u(y)Q(dy|x,a) = Eu[(x+a-w_0)^+] \;\; \forall (x,a) \in \mathbf{K},$$

donde u es una funcion medible. De esto, se observa que H6.5.1(a) implica a H6.2.1(c).

(b) Por otro lado, H6.5.1(b) implica que cada política estable φ es Harris recurrente con respecto a la medida $\lambda(B) := \mathbf{I}_B(0)$, $B \in \mathcal{B}(\mathbf{X})$. En consecuencia, $\mu_\varphi(\{0\}) > 0$, donde $\mu_\varphi(\cdot)$ es la medida invariante asociada a φ.

(c) Especificamente, la política estacionaria

$$f_K(x) := \begin{cases} K - x & \text{si} \quad 0 \leq x \leq K \\ \\ 0 & \text{si} \quad x > K \end{cases} \tag{6.33}$$

con $K \geq 0$, es Harris recurrente positiva. De hecho, $\mu_K(B) := \int_B (K-w)^+ G(dw)$ es la medida de probabilidad invariante asociada a f_K.

EJEMPLO A. Suponemos que la función de costo por etapa tiene la forma siguiente

$$C(x,a) = F_1(x+a) + F_2(a), \quad (x,a) \in \mathbf{K}, \tag{6.34}$$

donde $F_1(\cdot)$ y $F_2(\cdot)$ son funciones definidas en el intervalo $[0, +\infty)$ que satisfacen las siguientes condiciones.

H6.5.3(a) $F_1(\cdot)$ y $F_2(\cdot)$ son funciones semicontinuas inferiormente;

(b) existen sucesiones de reales positivos $\{y_n^1\}$ y $\{y_n^2\}$ que divergen a infinito tales que

$$\lim_{n\to\infty} \inf_{y > y_n^i} F_i(y) = +\infty, \text{ para } i = 1, 2;$$

(c) $EF_2[\min(y, w_0)] < +\infty \ \forall y \geq 0$.

Observe que las condiciones en H6.5.3 son lo suficientemente generales como para incluir problemas de inventarios en los cuales se tiene un costo fijo por ordenar ([6]).

Teorema 6.5.4. Si se satisfacen H6.5.1 y H6.5.3, entonces existe una política $\varphi^* \in \Phi_E \cap \Phi_H$ óptima en costo promedio por trayectorias. Si además se satisfacen las condiciones del Teorema 6.3.1, entonces existe una política estacionaria $f^* \in \mathbf{F}$ tal que $J(f^*, x) = j^* \ \forall x \in \mathbf{X}$.

Demostración del Teorema 6.5.4. Note que H6.5.3(a) implica que la función de costo en (6.34) satisface las condiciones H6.2.1(a)-(b). Para completar la demostración, de la Observación 6.5.2, solamente hace falta verificar que H6.2.2 se cumple. Para hacer esto, considere la política en (6.33) y note que

$$J(f_K, x) = F_1(K) + EF_2[\min(K, w_0)] < +\infty \ \forall \in \mathbf{X}. \ \blacksquare \tag{6.35}$$

EJEMPLO B. Ahora consideraremos un caso particular de (6.34), para el cual es posible encontrar explicitamente una política estacionaria óptima por trayectorias. Tomaremos $F_2(y) := by, \ y \geq 0$, de manera que la función (6.34) se convierte en

$$C(x, a) = F_1(x + a) + ba, \ \ (x, a) \in \mathbf{K}; \tag{6.36}$$

además, en lugar de las condiciones H6.5.3, suponemos que se cumplen las siguientes.

H6.5.5(a) $F_1(\cdot)$ es convexa y acotada inferiormente;

(b) $\lim_{y \to +\infty} F_1(y) = +\infty$.

Note que para la elección específica que aquí estamos considerando H6.5.5 implica H6.5.3, de manera que bajo las condiciones H6.5.1 y H6.5.5, los resultados en el Teorema 6.5.4 se cumplen. A continuación mostraremos que una política de la forma (6.33) es óptima en costo promedio por trayectorias. Defina

$$L(y) := F_1(y) + bE\min(y, w_0), \ y \geq 0, \quad \text{y} \quad \phi^* := \inf_{y \geq 0} L(y). \tag{6.37}$$

Observación 6.5.6.(a) De (6.35) y (6.37), se tiene que

$$J(f_K, x) = L(K) \quad \forall x \in \mathbf{X};$$

(b) Puesto que $L(\cdot)$ es continua y $\lim_{y\to+\infty} L(y) = +\infty$, existe $K^* \geq 0$ tal que $L(K^*) = \phi^*$.

Teorema 6.5.7. Si se cumplen las condiciones H6.5.1 y H6.5.5, entonces la política estacionaria

$$f_{K^*}(x) := \begin{cases} K^* - x & \text{si} \quad 0 \leq x \leq K^* \\ \\ 0 & \text{si} \quad x > K^* \end{cases}$$

donde K^* es una constante como en la Observación 6.5.6(b), satisface lo siguiente:

(a) $J(f_{K^*}, x) = j^* \quad \forall x \in \mathbf{X};$

(b) f_{K^*} es óptima en costo promedio por trayectorias.

Demostración del Teorema 6.5.7. Primero demostraremos que (a) $\to$ (b). De la Observación 6.5.2(c), tenemos que $f_{K^*} \in \Phi_E \cap \Phi_H$. Entonces de la Observación 6.2.7, tenemos que (f_{K^*}, μ_{K^*}) es un par mínimo donde $\mu_{K^*}(\cdot)$ es la medida invariante asociada a f_{K^*}. Por lo tanto, por el Teorema 6.4.1, f_{K^*} es óptima en costo promedio por trayectorias.

La demostración de la parte (a) sólo la bosquejaremos puesto que se obtiene con argumentos similares a los de la demostración del Teorema 3.4.6. Recuerde que para cada $\alpha \in (0, 1)$,

$$V_\alpha(\pi, x) = E_x^\pi \sum_{t=0}^{+\infty} \alpha^t C(x_t, a_t), \quad \pi \in \Pi, x \in \mathbf{X};$$

además, defina

$$V_\alpha(x) := \inf_\pi V_\alpha(\pi, x), \quad x \in \mathbf{X}.$$

Observe que $m_\alpha \leq V_\alpha(\cdot) \; \forall \alpha \in (0, 1)$, de manera que

$$\liminf_{\alpha\to 1^-} (1-\alpha) V_\alpha(x) \geq \lim_{\alpha\to 1^-} (1-\alpha) m_\alpha = j^* \quad \forall x \in \mathbf{X} \tag{6.38}$$

Por otra parte, bajo H6.5.1 y H6.5.5, por el Teorema 6.4.1 existe una política $\varphi^* \in \Phi_E \cap$

Φ_H tal que $J(\varphi^*, \mu^*) = j^*$. Por la o¡Observación 6.5.2(b), **tenemos que** $\mu^*(\{0\}) > 0$ **lo cual** combinado con el Teorema (Abeliano) A.E.5 y (6.38), implica que

$$j^* = \lim_{\alpha \to 1^-} (1-\alpha) V_\alpha(\varphi^*, 0) = \lim_{\alpha \to 1^-} (1-\alpha) V_\alpha(0) = j^*.$$

Por lo tanto, para demostrar la parte (a) es suficiente verificar que

$$\phi^* = L(K^*) = \lim_{\alpha \to 1^-} (1-\alpha) V_\alpha(0) = j^*. \tag{6.39}$$

Las igualdades en (6.39) se obtiene con argumentos similares a los de la demostración del Teorema 3.4.6(a), razon por la cual la omitimos. ■

Observación 6.5.8. En la Sección 3.3.4, se estudió el problema de inventarios con la función de costo (6.36) siguiendo el enfoque *Aproximación por Problemas Descontados*, pero en lugar de la condición H6.5.1(b) se usó la siguiente.

H6.5.1(b'). La función de distribución $G(\cdot)$ tiene una función de densidad continua y acotada.

Bajo las condiciones H6.5.1(a) y H6.5.1(b') se demostró en la Sección 3.3.4 que

$$J(f_{K^*}, x) = \phi^* = \lim_{\alpha \to 1} (1-\alpha) V_\alpha(x) \quad \forall x \in \mathbf{X},$$

lo cual implica que f_{K^*} es óptima en costo promedio por trayectorias.

EJEMPLO C. Otra posibilidad para evaluar el funcionamiento de un sistema de inventarios consiste en considerar costos cuadráticos, es decir,

$$C(x,a) = S_1(x - x^*)^2 + S_2(a - a^*)^2, \quad (x,a) \in \mathbf{K}, \tag{6.40}$$

donde $x^* \subset \mathbf{X}$ y $a^* \subset \mathbf{A}$ son fijos y representan los niveles de inventario y producción nominales; además S_1 y S_2 son constantes positivas. Para este ejemplo, además de H6.5.1, suponemos que cumple lo siguiente.

H6.5.9. El segundo momento de la variable aleatoria w_0 es finito, es decir, $\int_{\mathbf{X}} y^2 G(dy) < +\infty$.

Es claro que (6.40) cumple las condiciones en H6.2.1(a)-(b), mientras que la condición H6.5.9

garantiza que H6.2.2 se satisface. De hecho, cálculos directos muestran que

$$J(f,x) = S_1(x^*)^2 + S_2(a^*)^2 \quad \forall x \in \mathbf{X},$$

donde $f(\cdot) \equiv 0$. Por lo tanto, bajo H6.5.1 y H6.5.9, para el problema de inventarios con función de costo (6.40) se obtienen exactamente las mismas conclusiones del Teorema 6.5.4.

EJEMPLO D. En [58] se estudia un sistema de inventario en horizonte finito y consideran una variante de (6.40), en la cual se introduce un intervalo "sin costo" que contiene al nivel de inventario nominal x^*. De forma más precisa, la desviación del estado nominal se penaliza a través de la función

$$\widehat{C}(y) := \begin{cases} R_1(y-\alpha)^2 & \text{si} \quad 0 \le y < \alpha \\ 0 & \text{si} \quad \alpha \le y \le \beta \\ R_2(y-\beta)^2 & \text{si} \quad y > \beta, \end{cases}$$

donde $0 < \alpha < \beta$ y R_1, R_2 son constantes positivas, y la función de costo en una etapa la toman como

$$C(x,a) = E\widehat{C}(x + a - w_0) + S_2(a - a^*)^2, \quad (x,a) \in \mathbf{K}.$$

De nuevo es fácil verificar que se satisfacen las condiciones en el Teorema 6.4.2.

6.6 Conclusiones

El análisis por trayectorias del problema en costo promedio resulta muy atractivo desde el punto de vista de las aplicaciones, sin embargo, los trabajos en los que se realiza son escasos y se restringen al caso *discreto o de Borel y costos acotados* ([3], [2], [10], [15], [44]) y, hasta donde conocemos, [77] es el primer trabajo donde se consideran *espacios de Borel y costos no acotados*. Basándonos en la última referencia, en este capítulo se mostró (bajo condiciones débiles de continuidad y recurrencia) para problemas con costos estrictamente no acotados la existencia de pares mínimos y políticas óptimas por trayectorias. También se mostró que, para este tipo de costos sin imponer condiciones adicionales, el enfoque de Aproximaciones por

Problemas Descontados funciona parcialmente y, entonces, ofrece una buena alternativa para la determinación tanto del costo promedio óptimo como de políticas óptimas. Los resultados principales se ilustraron con problemas de inventarios y, en un caso especifico, se exhibió una política estacionaria óptima por trayectorias.

Capítulo 7

Conclusiones y Problemas Abiertos

En este trabajo se estudiaron procesos de Markov controlados en espacios de *Borel* y *costos no-acotados* con respecto a los siguientes criterios (de optimalidad) *no-descontados*: criterio en costo promedio (esperado), criterio en costo promedio por trayectorias, políticas fuertemente óptimas (en costo promedio), optimalidad en el sentido de Flynn, políticas dominantes y fuertemente dominantes, criterio en costo de oportunidad, criterio de Dutta y optimalidad en sesgo. Se proporcionaron condiciones suficientes para la existencia de políticas estacionarias óptimas con respecto a cada uno de estos criterios (excepto para polílticas fuertemente dominantes), se establecieron algunas relaciones entre ellos y los resultados obtenidos se ilustraron con problemas de control en sistemas de inventarios.

Para abordar los distintos problemas planteados se usaron los siguientes enfoques o métodos: a) Aproximaciones por Problemas Descontados (APD), b) Funciones de Lyapunov (FL); c) Algoritmo de Iteración de Políticas (AIP); d) Método de Momentos (MM) [o funciones estrictamente no-acotadas]. Estos enfoques, si bien es cierto que son distintos, son complementarios. Por ejemplo, en el Capítulo 4, se usaron los tres primeros en forma combinada para obtener una solución de la Ecuación de Optimalidad en Costo Promedio y en el Ejemplo B de la Sección 6.5 se hizo una combinación "parcial" del enfoque APD con el Método de Momentos para calcular explicitamente una política estacionaria óptima por trayectorias (cf. Teorema 6.3.5). Una posibilidad más sería combinar los métodos de Funciones de Lyapunov y de Momentos (e.g., el ejemplo de la Sección 4.6 tambíen satisface las condiciones del Teorema 6.4.2 el cual garantiza la existencia de políticas óptimas en costo promedio por trayectorias). Lo anterior sugiere un buen número de problemas, interesantes desde nuestro punto de vista, que no se abordaron en este trabajo y que se describen brevemente en la sección siguiente.

7.1 Problemas Abiertos

A. Aproximaciones por Problemas Descontados y Costos Estrictamente No Acotados.

Como se ha mencionado reiteradamente, el enfoque de Aproximaciones por Problemas Descontados se ha estudiado intensamente en los últimos años y se ha mostrado que puede adaptarse a contextos muy generales. Sin embargo, su flexibilidad se debe a que impone condiciones sobre "cantidades derivadas", no sobre los objetos básicos del modelo de control, lo cual representa una desventaja ya que, en general, resulta difícil verificar si tales condiciones se satisfacen en problemas específicos.

Conjeturamos que para el caso de costos estrictamente no acotados, las condiciones para que funcione el enfoque APD deben tomar una forma mucho más simple que las usadas en la literatura actual. De hecho, en los Teoremas 6.3.1 y 6.3.5 evidencian fuertemente que esto es factible.

B. Funciones de Lyapunov y Algoritmos de Aproximación.

Los esquemas de aproximación más importantes para el problema en costo promedio son el Algortimo de Iteración de Valores (AIV) y el Algoritmo de Iteración de Políticias (AIP). Hasta donde conocemos, para procesos de Markov controlados en espacios de Borel y costos no acotados, la convergencia del AIV sólo se ha demostrado en [54], [29], y la convergencia del AIP sólo se ha obtenido en [35], [49].

B.1 El tipo de condiciones que se usan en [54], [29] son bastante diferentes, pero en ambos trabajos se apoyan fuertemente en el Teorema de Ascoli-Arzelá ([64], p. 169, Theorem 40) y, entonces, requieren la equicontinuidad de las funciones aproximantes la cual es una condición muy restrictiva y difícil de verificar.

En el Lema 5.3.5 y la Observación 5.3.6, bajo las condiciones H4.3.1, H4.3.2 y H4.3.8, se muestra que existe una fuerte relación entre la convergencia del AIV y el costo de oportunidad para las políticas estacionarias óptimas en costo promedio. Más precisamente, se muestra que el AIV converge si y sólo si para cada política $f \in \mathbf{F}$ óptima en costo promedio existe una constante k_f tal que

$$CO(f,x) = \lim_{n\to\infty}[J_n(f,x) - J_n^*(x)] = k_f \quad \forall x \in \mathbf{X}.$$

Esta equivalencia, interesante por si misma, sugiere un esquema alternativo al uso del Teorema de Ascoli-Arzelá para mostrar la convergencia del AIV.

B.2. En [35], se demuestra que el AIP converge bajo dos conjuntos de condiciones. En el primero de ellos, hacen uso del las condiciones H4.3.1, H4.3.2 y H4.3.8 y, en el segundo, suponen que la función de costo es estrictamente no-acotada (en ambos casos, bajo condiciones adicionales). Por otra parte, en [49] se obtiene la convergencia del AIP usando una condición de Lyapunov menos restrictiva que H4.3.2, pero supone que la función de costo es estrictamente no acotada.

En el Teorema 5.3.4 mostramos que las condiciones H4.3.1, H4.3.2 y H4.3.8 garantizan la existencia de una política Dutta óptima en la clase de las políticas estacionarias, y en el Teorema 5.3.7 mostramos que las afirmaciones "f es Dutta óptima", "f es óptima en sesgo", " f es óptima en costo de oportunidad" y "f es dominante", son equivalentes. Un problema interesante es proporcionar condiciones para que el AIV converja e identifique a una política Dutta óptima. Al parecer, esto sólo se ha demostrado para el caso de espacios finitos en [80].

C. Funciones de Lyapunov y Análisis por Trayectorias.

En [50], Theorem 17.01, p. 411, se muestra que la condición H4.3.2 garantiza que la Ley de Grandes Números, el Teorema del Límite Central y la Ley del Logaritmo Iterado se cumplen para los procesos controlados inducidos por las políticas estacionarias. Esto sugiere, por una parte, la posibilidad de obtener versiones por trayectorias de los Teoremas 4.5.2, 5.3.3 y 5.3.7 y, por otra, versiones "adecuadas" de los Teoremas del Límite Central y del Logaritmo Iterado para procesos controlados en espacios de Borel. Resultados de este tipo se han obtenido en [47] para el caso de espacios finitos y en [2] para espacios de Borel y costos acotados.

D. Optimalidad por Trayectorias para Procesos Semi-Markovianos.

El análisis realizado recientemente en [5] y [46] muestran que es pausible extender los resultados principales de los Capítulos 4, 5 y 6 a procesos semi-markovianos controlados en espacios de Borel y costos no-acotados.

Apéndice A

Notación

Para un espacio topológico $(X, \mathcal{T})$, usaremos la siguiente notación:

- $\mathcal{B}(X)$ es la $\sigma-$álgebra de Borel de X, es decir, la mínima $\sigma-$álgebra que contiene a $\mathcal{T}$, y a sus elementos les llamaremos conjuntos de Borel.
- X es un *espacio de Borel* si es un conjunto de Borel de un espacio métrico separable y completo.
- $M(X)$ es la clase de las funciones (Borel-) medibles $u : X \to \mathbf{R}$.
- $M_+(X) = \{u \in M(X) : u \geq 0\}$.
- $L(X)$ es la clase de las funciones $u : X \to \mathbf{R}$ semicontinuas inferiormente y acotadas por abajo.
- $C_a(X)$ es el espacio vectorial formado por las funciones $u : X \to \mathbf{R}$ continuas acotadas dotado con la *norma del supremo*, es decir, para cada $u \in C_a(X)$, se define $||u|| := \sup_x |u(x)|$.
- $M_a(X)$ denota la clase de las funciones $u : X \to \mathbf{R}$, medibles y acotadas.
- Si d es una métrica en X consistente con la topología $\mathcal{T}$, entonces $\mathcal{U}_d(X)$ denotará espacio de las funciones uniformemente continuas con respecto a la métrica d y la topología relativa del espacio de Banach $C_a(X)$.

Apéndice B

Multifunciones y Teoremas de Selección

Definición A.B.1. Sean X y A espacios de Borel.

(a) Una *multifunción* Ψ de X a A, es una función tal que $\Psi(x)$ es un subconjunto no vacio de A para cada $x \in X$. La *gráfica* de Ψ es el subconjunto de $X \times A$ definido por

$$Gr(\Psi) := \{(x, a) : x \in X, a \in \Psi(x)\}.$$

(b) Ψ es *Borel-medible* si $Gr(\Psi)$ es un conjunto de Borel de $X \times A$.

(c) Ψ es semicontinua superiormente si $\{x \in X : \Psi(x) \subset G\}$ es un conjunto abierto en X para cada subconjunto abierto G de A.

Definición A.B.2. Sea Ψ una multifunción Borel medible de X a A. Una función medible $f : X \to A$ es un *selector medible* de Ψ, si $f(x) \in \Psi(x)$ para cada $x \in X$. La familia de los selectores medible se denotara por $\mathbf{F}$.

Sea Ψ una multifunción Borel medible. Para cada función medible $v : Gr(\Psi) \to \mathbf{R}$ se define

$$v^*(x) := \inf_{a \in \Psi(x)} v(x, a).$$

Teorema A.B.3.(*Teorema de Selección*) Suponga que Ψ es una multifunción Borel medible de X a A tal que $\Psi(x)$ es un subconjunto compacto para cada $x \in X$. Si $v(x, \cdot)$ es una función semicontinua inferiormente en $Gr(\Psi)$ para cada $x \in X$, entonces existe un selector medible $f^* \in \mathbf{F}$ tal que

$$v(x, f^*(x)) = v^*(x) = \min_{a \in \Psi(x)} v(x, a) \quad \forall x \in X.$$

y v^* es medible.

Definición A.B.4. Sea Ψ una multifunción Borel medible de X a A. Una función v es *compacta inferiormente* en $Gr(\Psi)$ si $\{a \in \Psi(x) : v(x, a) \leq r\}$ es un suconjunto compacto, para cada $x \in X$ y $r \in \mathbf{R}$.

Teorema A.B.5.(*Teorema de Selección*) Sea Ψ una multifunción Borel medible de X a A. Si v es compacta inferiormente en $Gr(\Psi)$ y acotada por abajo, entonces existe un selector medible $f^* \in \mathbf{F}$ tal que

$$v(x, f^*(x)) = v^*(x) = \min_{a \in \Psi(x)} v(x, a) \quad \forall x \in X.$$

y v^* es medible.

Demostración. Para la demostración de los Teoremas A.B.3 y A.B.5, ver [34] (Propositions D.5 y D.6, p. 182-183).

Apéndice C

Convergencia de Medidas de Probabilidad

Sea X un espacio métrico. Denotamos por $\mathbf{P}(X)$ la familia de medidas de probabilidad en X.

Definición A.C.1. Sean μ_n y μ medidas en $\mathbf{P}(X)$. Se dice que μ_n *converge débilmente* a μ ($\mu_n \xrightarrow{w} \mu$) si

$$\lim_{n\to\infty} \int_X u(y)\mu_n(dy) = \int_X u(y)\mu(dy),$$

para cada $u \in C_a(X)$.

Proposición A.C.2. Si $\mu_n \xrightarrow{w} \mu$ y $u \in L(X)$, entonces

$$\lim_{n\to\infty} \int_X u(y)\mu_n(dy) \geq \int_X u(y)\mu(dy).$$

Definición A.C.3. Sea $\mathcal{P}$ una familia de medidas en $\mathbf{P}(X)$.

(a) La familia $\mathcal{P}$ es *tensa* si para cada $\varepsilon > 0$ existe un subconjunto compacto K de X tal que para cada $\mu \in \mathcal{P} : \mu(K) > 1 - \varepsilon$.

(b) La familia $\mathcal{P}$ es *secuencialmente pre-compacta* si para cada subsucesión $\{\mu_n\} \subset \mathcal{P}$ existe una subsucesión $\{\mu_{n_i}\}$ y una medida $\mu \in \mathbf{P}(X)$, no necesariamente en $\mathcal{P}$, tal que $\mu_{n_i} \xrightarrow{w} \mu$.

Teorema A.C.4.(*Teorema de Prohorov*). Sea $\mathcal{P} \subset \mathbf{P}(X)$.

(a) Si $\mathcal{P}$ es tensa, entonces es secuencialmente pre-compacta.

(b) Suponga que X es separable y completo. Si $\mathcal{P}$ es secuencialmente compacta, entonces es tensa.

Demostración. Ver [8] (Theorems 6.1, 6.2, p. 37).

Definición A.C.5. Una función $v \in M_+(X)$ es un *momento* en X si existe una sucesión no decreciente de compactos $X_n \uparrow X$ tal que

$$\lim_{n\to\infty} \inf\{v(x) : x \notin X_n\} = +\infty.$$

Proposición A.C.6. Sea $\mathcal{P} \subset \mathbf{P}(X)$. Si existe un momento v en X tal que

$$\sup\left\{\int_X v(y)\mu(dy) : \mu \in \mathcal{P}\right\} < +\infty,$$

entonces $\mathcal{P}$ es tensa.

Demostración. El resultado es consecuencia inmediata de las definiciones.

Apéndice D

Procesos de Markov

Consideraremos un proceso de Markov en tiempo discreto $\{x_t\}$ con valores en un espacio de Borel X y probabilidad de transición $P(B|x)$, $x \in \mathbf{X}$ y $B \in \mathcal{B}(X)$. Las probabilidades de transición en $n-$pasos se definen recursivamente en la forma siguiente. Para cada $x \in X$ y $B \in \mathcal{B}(X)$:

$$P^{n+1}(B|x) \quad : \quad = \int_X P^n(B|y)P(dy|x), \quad n \in \mathbf{N},$$

$$P^0(B|x) \quad : \quad = \mathbf{I}_B(x).$$

Por el Teorema de Ionescu-Tulcea ([1], Theorem 2.7.2, p. 109), para cada medida de probabilidad ν en $(X, \mathcal{B}(X))$ existe una medida de probabilidad P_ν definida en el espacio medible $(\Omega, \mathcal{F})$, donde $\Omega := X^\infty$ y $\mathcal{F}$ es la $\sigma-$álgebra producto correspondiente, que satisface las siguientes propiedades: para cada $B \in \mathcal{B}(X)$, $x \in X$, $t, k \in \mathbf{N}_0$,

- $P_\nu[x_0 \in B] = \nu(B)$, para cada $B \in \mathcal{B}(X)$;
- $P_\nu[x_{t+k} \in B|x_0, x_1, \cdots, x_{k-1}, x_k = x] = P^t(B|x)$.

E_ν denotará la operación de esperanza con respecto a la medida de probabilidad P_ν. Si $\nu(B) = \mathbf{I}_B(x)$, $B \in \mathcal{B}(X)$, para $x \in X$ fijo, se escribirá P_x y E_x en lugar de E_ν y P_ν, respectivamente.

Definición A.D.1. Sea φ una medida $\sigma-$finita en $(X, \mathcal{B}(X))$. El proceso de Markov $\{x_t\}$ es *$\varphi-$irreducible* si para cada $x \in X$ y $B \in \mathcal{B}(X)$, con $\varphi(B) > 0$, existe un natural n tal que $P^n(B|x) > 0$. En este caso, también se dice que la medida φ es una medida de irreducibilidad para el proceso $\{x_t\}$.

Observación A.D.2. Es fácil verificar que si $\{x_t\}$ es $\varphi-$irreducible y $\varphi' \ll \varphi$, entonces también es $\varphi'-$irreducible.

Definición A.D.3. Una medida $\sigma-$finita ψ en $(X, \mathcal{B}(X))$ es una *medida máxima de irreducibilidad* para el proceso $\{x_t\}$ si cualquier otra medida de irreducibilidad φ es *absolutamente continua con respecto a* ψ, es decir, $\varphi \ll \psi$.

Teorema A.D.4. Suponga que el proceso $\{x_t\}$ es $\varphi-$irreducible. Entonces, existe una medida máxima de irreducibilidad ψ.

Demostración. Ver [50] (Proposition 4.2.2, p. 88; [55], Proposition 2.4, p. 13).

Cuando el proceso $\{x_t\}$ sea $\varphi-$irreducible, usaremos la siguiente notación:

$$\begin{aligned} \mathcal{B}^+(X) \quad &: \quad = \{B \in \mathcal{B}(X) : \psi(B) > 0\}, \\ \mathcal{M}^+(X) \quad &: \quad = \{u \in M_+(X) : \int_X u(y)\psi(dy) > 0\}, \end{aligned}$$

donde ψ es una medida máxima de irreducibilidad.

Teorema A.D.5. Suponga que el proceso $\{x_t\}$ es $\varphi-$irreducible. Entonces existe $m \in \mathbf{N}$, $S_m \in \mathcal{M}^+(X)$ y una medida no negativa ν_m en $(X, \mathcal{B}(X))$ tales que

$$P^m(\cdot|x) \geq S_m(x)\nu_m(\cdot) \quad \forall x \in X.$$

Demostración. Ver [55] (Theorem 2.1, p. 16).

Definición A.D.6. Suponga que el proceso $\{x_t\}$ es $\varphi-$irreducible y defina

$$d := m.c.d.\{m \in \mathbf{N} : P^m(\cdot|x) \geq S_m(x)\nu_m(\cdot) \ \ \forall x \in X\}.$$

A d se le llama el *periodo* del proceso $\{x_t\}$. Si $d = 1$ se dice que el proceso es *aperiódico*; en caso contrario se dice que el proceso es *periódico*.

Definición A.D.7. El proceso $\{x_t\}$ es *Harris recurrente* si es φ−irreducible y

$$P_x\left[\bigcap_{N=1}^{\infty}\bigcup_{k=N}^{\infty}\{x_k \in B\}\right] = 1 \quad \forall x \in X, B \in \mathcal{B}^+(X).$$

Definición A.D.8. Una medida σ−finita μ en $(X, \mathcal{B}(X))$ es una *medida invariante* para el proceso $\{x_t\}$ si

$$\mu(B) = \int_X P(B|x)\mu(dx), \quad \forall B \in \mathcal{B}(X).$$

Observación A.D.9. Suponga que $\{x_t\}$ es φ−irreducible y μ es una medida invariante. Entonces,

(a) $\varphi \ll \mu$;

(b) μ es una medida máxima de irreducibilidad.

Note que (b) se sigue de (a), considerando ψ en lugar de φ. Para demostrar (a), defina

$$K_{\frac{1}{2}}(\cdot|x) := \sum_{n=0}^{\infty}(\frac{1}{2})^n P^n(\cdot|x), \quad x \in X,$$

y note que

$$\mu(B) = \int_X K_{\frac{1}{2}}(B|x)\mu(dx), \quad \forall B \in \mathcal{B}(X).$$

Además, $\varphi(B) > 0$ implica que $K_{\frac{1}{2}}(B|x) > 0\ \forall x \in X$; entonces, $\mu(B) > 0$. Por lo tanto $\varphi \ll \mu$.

Teorema A.D.10. Si $\{x_t\}$ es Harris recurrente, entonces existe una medida invariante μ, la cual es única excepto por constantes multiplicativas.

Demostración. Ver [50] (Theorem 10.0.1, p. 230).

Definición A.D.11. El proceso $\{x_t\}$ es *Harris (recurrente) positivo* si es Harris recurrente y admite una medida invariante finita μ. Note que si $\mu(X) = 1$, entonces μ es la única medida (de probabilidad) invariante.

Definición A.D.12. (*Ley de Grandes Números para Procesos de Markov*). Suponga que el proceso $\{x_t\}$ es *Harris positivo* con medida de probabilidad invariante μ. Entonces

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} u(x_t) = \int_{\mathbf{X}} u(x)\mu(dx) \quad P_x - \text{casi seguramente},$$

para cada $x_0 = x$ y $u \in M(X)$ tal que $\int_X |u(y)|\mu(dy) < +\infty$.

Demostración. Ver [50] (Theorem 17.0.1, p. 411).

Para cada distribución de probabilidad $a = \{a_n\}$ en $\mathbf{N}_0$, se define

$$K_a(\cdot|x) := \sum_{n=0}^{\infty} a_n P^n(\cdot|x), \quad x \in X.$$

Definición A.D.13. Una probabilidad de transición $T(B|x)$, $x \in X$ y $B \in \mathcal{B}(X)$, es una componente continua de $K_a(\cdot|\cdot)$ si

(i) $K_a(\cdot|x) \geq T(\cdot|x) \;\; \forall x \in X$;

(ii) $T(B|\cdot)$ es semicontinua inferiormente para cada $B \in \mathcal{B}(X)$.

Definición A.D.14. El proceso $\{x_t\}$ es un *T-proceso* si existe una distribución $a = \{a_n\}$ tal que

(i) $K_a(\cdot|\cdot)$ admite una componente continua T;

(ii) $T(X|x) > 0$ para cada $x \in X$.

Definición A.D.15. Sea $a = \{a_n\}$ una distribución de probabilidad y ν una medida en $(X, \mathcal{B}(X))$ tal que $\nu(X) \leq 1$. Un conjunto $B \in \mathcal{B}(X)$ es $(\nu-)$ *petite* si

$$K_a(\cdot|x) \geq \nu(\cdot) \quad \forall x \in B.$$

Teorema A.D.16(a). Si cada subconjunto compacto de X es petite, entonces $\{x_t\}$ es un T-proceso;

(b) Si $\{x_t\}$ es un T-proceso $\varphi-$irreducible, entonces cada subconjunto compacto de X es petite.

Demostración. Ver [50] (Theorem 6.2.5, p. 134).

Teorema A.D.17. Si para cada $x \in X$ existe una distribución de probabilidad $a_x = \{a_n^x\}$ tal que $K_{a_x}(\cdot|\cdot)$ admite un componente continua $T_x(\cdot|\cdot)$ tal que $T_x(X|x) > 0$ para cada $x \in X$, entonces $\{x_t\}$ es un T-proceso.

Demostración. Ver [50] (Proposition 6.2.4, p. 134).

Apéndice E

Miscelánea

Aproximación de Funciones Continuas

Teorema A.E.1. Sea $(X, \mathcal{T})$ un espacio topológico separable. Suponga que existe una métrica d en X consistente con la topología $\mathcal{T}$. Entonces existe una métrica d^* en X (consistente con $\mathcal{T}$) tal que:

(a) el subespacio $\mathcal{U}_{d^*}(Y)$ es separable;

(b) para cada $v \in \mathcal{C}_b(Y)$, existen sucesiones $\{u_n\}$ y $\{u'_n\}$ en $\mathcal{U}_{d^*}(Y)$ tal que $u_n \uparrow v$ y $u'_n \downarrow v$.

Demostración. Ver [7] (Proposition 7.2, Corollary 7.6.1, Proposition 7.9, Lemma 7.7, p. 106, 113, 116, 125).

Ley de Grandes Números para Martingalas

Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad y $\{\mathcal{F}_t : t \in \mathbf{N}_0\}$ una *filtración* en $\mathcal{F}$, es decir, una sucesión creciente de sub $\sigma-$álgebras de $\mathcal{F}$.

Definición A.E.2. Sea $\{M_t\}$ una sucesión de variables aleatorias definidas en $(\Omega, \mathcal{F}, P)$.

(a) Se dice que $\{M_t\}$ es una sucesión *adaptada* a la filtración $\mathcal{F} = \{\mathcal{F}_t\}$, si M_t es $\mathcal{F}_t-$medible para cada $t \in \mathbf{N}_0$.

(b) $\{M_t\}$ es una $\mathcal{F}-$*martingala* si es adaptada y $E[M_{t+1}|\mathcal{F}_t] = M_t$, para cada $t \in \mathbf{N}_0$.

Teorema A.E.3. (*Ley de Grandes Números para Martingalas*). Sea $\{M_t\}$ una martingala tal que

(i) $E|M_t|^{2r} < +\infty$ para cada $t \in \mathbf{N}_0$, para algún $r \geq 1$;

(ii) $\sum_{t=0}^{\infty} \frac{1}{t^{1+r}} E|M_{t+1} - M_t|^{2r} < +\infty$.

Entonces,

$$\lim_{t \to \infty} \frac{1}{t} M_t = 0 \quad P - \text{casi seguramente.}$$

Demostración. Ver [71] (Corollary 2, p. 471).

Descomposición de Medidas

Teorema A.E.4. Sean X y A espacios de Borel y $\gamma \in \mathbf{P}(X \times A)$. Entonces,

$$\gamma(B_1 \times B_2) = \int_{B_1} \varphi(B_2|x)\mu(dx), \quad \forall B_1 \in X, B_2 \in A,$$

donde $\varphi(\cdot|\cdot)$ es una probabilidad de transición de X a A y $\mu \in \mathbf{P}(X)$.

Demostración. Ver [7] (Corollary 7.27.2, p. 139).

Teorema Abeliano

Teorema A.E.5. Sea $\{c_t\}$ una sucesión de números no-negativos. Entonces

(a) $\liminf_{n \to \infty} \frac{1}{n} \sum_{t=0}^{n-1} c_t \leq \liminf_{\alpha \to 1^-} (1 - \alpha) \sum_{t=0}^{\infty} \alpha^t c_t$;

(b) $\limsup_{\alpha \to 1^-} (1 - \alpha) \sum_{t=0}^{\infty} \alpha^t c_t \leq \limsup_{n \to \infty} \frac{1}{n} \sum_{t=0}^{n-1} c_t$;

(c) $\lim_{\alpha \to 1^-} (1 - \alpha) \sum_{t=0}^{\infty} \alpha^t c_t$ existe si sólo si $\lim_{n \to \infty} \frac{1}{n} \sum_{t=0}^{n-1} c_t$ existe.

Demostración. Ver [72].

Bibliografía

[1] Ash,R.B. (1972), *Real Analysis and Probability*, Academic Press, New York.

[2] Asriev, A.V., Rotar, V.I (1990), *On asymtotic optimality in probability and almost surely in dynamic control*, Stochastics and Stochastics Reports, 33, 1-16.

[3] Arapostathis A., Borkar, V.S., Férnandez-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, 282-344.

[4] Balder, E.J. (1989), *On compactness of the space policies in stochastic dynamic programming*, Stoch. Process. and their Appl. 1, 141-150.

[5] Bhatnagar, S., Borkar, V.S. (1995), *A convex analitic framework for ergodic control semi-Markov processes*, Mathematics of Operation Research 20, 923-936.

[6] Bertsekas, D.P. (1987), *Dynamic Programming: Deterministic and Stochastic Models*, Prentice-Hall, Englewood Cliffs, N.J.

[7] Bertsekas, D.P. and Shreve, S.E. (1978), *Stochastic Optimal Control: The Discrete Time Case*, Academic Press, N.Y.

[8] Billingsley, P. (1968), *Convergence of Probability Measures*, Wiley, N.Y.

[9] Blackwell, D. (1962), *Discrete dynamic programming*, Ann. Math. Statist. 33, 719-726.

[10] Borkar, V.S. (1991), *Topics in Controlled Markov Chain*, Pitman Research Notes in Mathematics Series 240, Longman Scientific & Technical UK.

[11] Brown, B.W. (1965), *On the iterative method of dynamic programming on a finite space discrete time Markov processes,* Ann. Math. Statist. 33, 719-726.

[12] Cavazos-Cadena, R. (1988), *Necessary and sufficient conditions for a bounded solution to the optimality equation in average reward Markov decision chains*, Systems Control Letters, 10, 71-78.

[13] Cavazos-Cadena, R. (1991), *A counterexample on the optimality equation in a class of average Markov decision chains with average cost criterion*, Syst. control Lett. 16, 23-27.

[14] Cavazos-Cadena, R. (1993), *A note in vanishing interest rate approach in average Markov decision chains with continuos and bounded cost*, Syst. Control Lett. 24, 373-383.

[15] Cavazos-Cadena, R. and Fernández-Gaucherand, E. (1995), *Denumerable controlled Markov chains with average reward criterion: sample path optimality*, ZOR–Mathematical Methods of Operation Research 41, 89-108.

[16] Cavazos-Cadena, R. and Fernández-Gaucherand, E. (1996), *Denumerable Markov control chains with strong average optimality criterion: bounded and unbounded cost*, ZOR–Math. Methods of Oper. Research 43, 261-300.

[17] Cavazos-Cadena, R., Hernández-Lerma, O. (1992), *Equivalence of Lyapunov stability criteria in a classs of Markov decision processes*, J. App. Math. and Optimization 26, 113-137.

[18] Cavazos-Cadena, R. and Sennott, L.I. (1992), *Comparing recent assumptions for the existence for average optimal stationary policies*, Oper. Research Lett. 11, 33-37.

[19] Denardo, E.V., Rothblum, U.G. (1979), *Overtaking optimality for Markov decision chains*, Math. Oper. Res. 4, 144-152.

[20] Dynkin, E.B. and Yuskevich, A.A. (1979), *Controlled Markov Processes*, Springer-Verlag, N.Y.

[21] Dutta, P.K. (1990), *What do discount optima converge?, A theory of discount rate asymptotics in economic models*, J. Economic Theory, 55-94.

[22] Ephremides, A. and Verdú, S. (1989), *Control and optimization methods in communication networks*, IEEE Trans. Autom. Control, 34, 930-942.

[23] Foster, F.G. (1953), *On the stochastic matrices associated with certain queueing processes*, Ann. Math. Statist. 24, 355-360.

[24] Flynn, J. (1976), *Conditions for the equivalence of optimality criteria in dynamic programming*, Ann. Statist., 4, 936-953.

[25] Flynn, J. (1980), *On optimality criteria for dynamic programs with long finite horizons*, J. Math. Anal. Appl., 76, 202-208.

[26] Gale, D. (1965), *On optimal development in a multi-sector economy*, Review of Econom. Studies, 34, 1-19.

[27] Glynn, P.W., Meyn, S.P. (1996), *A Lyapunov bound for solution of Poisson's equation*, Ann. Prob. 24, 916-931.

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

[29] Gordienko, E., Hernández-Lerma, O. (1995), *Average cost Markov control processes with weghted norms: value iteration*, Appl. Math. 23, 219-237.

[30] Hernández-Lerma, O. (1989), *Adaptive Markov Control Processes*, Springer-Verlag, N.Y.

[31] Hernández-Lerma, O. (1991), *Average optimality in dynamic programming on Borel spaces—Unbounded cost and controls*, Syst. Control Lett. 17, 237-242.

[32] Hernández-Lerma, O. (1993), *Existence of average optimal policies in Markov control processes with strictly unbounded costs*, Kybernetika 29, 1-17.

[33] Hernández-Lerma, O. and Lasserre, J.B. (1990), *Average cost optimal policies for Markov control processes with Borel state space and unbounded costs*, Syst. control Lett. 15, 349-356.

[34] Hernández-Lerma, O. and Lasserre, J.B. (1996), *Discrete-Time Markov Control Preccesses: Basic Optimality Criteria*, Springer-Verlag, N.Y.

[35] Hernández-Lerma, O. and Lasserre, J.B. (1997), *Policy iteration for average cost Markov control processes on Borel spaces*, Acta Appl. Math. 47, 125-154.

[36] Hernández-Lerma, O., Montes-de-Oca, R., Cavazos-Cadena, R. (1991), *Recurrence conditions for Markov decision processes with Borel state space: a survey*, Ann. Oper. Res., 28, 29-46.

[37] Hernández-Lerma, O., Muñoz-de-Osak, M. (1992), *Discrete.time Markov control processes with discounted unbounded cost: optimality criteria*, Kybernetika 28, 191-212.

[38] Hernández-Lerma, O., Vega-Amaya, O. (1996), *Infinite-horizon Markov control processes with undiscounted cost criteria: From average to overtaking optimality*, to appear in Appl. Math.

[39] Hinderer, K. (1970), *Foundations of Non-Stationary Dynamic Programming with discrete Time Parameter*, Lecture Notes Oper. Res. 33, springer-Verlag.

[40] Hordijk, A. (1977), *Dynamic Programming and Potential Theory*, 2nd ed., Mathematical Centre Tracts 51, Mathematisch Centrum, Amsterdam.

[41] Hu, Q.Y. (1992), *Discounted and average Markov decision processes with unbounded rewards: New conditions*, J. Math.Anal. and Appl. 171, 111-124.

[42] Kurano, M. (1989), *The existence of a minimium pair of state and policy for Markov decision processes under the hypothesis of Doeblin*, SIAM J. Control and Optim. 27, 296-307.

[43] Leizarowitz, A. (1987), *Infinite horizon optimization for finite state Markov chains*, SIAM J. control Optim. 25, 1601-1618.

[44] Leizarowitz, A. (1996) *Overtaking and almost-sure optimality for infinite horizon Markov decision processes*, ZOR–Mathematical Methods of Operation Research 21, 158-181.

[45] Luque-Vásquez, F. and Hernández-Lerma, O. (1995), *A counterexample on the semicontinuity of minima,* Proc. Amer. Math. Soc. 123, 3175-3176.

[46] Luque-Vásquez, F. (1997), *Modelos de Control Semi-Markoviano en Espacios de Borel*, Tesis de Doctorado en Ciencias (Matemáticas), Facultad de Ciencias, UNAM.

[47] Mandl, P., Lausmanová, M. (1991), *Two extensions of asymptotic methods in controlled Markov chains,* Ann. Oper. Res. 28, 67-80.

[48] Meyn, S.P. (1989), *Ergodics theorems for discrete time stochastic systems using a stochastic Lyapunov function*, SIAM J. Control and Optim. 27, 1409-1439.

[49] Meyn, S.P. (1995), *The policy improvement algorithm for Markov decision processes with general state space,* Preprint Coordinated Science Laboratory, University of Illinois, Urbana, Il.

[50] Meyn, S.P., Tweedie, R.L. (1993), *Markov Chains and Stochastic Stability*, Springer-Verlag, London.

[51] Montes-de-Oca, R. (1994), *The average cost optimality equation for Markov control processes on Borel spaces*, Syst. Control Lett. 22, 351-357.

[52] Montes-de-Oca, R.; Hernández-Lerma, O. (1996), *Value iteration in average cost Markov control processes on Borel spaces,* Acta Appl. Math. 46, 203-221.

[53] Montes-de-Oca, R.; Hernández-Lerma, O. (1994), *Conditions for average optimality in Markov control processes with unbounded costs and controls,* J. Math. Syst., Estimation and Control 4, 1-19.

[54] Montes-de-Oca, R., Minjarez-Sosa, J.A. and Hernández-Lerma, O. (1994), *Conditions for average optimality in Markov control processes*, Bol. Soc. Mat. Mexicana 39, 39-50.

[55] Nummelin, E. (1983), *General Irreducible Markov Chains and Non-Negative Operators*, Cambridge University Press, Cambridge.

[56] Nowak, A.S. (1992), *Stationary overtaking optimal strategies in Markov decision processes with general staate space*, Preprint, Institute of Mathematics, Technical University of Wroclaw, Poland.

[57] Nowak, A.S., Vega-Amaya, O. (1997), *A counterexample on overtaking optimality*, Reporte preliminar.

[58] Parlar, M., Rempala, R. (1992), *Stochastic inventory problem with piecewise quadratic holding cost function containing a cost-free interval*, Journal of Optimization Theory and Applications, 75, 133-153.

[59] Puterman, M.L. (1994), *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, Wiley, N.Y.

[60] Ramsey, F.P. (1928), *A mathematical theory of savings*, Economic J. 38, 543-559.

[61] Rieder, U. (1976), *On optimal policies and martingales in dynamic programming*, J. Appl. Probab. 13, 507-518.

[62] Ross, S.M. (1968), *Non-discounted denumerable Markovian decision models*, Ann. Math. Statist. 39, 412-423.

[63] Ross, S.M. (1983), *Introduction to Solchastic Dynamic Programming*, Academic Press, N.Y..

[64] Royden, H.L. (1988), *Real Analysis*, MacMillan, New York.

[65] Schäl, M. (1975), *Conditions for optimality and for the limit of n-stage optimal policies to be optimal*, Z. Whars. verw. Gerb.,32, 179-196.

[66] Schäl, M. (1993), *Average optimality in dynamic programming with general state space*, Math. Oper. Res. 18, 163-172.

[67] Schäl, M., Sudderth, W. (1987), *Stationary policies and Markov policies in Borel dynamic programming*, Probab. Theory and Rel. Fields 74, 91-111.

[68] Sennott, L.I. (1989), Average cost optimal stationary policies in infinite markov decision processes with unbounded cost, Res. 37, 626-633.

[69] Sennott, L:I. (1989), *Average cost semi-Markov decision processes and the control of systems*, Prob. Eng. Inform. Sci. 3, 247-272.

[70] Sennott, L.I. (1995), *Another set of conditions for average optimality in Markov control processes*, Syst. control Lett. 24, 147-151.

[71] Shiryayev, A.N. (1984), *Probability*, Springer-Verlag, N.Y.

[72] Sznadjer, R. and Filar, J.A: (1992), *Some comments on a theorem of Hardy and Littlewood*, J. Optim. Theory Appl. 75, 201-208.

[73] Stidham, S. and Weber, R. (1993), *A survey of Markov decision models for control of networks of queuees*, Queueing Syst., 13, 206-217.

[74] Tapiero, C. S. (1994), *Applicable stochastic control: From theory to practice*, European Journal of Operational Reseach **73**, 209-225.

[75] Taylor, H.M. (1965), *Markovian sequential replacement processes*, Ann. Math. Statist. 38, 1677-1694.

[76] Vega-Amaya, O. (1993), *Average optimality in semi-Markov control models on Borel spaces: Unbounded costs and controls*, Bol. Soc. Mat. Mex. 38, 47-60; Erratum (1994) 39, 68.

[77] Vega-Amaya, O. (1997), *Sample path optimality of Markov control processes with strictly unbounded cost*, sometido para su publicación en Applicationes Mathematicae.

[78] Vega-Amaya, O., Luque-Vásquez, F. (1994), *Modelos de control semi-Markoviano*, Aportaciones Matemáticas, Soc. Mat. Mex. 11, 171-183.

[79] Vega-Amaya, O. and Montes-de-Oca, R. (1996), *Application of average dynamic programming to inventory systems*, Math.Methods Oper. Res., to appear.

[80] Veinott, A.F. Jr. (1966), On finding optimal policies in discrete dynamic programming with no discounting, Ann. Math. Statist., 37, 1284-1294.

[81] von Weizsacker, C.C. (1965), *Existence of optimal programs of accumulation for infinite horizon*, Review of Econom. Studies, 32, 85-104.

[82] Yushkevich, A.A. (1973), *On a class of strategies in general Markov decision models*, Theory Probab. Appl., 18, 777-779.