UNIVERSIDAD DE SONORA

División de Ciencias Exactas y Naturales

Departamento de Matemáticas

Optimalidad Asintótica en Procesos de Control de Markov

T E S I S

Que para obtener el título de:

Licenciado en Matemáticas

Presenta:

Estephania Pivac Alcaraz

Director de tesis: Dr. Jesús Adolfo Minjárez Sosa

Hermosillo, Sonora, México Diciembre de 2015

SINODALES

Dr. Eugene Gordienko
Departamento de Matemáticas,
Universidad de Sonora.

M.C. Carmen Geraldi Higuera Chan
Departamento de Matemáticas,
Universidad de Sonora.

Dr. Jesús Adolfo Minjárez Sosa
Departamento de Matemáticas,
Universidad de Sonora.

Dr. Oscar Vega Amaya
Departamento de Matemáticas,
Universidad de Sonora.

"Que Dios reparta suerte y va por ustedes."

Con inmenso amor,
a mi mamá.

Agradecimientos

Quiero comenzar agradeciendo a Dios por darme la oportunidad de lograr esta meta y estar un paso mas cerca de lograr mi sueño, por todo el conjunto de accidentes que a lo largo de mi vida me llevaron a tomar las decisiones que he tomado y a ser lo que soy, y principalmente quiero agradecerle por la familia que me dió, porque mi familia es mi inspiración y mi fuerza para seguir adelante.

Quiero agradecer a mi familia por todo su amor, por apoyarme incondicionalmente, por animarme, por estar en las buenas y en las malas, y por nunca dudar de mí. En especial le doy gracias a mi mamá que hizo hasta lo imposible por mandarme a estudiar, por ser la única persona que desde el inicio me apoyó en mi decisión de estudiar matemáticas, y simple y sencillamente por ser la mejor mamá que Dios pudo darme. A mis hermanas: a Leonelita por todo su amor y su ternura, a Camile por su apoyo y sus ocurrencias, y a Carolina por ser mi defensora y confidente. A mi papá por todo el apoyo que me brindó. A mi abuelita por su cariño y por siempre sentirse orgullosa de mí. A mi primo Giovanni por ser el hermano mayor que nunca tuve, por su ejemplo y sus consejos.

Le doy gracias a mi director de tesis, Dr. Adolfo Minjarez Sosa, por su tiempo, su esfuerzo y su paciencia. Muchas gracias por su apoyo, para mí fue todo un honor trabajar con usted. También agradezco a mis sinodales: Dr. Eugene Gordienko, M.C. Carmen Geraldi Higuera Chan y Dr. Oscar Vega Amaya, por su tiempo y su disposición a lo largo de la revisión del trabajo.

También quiero darle las gracias a todos mis profesores por su motivación y sus enseñanzas; a mis amigos y a mis compañeros de carrera, por ayudarme, por todas sus risas, su tiempo, su confianza y por hacer que disfrutara aún más esta etapa.

Agradezco a Carolina Martinez y a Carmen Higuera, por su compañía, su ayuda y su amistad, soy muy afortunada de tenerlas en mi vida. Por último le doy gracias a Pedro Morghen Lopez, por ser y estar, por motivarme y apoyarme, por sus consejos, por su cariño y por hacerme feliz.

Estephania Pivac Alcaraz

Hermosillo, Sonora. Diciembre 2015

Índice general

Introducción

La Teoría de Control Óptimo trata con problemas de optimización de sistemas que evolucionan en el tiempo. Este hecho es lo que la diferencia de los problemas de optimización comunes, y por lo tanto sus aplicaciones se presentan en problemas donde es necesario controlar sistemas dinámicos que aparecen, por ejemplo, en áreas como Economía, Finanzas, Biología e Ingeniería, y cuyo objetivo es determinar las acciones o decisiones de control que debe tomar un controlador para optimizar dicho sistema.

Las acciones durante la evolución del sistema se eligen por medio de reglas o sucesiones de funciones llamadas políticas de control, cuyo comportamiento o respuesta al sistema lo mide un funcional de costo al cual se le conoce como Índice de funcionamiento.

Al problema de encontrar una política que minimice un índice de funcionamiento se le llama Problema de Control Óptimo. Si el sistema involucra elementos aleatorios para describir la dinámica diremos que tenemos un Problema de Control Óptimo Estocástico el cual es el tema de esta tesis.

Una clase de sistemas de control estocástico la constituyen los Procesos de Control de Markov cuya teoría se basa en la técnica de Programación Dinámica Estocástica. El estudio de estos procesos se puede dividir de acuerdo al tipo de espacios de estados y de control: numerables o espacios generales; y según el criterio de optimalidad que define el índice de funcionamiento: costo total esperado, costo descontado y costo promedio por etapa.

En este trabajo nos centraremos en el estudio de los Procesos de Control de Markov con espacios numerables bajo el criterio de costo descontado acotado.

En términos generales, el estudio del problema de control óptimo bajo

el criterio de costo descontado consiste primero en mostrar que la función de costo óptimo, digamos V^*, satisface una ecuación de la forma

$$V^* = TV^*,$$

donde T es un operador, a la cual se le conoce como Ecuación de Optimalidad. Después, como segundo paso, se debe de resolver un problema de optimización para calcular la política óptima. La solución a ambos problemas no es trivial, y por lo tanto es necesario desarrollar métodos de aproximación tanto para V^* como para la política óptima. Estos métodos de aproximación consisten en definir apropiadamente una sucesión de funciones $\{u_n\}$ tal que $u_n \to V^*$, y más aún, en cada paso de la aproximación resolver un problema de optimización para encontrar una función f_n que determine el control en la etapa n:

$$a_n = f_n(\cdot).$$

El objetivo principal de este trabajo es proponer algunos métodos de aproximación de V^* y analizar la optimalidad de las políticas de control $\pi = \{f_n\}$ que producen dichos métodos.

Sin embargo, una característica que tiene el índice de costo descontado es que los controles seleccionados en las primeras etapas, digamos $n \leq m$ para algún $m \in \mathbb{N}$, son los que tienen mayor peso. Pero por otro lado, la información más confiable sobre V^* que proporcionan los algoritmos de aproximación ($u_n \to V^*$) se obtiene en las últimas etapas, $n > m$. Esta contradicción implica que en general este tipo de políticas no necesariamente son óptimas, y por lo tanto su optimalidad será analizada en un sentido asintótico, dando lugar a las llamadas políticas asintóticamente óptimas.

La optimalidad asintótica ha sido estudiada en otros contextos, como por ejemplo en problemas de control adaptado [ver [2], [4], [8], [9], [13]]. Regularmente en estos problemas se supone que la distribución de la variable aleatoria que define la dinámica del sistema es desconocida. Entonces se deben implementar métodos de estimación y control para definir las políticas. De igual forma, estos métodos de estimación proporcionan información más confiable sobre la distribución desconocida en las últimas etapas, lo cual implica que estas políticas no sean óptimas.

El trabajo está estructurado de la siguiente manera:

En el primer capítulo se presentan los elementos básicos necesarios para definir el problema de control óptimo (PCO).

En el segundo capítulo se establecen las condiciones bajo las cuales existen políticas óptimas para el PCO y se presentan algoritmos de aproximación.

En el tercer capítulo se estudia la optimalidad asintótica y la optimalidad en el límite.

Finalmente comentaremos sobre la bibliografía utilizada.
Para los elementos básicos se utilizaron [1], [4] y [12]; para los resultados de optimalidad asintótica y algoritmos de aproximación se usaron [5] y [6], y el resto de la bibliografía nos sirvió de consulta y apoyo para desarrollar el trabajo.

Capítulo 1

Procesos de control de Markov

En este capítulo presentamos los elementos necesarios para definir el problema de control óptimo. Con este fin, primero describiremos el modelo de control de Markov así como la familia de políticas de control. Presentaremos también los índices de funcionamiento más comunes que miden el comportamiento de las políticas de control. Finalmente presentamos un ejemplo de un sistema de producción/inventario.

1.1. Modelo de control markoviano

Definición 1.1.1 *Un modelo de control markoviano (MCM) en tiempo discreto, es un arreglo*

$$(\mathbb{X}, \mathbb{A}, \{A(x) : x \in \mathbb{X}\}, P, c), \tag{1.1}$$

que consta de los siguientes elementos:

- $\mathbb{X}$ *representa el espacio de estados, y supondremos que es un conjunto numerable.*
- $\mathbb{A}$ *es el espacio de controles o acciones, y supondremos que es un conjunto numerable.*
- $\{A(x) : x \in \mathbb{X}\}$ *es la familia de conjuntos de controles (o acciones) admisibles. Es decir, cada estado $x \in \mathbb{X}$ tiene asociado un conjunto no vacío $A(x) \subset \mathbb{A}$, cuyos elementos son los controles admisibles cuando*

el sistema se encuentra en el estado x.
Para cada $x \in \mathbb{X}$ y $a \in A(x)$ definimos

$$\mathbb{K} := \{(x, a) : x \in \mathbb{X}, a \in A(x)\} \tag{1.2}$$

al cual llamaremos el conjunto de pares estado-acción admisibles.

- *P representa la ley de transición entre los estados. Es decir,*

  $$P_{x,y}(a) := P[x_{t+1} = y | x_t = x, a_t = a].$$

  Además,

  (i) $P_{x,y}(a) \geq 0 \quad \forall\ x, y \in X,\ a \in A(x)$;

  (ii) $\sum_{y \in \mathbb{X}} P_{x,y}(a) = 1.$

- *$c : \mathbb{K} \to \mathbb{R}$ representa la función de costo por etapa.*

Un modelo de control Markoviano representa un sistema estocástico controlado que se observa en cada tiempo $t \in \mathbb{N}_0$ que llamaremos etapas de decisión. Denotando por $x_t = x \in \mathbb{X}$ y $a_t \in A(x)$, el estado del sistema y el control (o acción) aplicado al tiempo t, respectivamente, la evolución del sistema la podemos describir de la siguiente manera. Si el sistema se encuentra en el estado $x_t = x \in \mathbb{X}$ al tiempo t y se aplica el control $a_t = a \in A(x)$, entonces dos cosas ocurren:

1. Se produce un costo $c(x, a)$.
2. El sistema evoluciona al estado $x_{t+1} \in \mathbb{X}$ de acuerdo a la ley de transición P.

Una vez que el sistema se encuentra en el estado $x_{t+1} = x'$, se elige un nuevo control $a' \in A(x')$ y el proceso anterior se repite.

Si el número de etapas de decisión es finito, decimos que el MCM tiene horizonte de planeación finito y en otro caso decimos que el horizonte de planeación es infinito.

Observación 1.1.2
En muchas aplicaciones, la evolución del sistema está determinada por una ecuación en diferencias de la forma

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

donde $\{\xi_t\}$ *es una sucesión de variables aleatorias independientes e identicamente distribuidas, con valores en algún conjunto numerable* $\mathbb{S}$*, y* $F : \mathbb{X} \times \mathbb{A} \times \mathbb{S} \to \mathbb{X}$ *es una función conocida.*

Si ϕ *es la función de probabilidad común de las variables aleatorias* ξ_t*, es decir,*

$$\phi(k) = P[\xi_t = k] \quad \forall\ k \in \mathbb{S},\ t \in \mathbb{N}_0, \tag{1.4}$$

entonces para cada $(x, a) \in \mathbb{K}$ *tenemos*

$$P_{x,x'}(a) = P[x_{t+1} = x' | x_t = x, a_t = a] = \sum_{k \in S_{x'}} \phi(k),$$

donde

$$S_{x'} := \{s \in \mathbb{S} : F(x, a, s) = x'\}.$$

De lo anterior podemos concluir que si la evolución del sistema está representada por la ecuación en diferencias (1.3) *entonces la ley de transición entre los estados esta determinada por la función de probabilidad* ϕ*.*

1.2. Políticas de control admisibles

Definición 1.2.1 *Dado un MCM y* $t \in \mathbb{N}_0$*, definimos el espacio de historias admisibles hasta la etapa* t *mediante*

$$\mathbb{H}_0 := \mathbb{X} \quad y$$

$$\mathbb{H}_t := \mathbb{K}^t \times \mathbb{X} \quad para \ \ t \in \mathbb{N}.$$

Un elemento en $\mathbb{H}_t$ *es un vector o t-historia de la forma*

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

con $(x_k, a_k) \in \mathbb{K}$ *para* $k = 0, 1, ..., t-1$ *y* $x_t \in \mathbb{X}$*.*

Definimos el conjunto

$$\mathbb{F} := \{f : \mathbb{X} \to \mathbb{A} \mid f(x) \in A(x), x \in \mathbb{X}\}.$$

A cada elemento de $\mathbb{F}$ se le llama selector.

Definición 1.2.2

(a) *Una política de control es una sucesión* $\pi = \{f_t\}$ *de funciones* $f_t : \mathbb{H}_t \to \mathbb{A}$ *tal que* $f_t(h_t) \in A(x_t) \ \ \forall\ h_t \in \mathbb{H}_t,\ t \in \mathbb{N}_0$.

(b) *Una política de control Markoviana es una sucesión* $\pi = \{f_t\}$, *donde* $f_t \in \mathbb{F} \ \ \forall\ t \in \mathbb{N}_0$.

(c) *Una política Markoviana es estacionaria si existe* $f \in \mathbb{F}$ *tal que* $f_t = f \ \ \forall\ t \in \mathbb{N}_0$.

De acuerdo a la Definición 1.2.2, una política determina el control o decisión que se aplica en cada etapa; es decir, $a_t = f_t(h_t)$, $a_t = f_t(x_t)$ o $a_t = f(x_t)$, según sea el caso.

Denotaremos por Π al conjunto de todas las políticas e identificaremos el conjunto de políticas estacionarias con el conjunto $\mathbb{F}$.

En el caso particular de un MCM con horizonte de planeación finito N, una política es de la forma $\pi = \{f_0, f_1, ..., f_{N-1}\}$.

En un MCM con horizonte de planeación $N < \infty$, definimos el espacio muestral como

$$\Omega_N := \mathbb{K}^N \times \mathbb{X},$$

cuyos elementos son las **trayectorias** de la forma

$$\omega = (x_0, a_0, ..., x_{N-1}, a_{N-1}, x_N)$$

con $(x_k, a_k) \in \mathbb{K}$ si $k = 0, 1, ..., N-1$ y $x_N \in \mathbb{X}$; y en el caso en que $N = \infty$, el espacio muestral se define como $\Omega = \mathbb{K}^\infty$, y las **trayectorias** son de la forma $\omega = (x_0, a_0, ...)$, con $(x_k, a_k) \in \mathbb{K}$.

Para un estado inicial $x \in \mathbb{X}$ y una política $\pi = \{f_0, f_1, ...\} \in \Pi$, existe una probabilidad denotada por P_x^π definida en una familia de subconjuntos de Ω tal que las variables x_k y a_k satisfacen

$$P_x^\pi[x_0 = x] = 1,$$

$$P_x^\pi[x_{t+1} = y | h_t, a_t] = P_{x_t,y}(a_t) \tag{1.5}$$

En el caso de horizonte finito $N < \infty$, la probabilidad P_x^π se define explícitamente por

$$P_x^\pi(x_0, a_0, ..., x_{N-1}, a_{N-1}, x_N) = \delta_x(x_0) P_{x_0,x_1}(a_0) \cdots P_{x_{N-1}x_N}(a_{N-1})$$

donde $a_k = f_k(x_0, a_0, ..., x_k)$, $k = 0, 1, ..., N-1$ y $\delta_x(\cdot)$ representa la probabilidad concentrada en x.

Observaciones 1.2.3

(a) *Denotamos por E_x^π al operador esperanza respecto a P_x^π, es decir, si W es una variable aleatoria definida en Ω, su valor esperado está dado por*

$$E_x^\pi[W] = \sum_{\omega \in \Omega} W(\omega) P_x^\pi(\omega).$$

(b) *Si v es una función de x_{t+1} entonces*

$$E_x^\pi[v(x_{t+1})|h_t, a_t] = \sum_{y \in \mathbb{X}} v(y) P_{x_t, y}(a_t) \tag{1.6}$$

1.3. Índice de funcionamiento

Como se había mencionado antes, un índice de funcionamiento, también llamado criterio de optimalidad, es una función que de cierta manera "mide" el comportamiento del sistema al utilizar diferentes políticas de control, dado el estado incial.

A continuación se presentan tres índices de funcionamiento usuales cuando se utiliza la política π dado que el estado inicial es $x_0 = x$.

Definición 1.3.1 *Sean $x \in \mathbb{X}$ y $\pi \in \Pi$. Se define:*

(a) *El **costo total esperado** hasta la N-ésima etapa por*

$$J_N(\pi, x) := E_x^\pi[\sum_{t=0}^{N-1} c(x_t, a_t) + c_N(x_N)],$$

donde $c_N(x)$ es una función definida para cada $x \in \mathbb{X}$, y representa un "costo terminal".

(b) *El* ***costo total esperado*** α***-descontado*** *mediante*

$$V_\alpha(\pi, x) := E_x^\pi[\sum_{t=0}^{\infty} \alpha^t c(x_t, a_t)], \tag{1.7}$$

donde $\alpha \in (0, 1)$ *representa el factor de descuento.*

(c) *El* ***costo promedio esperado*** *por*

$$J(\pi, x) := \limsup_{N\to\infty} \frac{1}{N} E_x^\pi[\sum_{t=0}^{N-1} c(x_t, a_t)].$$

En este trabajo nos centraremos en el estudio de MCM con índice de costo descontado. La motivación para el estudio de este índice de funcionamiento proviene del análisis de problemas en las áreas de economía y finanzas donde V_α tiene una interpretación monetaria. En este caso, se incluye un factor de descuento α al costo, ya que cierta cantidad de dinero en el presente tiene menor valor en el futuro. De hecho, en muchos problemas, el factor de descuento α se interpreta como $\alpha = \frac{1}{1+i}$, donde i representa la tasa de interés. Entonces, α^t representa el valor presente de la moneda t períodos después, es decir: un costo de L unidades en el tiempo t equivale a un costo presente de $\alpha^t L$ unidades.

1.4. Problema de control óptimo

Con todos los elementos descritos anteriormente, podemos plantear el problema de control óptimo (PCO) de la siguiente manera:

Dado un modelo de control de Markov $(\mathbb{X}, \mathbb{A}, \{A(x) : x \in \mathbb{X}\}, P, c)$ una familia de políticas de control admisibles Π y uno de los índices de funcionamiento de la Definición 1.3.1 al cual representamos por $w(\pi, x)$, el PCO consiste en encontrar una política $\pi^* \in \Pi$ tal que

$$w(\pi^*, x) = \inf_{\pi\in\Pi} w(\pi, x), \quad \forall\ x \in \mathbb{X}. \tag{1.8}$$

Por ejemplo, para el caso descontado, el PCO es encontrar $\pi^* \in \Pi$ tal que

$$V_\alpha(\pi^*, x) = \inf_{\pi\in\Pi} V_\alpha(\pi, x) =: V^*(x), \quad x \in \mathbb{X}.$$

En este caso a π^* se le llama política α-óptima y $V^*(\cdot)$ es la función de valor óptimo.

1.5. Ejemplo: sistema de producción/inventario

Consideremos el siguiente sistema de producción/inventario.

Al inicio de cada período, se observa el nivel de inventario (cantidad) de determinado artículo que hay en un almacén o tienda, y en base a esta información se decide ordenar a la unidad de producción, una cantidad adicional de artículos o conservar el nivel de inventario. Suponemos que se satisfacen las siguientes condiciones:

1. El almacén tiene una capacidad finita de C unidades.
2. La solicitud de artículos adicionales se hace al inicio de cada período y se surte inmediatamente.
3. Los costos y precio de venta del artículo no varían en diferentes períodos.

Definamos las siguientes variables. Para cada $t \in \mathbb{N}_0$:

- x_t representa el nivel de inventario del artículo al inicio de la etapa t.
- a_t representa la cantidad de artículos solicitados a la unidad de producción, a fin de abastecer la unidad de inventario al inicio de la etapa t.
- ξ_t representa la demanda durante la etapa t, y suponemos que $\{\xi_t\}$ es una sucesión de variables aleatorias i. i. d. (con valores en $\mathbb{N}_0$), y función de probabilidad común ϕ.

De lo anterior observamos que:

- $\mathbb{X} = \mathbb{A} = \{0, 1, ..., C\}$.
- Dado que el sistema tiene capacidad finita C, si $x_t = x$, entonces solo tiene sentido solicitar a la unidad de producción una cantidad de artículos $a_t = a \in A(x) = \{0, 1, ..., C - x\}$. Así, cada $x \in \mathbb{X}$ tiene asociado un conjunto no vacío $A(x) \subset A$ de controles admisibles cuando el sistema se encuentra en el estado x.
- Entonces, la dinámica de las variables de estado puede modelarse mediante el sistema de ecuaciones en diferencias (ver observación 1.1.2)

$$x_{t+1} = (x_t + a_t - \xi_t)^+ = \text{máx}(x_t + a_t - \xi_t, 0), \tag{1.9}$$

con $t \in \mathbb{N}_0$ y $x_0 = \tilde{x} \in \mathbb{X}$.

- Supongamos que la evolución de este sistema se ha observado hasta la etapa t, de manera que se conoce la historia correspondiente mediante los valores específicos de $x_0, a_0, x_1, a_1, ..., x_t, a_t$, y supongamos además que en particular, $x_t = x$ y $a_t = a$.

  Notemos que la probabilidad de que $x_{t+1} = y$ dada la historia hasta la etapa t (t-historia), depende únicamente del último estado observado ($x_t = x$) y del control respectivo ($a_t = a$), sin importar la $(t-1)$-historia del sistema, ni el valor de t.
  En efecto, usando (1.9) tenemos que para todo $x, y \in \mathbb{X}$, $a \in A(x)$ y $t \in \mathbb{N}_0$

$$\begin{aligned} P[x_{t+1} = y | x_0, a_0, x_1, a_1, ..., x_{t-1}, a_{t-1}, x_t = x, a_t = a] \\ &= P[(x_t + a_t - \xi_t)^+ = y | h_{t-1}, a_{t-1}, x_t = x, a_t = a] \\ &= P[(x_t + a_t - \xi_t)^+ = y | x_t = x, a_t = a] \\ &= P_{x,y}(a), \end{aligned}$$

  lo cual representa la ley de transición del sistema.

  Además, como las v.a's ξ_t son i.i.d. con función de probabilidad común $\phi(\cdot)$, se sigue que

$$\begin{aligned} P[x_{t+1} = y | x_t = x, a_t = a] &= P[(x_t + a_t - \xi_t)^+ = y | x_t = x, a_t = a] \\ &= P[(x + a - \xi_t)^+ = y] \\ &= \sum_{k \in S_y} \phi(k), \end{aligned} \tag{1.10}$$

  donde $S_y := \{s \in \mathbb{N}_0 \ : \ (x_t + a_t - s)^+ = y\}$.

- Finalmente, definiendo las constantes λ y h como sigue

$$\begin{aligned} \lambda &: \text{precio (unitario) de producción,} \\ h &: \text{costo (unitario) de almacenamiento,} \end{aligned} \tag{1.11}$$

  tenemos que la función de costo por etapa, queda determinada por

$$c(x, a) = \lambda a + h E_\phi \left[(x + a - \xi_t)^+ \right].$$

Capítulo 2

Criterio de costo descontado

2.1. Introducción

En este capítulo vamos a analizar el problema de control óptimo asociado al criterio de costo total esperado α-descontado. Este análisis se realizará por medio de la ecuación de optimalidad, para lo cual demostraremos que la función de valor óptimo es la única solución. En base a este hecho mostramos la existencia de políticas óptimas. Finalmente introduciremos algunos algoritmos de aproximación para la función de valor.

Para una fácil referencia introducimos de nuevo los elementos que definen el PCO en el caso descontado.

Para $x \in \mathbb{X}$ y $\pi \in \Pi$, definimos el costo total esperado α-descontado al usar la política π cuando el estado inicial es $x_0 = x$, por

$$V_\alpha(\pi, x) := E_x^\pi[\sum_{t=0}^{\infty} \alpha^t c(x_t, a_t)] \tag{2.1}$$

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

En este caso, el PCO respectivo consiste en encontrar una política $\pi^* \in \Pi$ tal que minimice la función $V_\alpha(\pi, x)$, es decir,

$$V_\alpha(\pi^*, x) = \inf_{\pi \in \Pi} V_\alpha(\pi, x) \qquad \forall\ x \in \mathbb{X}.$$

Además la función de valor óptimo, a la cual llamaremos *función de valor α-óptimo*, cuando el estado inicial es $x_0 = x$ queda definida como

$$V^*(x) := \inf_{\pi \in \Pi} V_\alpha(\pi, x), \qquad x \in \mathbb{X}. \tag{2.2}$$

Entonces, llamaremos a π^* una *política α-óptima.*

2.2. Condiciones

En lo que resta del trabajo vamos a suponer que se cumplen las siguientes condiciones:

Hipótesis 2.2.1

(a) *Para cada $x \in \mathbb{X}$, $A(x)$ es un conjunto finito.*

(b) *Existe una constante $M > 0$ tal que*

$$|c(x, a)| \leq M, \qquad \forall \ \ (x, a) \in \mathbb{K}. \tag{2.3}$$

La Hipótesis 2.2.1 implica que $V_\alpha(\cdot, \cdot)$ está bien definido, como lo establece el siguiente resultado.

Proposición 2.2.2 *La Hipótesis 2.2.1 (b) implica que el índice $V_\alpha(\pi, x)$ en* (2.1) *está acotado.*

Demostración. Para cada $t \in \mathbb{N}_0$ definamos las v.a.'s X_t y Y_t como sigue

$$X_t := \alpha^t c(x_t, a_t) \ \text{ y } \ Y_t := |X_t|.$$

Entonces

$$P[X_t \leq Y_t] = 1 \quad \forall\, t \in \mathbb{N}_0,$$

lo cual implica, para cada $x \in \mathbb{X}$ y $\pi \in \Pi$

$$|E_x^\pi[X_t]| \leq E_x^\pi[Y_t] \quad \forall\, t \in \mathbb{N}_0.$$

Ahora, de (2.3) se sigue que

$$E_x^\pi[Y_t] \leq M\alpha^t < \infty \quad \forall\, t \in \mathbb{N}_0,$$

y por lo tanto, para cada $x \in \mathbb{X}$ y $\pi \in \Pi$:

$$\begin{aligned} |V_\alpha(\pi, x)| &= \left| \sum_{t=0}^{\infty} E_x^\pi [\alpha^t c(x_t, a_t)] \right| \\ &\leq M \sum_{t=0}^{\infty} \alpha^t = \frac{M}{1-\alpha} < \infty \qquad \forall\ x \in \mathbb{X}\ y\ \pi \in \Pi, \end{aligned} \tag{2.4}$$

debido a que $\alpha \in (0, 1)$. ■

Hemos probado que si la función de costo es una función acotada, entonces el índice de funcionamiento también es acotado. Esta propiedad nos facilitará el análisis puesto que nos permite apoyarnos en la teoría de funciones sobre espacios lineales normados para establecer los principales resultados de optimalidad α-descontada.

Denotaremos por $B(\mathbb{X})$ al espacio lineal normado de todas las funciones acotadas $v : \mathbb{X} \to \mathbb{R}$ con la norma

$$\|v\| := \sup_{x \in \mathbb{X}} |v(x)|. \tag{2.5}$$

Observación 2.2.3 *Como consecuencia se tiene que:*

(a) $B(\mathbb{X})$ *es un espacio de Banach (ver Apéndice B, Teorema B.0.8).*

(b) *De* (2.2) *y la Proposición 2.2.2 se tiene que* $V^* \in B(\mathbb{X})$, *y*

$$|V^*(x)| \leq \frac{M}{1-\alpha} \quad \forall\ x \in \mathbb{X}. \tag{2.6}$$

2.3. Ecuación de Optimalidad

Definición 2.3.1 *Diremos que una función* $u \in B(\mathbb{X})$ *es una solución de la ecuación de optimalidad* α*-descontada (EO) si*

$$u(x) = \min_{a \in A(x)} \left\{ c(x, a) + \alpha \sum_{y \in \mathbb{X}} u(y) P_{x,y}(a) \right\} \qquad \forall\ x \in \mathbb{X}. \tag{2.7}$$

El objetivo es demostrar que, bajo la Hipótesis 2.2.1, la función de valor α-óptimo satisface la EO, lo cual nos permitirá mostrar la existencia de políticas óptimas para el modelo MCM (1.1). Para tal fin introduciremos nueva notación, así como también algunos resultados preliminares.

Primeramente, para cada $u \in B(\mathbb{X})$ definimos los operadores

$$Tu(x) := \min_{a \in A(x)} \left\{ c(x,a) + \alpha \sum_{y \in \mathbb{X}} u(y) P_{x,y}(a) \right\} \qquad \forall\ x \in \mathbb{X}, \tag{2.8}$$

y para $f \in \mathbb{F}$

$$T_f u(x) := c(x,f) + \alpha \sum_{y \in \mathbb{X}} u(y) P_{x,y}(f). \tag{2.9}$$

Asimismo, definimos

$$T^t u := T[T^{t-1}u], \quad t \in \mathbb{N}, \tag{2.10}$$

y para cada $f \in \mathbb{F}$,

$$T_f^t u := T_f[T_f^{t-1}u], \quad t \in \mathbb{N},$$

donde $T^0 u = u$ y $T_f^0 u = u$.

Observaciones 2.3.2 **(a)** *Observemos que en términos del operador T, la EO queda expresada como*

$$u = Tu, \qquad u \in B(\mathbb{X}).$$

(b) *La Hipótesis 2.2.1 (a) garantiza que existe $f \in \mathbb{F}$ tal que*

$$Tu = T_f u, \qquad u \in B(\mathbb{X}).$$

(c) *De la hipótesis 2.2.1 (b), se tiene que para cada $u \in B(\mathbb{X})$ y $t \in \mathbb{N}_0$*

$$T^t u \in B(\mathbb{X})$$

y además

$$T_f^t u \in B(\mathbb{X}), \qquad f \in \mathbb{F}.$$

Los dos resultados que se muestran a continuación resaltan algunas propiedades importantes de ambos operadores, T y T_f, previamente definidos.

Proposición 2.3.3 *Bajo la Hipótesis 2.2.1 (b), T y T_f $(f \in \mathbb{F})$ son operadores de contracción módulo α sobre $B(\mathbb{X})$ con la norma (2.5), esto es, para cada par de funciones $u, v \in B(\mathbb{X})$:*

(a) $||Tu - Tv|| \leq \alpha ||u - v||$, *y*

(b) $||T_f u - T_f v|| \leq \alpha ||u - v||$.

Demostración.

(a) Primero tenemos que para cada $u, v \in B(\mathbb{X})$, $x \in \mathbb{X}$ y $a \in A(x)$ se cumple

$$c(x,a) + \alpha \sum_{y \in \mathbb{X}} u(y) P_{x,y}(a)$$

$$= c(x,a) + \left[\alpha \sum_{y \in \mathbb{X}} v(y) P_{x,y}(a) - \alpha \sum_{y \in \mathbb{X}} v(y) P_{x,y}(a) \right] + \alpha \sum_{y \in \mathbb{X}} u(y) P_{x,y}(a)$$

$$\leq c(x,a) + \alpha \sum_{y \in \mathbb{X}} v(y) P_{x,y}(a) + \alpha \sum_{y \in \mathbb{X}} |u(y) - v(y)| P_{x,y}(a)$$

$$\leq c(x,a) + \alpha \sum_{y \in \mathbb{X}} v(y) P_{x,y}(a) + \alpha \sup_{y \in \mathbb{X}} |u(y) - v(y)|.$$

Ahora, tomando el mínimo sobre $A(x)$ en ambos lados de la desigualdad y de acuerdo con (2.5) y (2.8), vemos que para cada $x \in \mathbb{X}$:

$$Tu(x) \leq Tv(x) + \alpha ||u - v||,$$

Lo cual es equivalente a la expresión

$$Tu(x) - Tv(x) \leq \alpha ||u - v|| \qquad \forall\ x \in \mathbb{X}. \tag{2.11}$$

Luego, siguiendo un procedimiento completamente análogo es posible observar que además se tiene

$$Tv(x) - Tu(x) \leq \alpha ||u - v|| \qquad \forall\ x \in \mathbb{X}, \tag{2.12}$$

de manera que de (2.11) y (2.12) obtenemos

$$|Tu(x) - Tv(x)| \leq \alpha ||u - v|| \qquad \forall\ x \in \mathbb{X}. \tag{2.13}$$

Finalmente, tomando el supremo sobre $\mathbb{X}$ en (2.13) se obtiene la afirmación de la parte (a).

(b) Para la parte (b) la demostración sigue un esquema similar al previo. ■

Observación 2.3.4 *Una consecuencia de la Proposición 2.3.3 y el inciso (a) del Teorema de Punto Fijo de Banach B.0.10 (Ver Apéndice B), es que ambos operadores, T y T_f, tienen un único punto fijo en $B(\mathbb{X})$.*

2.4. Existencia de políticas óptimas

Los resultados que se presentan en esta sección están orientados a mostrar que la única solución a la EO es la función de valor óptimo, así como a mostrar la existencia de políticas óptimas.

Proposición 2.4.1

(a) *El punto fijo del operador T_f es $V_\alpha(f,\cdot)$, es decir,*

$$V_\alpha(f,x) = T_f V_\alpha(f,x) \qquad \forall \;\; x \in \mathbb{X}. \tag{2.14}$$

(b) *Una política $\pi = \{f_t\}$ es α-óptima si, y solo si, $V_\alpha(\pi,x)$ es punto fijo del operador T.*

Demostración.

(a) Nótese que utilizando (2.1), junto con los Teoremas A.1.3, A.1.4 (c) y A.2.4 (véase Apéndice A) se obtiene lo siguiente

$$\begin{aligned} V_\alpha(f,x) \;:\; &= E_x^f\left[\sum_{t=0}^{\infty}\alpha^t c(x_t,a_t)\right] \\ &= c(x,f) + \alpha E_x^f\left[\sum_{t=1}^{\infty}\alpha^{t-1}c(x_t,a_t)\right] \\ &= c(x,f) + \alpha E_x^f\left[E_x^f\left[\sum_{t=1}^{\infty}\alpha^{t-1}c(x_t,a_t)|x_1,a_1\right]\right] \\ &= c(x,f) + \alpha E_x^f\left[V_\alpha(f,x_1)\right] \\ &= c(x,f) + \alpha\sum_{y\in\mathbb{X}} V_\alpha(f,y)P_{x,y}(f) \qquad \forall \;\; x\in\mathbb{X},\; f\in\mathbb{F}. \end{aligned}$$

Es decir, se cumple (2.14), y en consecuencia, $V_\alpha(f,\cdot)$ es el punto fijo de T_f.

(b) Primero, supongamos que

$$u(x) = V_\alpha(\pi,x)$$

es punto fijo de T para alguna $\pi \in \Pi$.

Entonces,

$$u(x) = \min_{a \in A(x)} \left\{ c(x,a) + \alpha \sum_{y \in \mathbb{X}} u(y) P_{x,y}(a) \right\}. \tag{2.15}$$

Sea $\pi' = \{f'_t\}$ una política arbitraria. Obsérvese que, de acuerdo con (1.6), se cumple lo siguiente

$$E_x^{\pi'}[\alpha^{t+1} u(x_{t+1}) | h_t, a_t] \;=\; \alpha^{t+1} \sum_{y \in \mathbb{X}} u(y) P_{x_t,y}(f'_t(h_t)).$$

De aquí y por (2.15), nótese que

$$\begin{aligned} E_x^{\pi'}[\alpha^{t+1} u(x_{t+1}) | h_t, a_t] &= \alpha^{t+1} \sum_{y \in \mathbb{X}} u(y) P_{x_t,y}\left(f'_t(h_t)\right) \pm \alpha^t c\left(x_t, f'_t(h_t)\right) \\ &= \alpha^t \left[c\left(x_t, f'_t(h_t)\right) + \alpha \sum_{y \in \mathbb{X}} u(y) P_{x_t,y}\left(f'_t(h_t)\right) \right] \\ &\quad -\alpha^t c\left(x_t, f'_t(h_t)\right) \\ &\geq \alpha^t u(x_t) - \alpha^t c\left(x_t, f'_t(h_t)\right). \end{aligned}$$

Es decir,

$$\alpha^t c(x_t, f'_t(h_t)) \geq \alpha^t u(x_t) - E_x^{\pi'}[\alpha^{t+1} u(x_{t+1}) | h_t, a_t],$$

de lo cual, por los Teoremas A.1.4(c) y A.2.4 se tiene

$$E_x^{\pi'}[\alpha^t c(x_t, a_t)] \geq \alpha^t E_x^{\pi'}[u(x_t)] - \alpha^{t+1} E_x^{\pi'}[u(x_{t+1})],$$

expresión en la que, sumando de ambos lados desde $t = 0$ hasta n, vemos que

$$E_x^{\pi'}[u(x_0)] - \alpha^{n+1} E_x^{\pi'}[u(x_{n+1})] \leq E_x^{\pi'}[\sum_{t=0}^{n} \alpha^t c(x_t, a_t)].$$

Tomando límite cuando $n \to \infty$, debido a que $u(x)$ es acotada y $\alpha \in (0,1)$, de obtiene que

$$u(x) \leq V_\alpha(\pi', x),$$

ésto es,

$$V_\alpha(\pi, x) \leq V_\alpha(\pi', x).$$

Como π' es arbitraria,

$$V_\alpha(\pi, x) = V^*(x)$$

Por consiguiente, π es una política α-óptima.

Ahora, supóngase que π es una política α-óptima, es decir,

$$u(x) = V_\alpha(\pi, x) = V^*(x).$$

Demostraremos que

$$u \geq Tu \qquad y \qquad u \leq Tu. \tag{2.16}$$

Para demostrar la primera desigualdad en (2.16) considérese la expresión

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

de la cual, por el Teorema A.2.4 vemos que

$$u(x) = c(x, f_0) + \alpha E_x^\pi [V_\alpha(\pi', x_1)],$$

donde $\pi' = \{f_t\}_{t \in \mathbb{N}}$, es decir, $\pi = \{f_0, \pi'\} = \{f_0, f_1, \ldots\}$.

De aquí,

$$u(x) \geq c(x, f_0) + \alpha E_x^\pi [u(x_1)]$$

por lo que se obtiene

$$u(x) \geq \min_{a \in A(x)} \left\{ c(x,a) + \alpha \sum_{y \in \mathbb{X}} u(y) P_{x,y}(a) \right\},$$

lo cual demuestra que $u \geq Tu$.

Con el fin de demostrar la segunda desigualdad en (2.16), sea $g \in \mathbb{F}$ arbitraria, y definiendo la política

$$\pi' = \{g, \pi\}$$

se tiene que

$$u(x) \leq V_\alpha(\pi', x),$$

de donde

$$u(x) \leq c(x, g) + \alpha E_x^{\pi'}[V_\alpha(\pi, x_1)].$$

Como $u(x_1) = V_\alpha(\pi, x_1)$, entonces

$$u(x) \leq c(x, g) + \alpha \sum_{y \in \mathbb{X}} u(y) P_{x,y}(g),$$

y, dado que $g \in \mathbb{F}$ es arbitraria, entonces

$$u(x) \leq \min_{a \in A(x)} \left\{ c(x, a) + \alpha \sum_{y \in \mathbb{X}} u(y) P_{x,y}(a) \right\},$$

lo cual conduce a que $u \leq Tu$. ■

Teorema 2.4.2 *Si se satisface la hipótesis 2.2.1, entonces:*

(a) *V^* es la única solución acotada de la EO.*

(b) *$\pi = \{f\}$ es una política α-óptima si y sólo si f minimiza el lado derecho de la EO, es decir,*

$$V^*(x) = c(x, f) + \alpha \sum_{y \in \mathbb{X}} V^*(y) P_{x,y}(f).$$

Demostración.

(a) Debido a que T es un operador de contracción y $B(\mathbb{X})$ es un espacio de Banach, entonces por el Teorema de Punto Fijo existe $u \in B(\mathbb{X})$ tal que

$$\begin{aligned} u(x) &= Tu(x) \\ &= \min_{a \in A(x)} \left\{ c(x, a) + \alpha \sum_{y \in \mathbb{X}} u(y) P_{x,y}(a) \right\}. \end{aligned}$$

Sea $g \in \mathbb{F}$ tal que

$$u(x) = T_g u(x).$$

Luego, por la Proposición 2.4.1 (a) se tiene

$$u(x) = V_\alpha(g, x),$$

lo cual implica que $\pi = g$ es una política α-óptima, y entonces

$$u(x) = V^*(x).$$

(b) Primero supongamos que $\pi = \{f\}$ es una política α-óptima. Entonces, de la Proposición 2.4.1 (b) se tiene

$$TV_\alpha(f, x) = V_\alpha(f, x) = V^*(x), \tag{2.17}$$

es decir,

$$\min_{a \in A(x)} \left\{ c(x, a) + \alpha \sum_{y \in \mathbb{X}} V_\alpha(f, y) P_{x,y}(a) \right\}$$

$$= \min_{a \in A(x)} \left\{ c(x, a) + \alpha \sum_{y \in \mathbb{X}} V^*(y) P_{x,y}(a) \right\}$$

y, dado que por la Proposición 2.4.1 (a)

$$V_\alpha(f, x) = T_f V_\alpha(f, x),$$

por (2.17) tenemos

$$c(x, f) + \alpha \sum_{y \in \mathbb{X}} V^*(y) P_{x,y}(f) = \min_{a \in A(x)} \left\{ c(x, a) + \alpha \sum_{y \in \mathbb{X}} V^*(y) P_{x,y}(a) \right\}.$$

Por consiguiente, f minimiza el lado derecho de la EO.

Supongamos ahora que f minimiza el lado derecho de la EO. En tal situación

$$V^*(x) = T_f V^*(x),$$

y, como por la Proposición 2.4.1 (a) se tiene

$$V^*(x) = V_\alpha(f, x),$$

entonces $\pi = \{f\}$ es una política α-óptima. ■

2.5. Aproximaciones

Como podemos observar, el Teorema 2.4.2 establece un procedimiento para resolver el PCO. En efecto, el primer paso es calcular la función de valor óptimo V^* como solución de la EO, y después, como un segundo paso resolver un problema de minimización para calcular la política óptima. Ambos problemas no son fáciles de resolver, y por lo tanto, desde el punto de vista de las aplicaciones, es importante contar con algoritmos de aproximación tanto para V^* como para la política óptima.

En esta sección presentamos tres algoritmos de aproximación para V^*. El primero es el llamado Algoritmo de Iteración de Valores, y los dos últimos están definidos mediante sucesiones de costos. El problema de aproximación de políticas óptimas lo analizaremos en el próximo capítulo.

2.5.1. Algoritmo de Iteración de valores

Observe que de la Proposición 2.3.3, el Teorema 2.4.2, y la Observación B.0.11 (Ver Apéndice B), para toda $u \in B(\mathbb{X})$ y $t \in \mathbb{N}_0$,

$$\|T^t u - V^*\| \leq \alpha^t \|u - V^*\|. \tag{2.18}$$

Definamos la sucesión de funciones de Iteración de Valores (IV) $\{v_t\}$ como

$$v_0 := 0, \tag{2.19}$$

$$v_t(x) := T v_{t-1}(x) = T^t v_0(x), \quad x \in \mathbb{X}. \tag{2.20}$$

De aquí,

$$v_t(x) = \min_{a \in A(x)} \left\{ c(x,a) + \sum_{y \in \mathbb{X}} v_{t-1}(y) P_{x,y}(a) \right\}, \quad t \in \mathbb{N},\ x \in \mathbb{X}. \tag{2.21}$$

En el resultado siguiente se garantiza la convergencia del algoritmo de Iteración de Valores a la función de valor α-óptimo.

Teorema 2.5.1 *Bajo la Hipótesis 2.2.1,*

$$\|v_t - V^*\| \to 0 \quad \textit{cuando} \quad t \to \infty. \tag{2.22}$$

Además, si

$$0 \leq c(x,a) \leq M \quad \forall \ \ (x,a) \in \mathbb{K}, \tag{2.23}$$

entonces, $v_t \nearrow V^*$, *cuando* $t \to \infty$.

Demostración.
La convergencia (2.22) de la sucesión $\{v_t\}$ a V^* se sigue de la desigualdad (2.18) tomando $u = v_0 = 0$ y de la relación (2.6). Es decir, de (2.18), (2.20) y (2.6),

$$\|v_t - V^*\| \leq \alpha^t \|V^*\| \leq \frac{\alpha^t M}{1-\alpha}.$$

Tomando límite cuando $t \to \infty$, obtenemos

$$\|v_t - V^*\| \to 0 \tag{2.24}$$

ya que $\alpha \in (0, 1)$.
Por otro lado, observemos que bajo (2.23) el operador T es monótono, es decir, si $u, v \in B(\mathbb{X})$ son tales que $u \leq v$, entonces es fácil ver que

$$Tu \leq Tv. \tag{2.25}$$

Además, como $v_0 = 0$, tenemos

$$v_0 = 0 \leq \min_{a \in A(x)} \left\{ c(x,a) + \alpha \sum_{y \in \mathbb{X}} v_0(y) P_{x,y}(a) \right\} = \min_{a \in A(x)} \{c(x,a)\} = v_1.$$

Entonces, por (2.25) se sigue que

$$Tv_0 \leq Tv_1$$

es decir,

$$v_1 \leq v_2.$$

Procediendo de manera inductiva se demuestra que

$$v_t \leq v_{t+1} \quad \forall\ t \in \mathbb{N}_0.$$

Por lo tanto, $v_t \nearrow V^*$. ■

2.5.2. Aproximación por medio de sucesiones de costos

A continuación vamos a presentar dos tipos de aproximaciones a la función de valor óptimo V^*, las cuales se definen por medio de sucesiones de costos que convergen a la función de costo por etapa c.

Para presentar estos algoritmos de aproximación, sea $\{c^n\}$ una sucesión de costos acotados, tales que $c^n : \mathbb{K} \to \mathbb{R}^+ \cup \{0\}$ y $c^n \uparrow c$.

Para cada $u \in \mathbb{B}(X)$ definimos el operador

$$T_n u(x) := \min_{a \in A(x)} \left\{ c^n(x,a) + \alpha \sum_{y \in \mathbb{X}} u(y) P_{x,y}(a) \right\}, \quad n \in \mathbb{N}. \tag{2.26}$$

Aproximaciones por sucesiones de costos acotados

Para cada $n \in \mathbb{N}$, definimos el índice de funcionamiento respecto al costo c^n, y su correspondiente función de valor como:

$$U_n(\pi, x) := E_x^\pi \left[\sum_{t=0}^{\infty} \alpha^t c^n(x_t, a_t) \right], \tag{2.27}$$

y

$$U_n^*(x) := \inf_{\pi \in \Pi} U_n(\pi, x). \tag{2.28}$$

Observe que del Teorema 2.4.2, para cada n, la función de valor U_n^* es la única solución acotada a la EO $U_n^* = T_n U_n^*$, es decir,

$$U_n^*(x) = \min_{a \in A(x)} \left\{ c^n(x,a) + \alpha \sum_{y \in \mathbb{X}} U_n^*(y) P_{x,y}(a) \right\} \quad \forall\ x \in \mathbb{X}. \tag{2.29}$$

El objetivo es demostrar que U_n^* converge a V^*. Para este fin, primero mostraremos el siguiente lema:

Lema 2.5.2 *Sean u y $\{u_n\}$ funciones acotadas en $\mathbb{K}$. Si $u_n \uparrow u$, entonces*

$$\lim_{n \to \infty} \min_{a \in A(x)} u_n(x,a) = \min_{a \in A(x)} u(x,a) \quad \forall\ x \in \mathbb{X}.$$

Demostración.
Para cada $x \in \mathbb{X}$ definimos

$$l(x) = \lim_{n \to \infty} \min_{a \in A(x)} u_n(x,a)$$

y

$$u^*(x) = \min_{a \in A(x)} u(x,a).$$

Como $u_n \uparrow u$, tenemos que

$$l(\cdot) \leq u^*(\cdot). \tag{2.30}$$

Para probar la desigualdad inversa, fijamos un estado arbitrario $x \in \mathbb{X}$, y consideramos los conjuntos

$$\begin{aligned} A_0 &:= \{a \in A(x) \mid u(x,a) = u^*(x)\} \\ A_n &:= \{a \in A(x) \mid u_n(x,a) \leq u^*(x)\} \quad \forall\, n \in \mathbb{N} \end{aligned}$$

Como $A(x)$ es un conjunto finito, cada uno de estos conjuntos es no vacío. Además como $u_n \uparrow u$, A_n decrece al conjunto A_0, esto es, $A_n \downarrow A_0$.
Para cada $n \geq 1$, sea $a_n \in A_n$ tal que $u_n(x,a_n) = \min\limits_{a \in A(x)} u_n(x,a)$, el cual existe porque $A(x)$ es finito.
Más aún, existen $a_0 \in A_0$ y una subsucesión $\{a_{n_i}\}$ de $\{a_n\}$ tal que $a_{n_i} \to a_0$.
Ahora, usando de nuevo que u_n es creciente tenemos que para cualquier $n \geq 1$ dada

$$u_{n_i}(x,a_{n_i}) = \min_{a \in A(x)} u_{n_i}(x,a) \geq u_n(x,a_{n_i}) \quad \forall\, n_i \geq n.$$

Tomando el límite cuando $i \to \infty$, tenemos que

$$\begin{aligned} l(x) &= \lim_{i\to\infty} \min_{a \in A(x)} u_{n_i}(x,a) \\ &= \lim_{i\to\infty} u_{n_i}(x,a_{n_i}) \\ &\geq u_n(x, \lim_{i\to\infty} a_{n_i}) \\ &\geq u_n(x,a_0) \end{aligned}$$

Por lo tanto, como $u_n \uparrow u$,

$$l(x) \geq u(x,a_0) = u^*(x). \tag{2.31}$$

Finalmente, como x es arbitraria, por (2.30) y (2.31), se sigue que

$$\lim_{n\to\infty} \min_{a \in A(x)} u_n(x,a) = l(x) = u^*(x) = \min_{a \in A(x)} u(x,a) \quad \forall\, x \in \mathbb{X}. \ \blacksquare$$

Proposición 2.5.3 *Bajo la Hipótesis 2.2.1 (a),*

$$U_n^* \uparrow V^*$$

Demostración.
Como $c^n \uparrow c$, se tiene que $c^n(x,a) \leq c(x,a)$ para todo $(x,a) \in \mathbb{K}$, lo cual implica que $U_n^* \leq V^*$ para toda $n \in \mathbb{N}$ y $\{U_n^*\}$ es creciente. Entonces existe una función $u \in B(\mathbb{X})$ tal que $u \leq V^*$ y $U_n^* \uparrow u$.
Demostraremos que $Tu = u$.
Dado que $U_n^* = T_n U_n^*$, se sigue que $u = \lim_{n\to\infty} U_n^* = \lim_{n\to\infty} T_n U_n^*$. Veamos que $\lim_{n\to\infty} T_n U_n^* = Tu$.
Aplicando el Lema 2.5.2 tenemos

$$\begin{aligned} u = \lim_{n\to\infty} T_n U_n^* &= \lim_{n\to\infty} \left[\min_{a\in A(x)} \left(c^n(x,a) + \alpha \sum_{y\in\mathbb{X}} U_n^*(y) P_{x,y}(a) \right) \right] \\ &= \min_{a\in A(x)} \left[\lim_{n\to\infty} \left(c^n(x,a) + \alpha \sum_{y\in\mathbb{X}} U_n^*(y) P_{x,y}(a) \right) \right] \\ &= \min_{a\in A(x)} \left[c(x,a) + \alpha \lim_{n\to\infty} \sum_{y\in\mathbb{X}} U_n^*(y) P_{x,y}(a) \right] \end{aligned}$$

Y aplicando el Teorema de convergencia monótona (ver Apéndice C, Teorema C.0.13), obtenemos

$$\begin{aligned} u &= \min_{a\in A(x)} \left[c(x,a) + \alpha \sum_{y\in\mathbb{X}} \lim_{n\to\infty} U_n^*(y) P_{x,y}(a) \right] \\ &= \min_{a\in A(x)} \left[c(x,a) + \alpha \sum_{y\in\mathbb{X}} u(y) P_{x,y}(a) \right] \\ &= Tu. \end{aligned}$$

Entonces, $Tu(x) = u(x)$, pero como V^* es la única solución acotada de la EO, llegamos a que $V^* = u$, y por lo tanto $U_n^* \uparrow V^*$. ■

Aproximaciones recursivas por costos acotados.

Este algoritmo es una combinación de los algoritmos de Iteración de valores y el de Sucesiones de costos presentado anteriormente. Sea $\{v_n'\}$ una sucesión definida recursivamente como $v_0' := 0$ y $v_n'(x) := T_n v_{n-1}'$ para

$n \geq 1$, es decir,

$$v'_n(x) := \min_{a \in A(x)} \left[c^n(x,a) + \alpha \sum_{y \in \mathbb{X}} v'_{n-1}(y) P_{x,y}(a) \right], \quad n \geq 1. \tag{2.32}$$

La convergencia de esta sucesión a la función de valor es mostrada en el siguiente resultado.

Proposición 2.5.4 *Bajo la Hipótesis 2.2.1 (a),*

$$v'_n \uparrow V^*$$

Demostración.
La demostración es similar a la de la Proposición 2.5.3.
Como $c^n \uparrow c$, se tiene que $c^n(x,a) \leq c(x,a)$ para todo $(x,a) \in \mathbb{K}$, lo cual implica que $v'_n \leq V^*$ para toda $n \in \mathbb{N}$. Y como $\{c^n\}$ es una sucesión creciente, es claro que $\{v'_n\}$ es creciente. Entonces existe una función $u \in B(\mathbb{X})$ tal que $u \leq V^*$ y $v'_n \uparrow u$.

Demostraremos que $Tu = u$.

Dado que $v'_n = T_n v'_{n-1}$, se sigue que $u = \lim_{n\to\infty} v'_n = \lim_{n\to\infty} T_n v'_{n-1}$. Por lo tanto es suficiente demostrar que $\lim_{n\to\infty} T_n v'_{n-1} = Tu$.
Aplicando el Lema 2.5.2 tenemos

$$\begin{aligned} u = \lim_{n\to\infty} T_n v_{n-1} &= \lim_{n\to\infty} \left[\min_{a\in A(x)} \left(c^n(x,a) + \alpha \sum_{y\in\mathbb{X}} v'_{n-1}(y) P_{xy}(a) \right) \right] \\ &= \min_{a\in A(x)} \left[\lim_{n\to\infty} \left(c^n(x,a) + \alpha \sum_{y\in\mathbb{X}} v'_{n-1}(y) P_{xy}(a) \right) \right] \\ &= \min_{a\in A(x)} \left[c(x,a) + \alpha \lim_{n\to\infty} \sum_{y\in\mathbb{X}} v'_{n-1}(y) P_{xy}(a) \right] \end{aligned}$$

Ahora, aplicando el Teorema de Convergencia Monótona (ver Apéndice C,

Teorema C.0.13), obtenemos

$$\begin{aligned} u &= \min_{a \in A(x)} \left[c(x,a) + \alpha \sum_{y \in \mathbb{X}} \lim_{n \to \infty} v'_{n-1}(y) P_{xy}(a) \right] \\ &= \min_{a \in A(x)} \left[c(x,a) + \alpha \sum_{y \in \mathbb{X}} u(y) P_{xy}(a) \right] \\ &= Tu. \end{aligned}$$

Entonces, $Tu(x) = u(x)$, pero como V^* es la única solución acotada de la EO, llegamos a que $V^* = u$, y por lo tanto $v'_n \uparrow V^*$. ■

Capítulo 3

Optimalidad asintótica

Como se vio en el Capítulo 2, el procedimiento estándar para resolver un PCO bajo el criterio de costo descontado es:

1. Mostrar que la función de valor óptimo V^* es una solución de la EO:
$$V^* = TV^*.$$
2. Resolver un problema de minimización para calcular políticas óptimas.

Para resolver el Problema 1 se propusieron algoritmos de aproximación los cuales consisten en definir sucesiones de funciones, denotado en forma genérica como $w = \{w_n\}$, tal que $w_n \to V^*$. En cada paso del algoritmo nos podemos plantear un problema de minimización, cuya solución produce una política, digamos, $\pi_w = \{f_0, f_1, \ldots\}$, donde $f_n \in \mathbb{F}$ es el minimizador correspondiente a la función w_n. Entonces la pregunta que surge es si la política π_w resuelve el Problema 2. La respuesta no necesariamente es afirmativa debido a las características del criterio de costo descontado. En efecto, observemos que el índice $V_\alpha(\cdot, \cdot)$ depende fuertemente de las decisiones que se toman en las primeras etapas, precisamente donde la información acerca de V^* que proporcionan los algoritmos de aproximación es deficiente. Por lo tanto una política como π_w, no necesariamente es óptima.

Para analizar la optimalidad de este tipo de políticas es necesario introducir un tipo de optimalidad más debil, a la cual se le conoce como *optimalidad asintótica*.

El objetivo del presente capítulo es demostrar la optimalidad asintótica de las políticas que produce el algoritmo de Iteración de valores, y los algoritmos definidos por medio de sucesiones de costos. Más aún, demostraremos

que el límite de estas políticas definen una política óptima en el sentido usual.

Existen varias maneras de analizar la optimalidad asintótica, una de ellas es por medio de la llamada *Función de Discrepancia*, la cual introducimos en la Sección 1, así como sus propiedades. En las siguientes secciones estableceremos los resultados principales del trabajo.

3.1. La función de discrepancia

Definición 3.1.1

a) *La función de discrepancia* $\Phi : \mathbb{K} \to \mathbb{R}$ *está definida por*

$$\Phi(x,a) := c(x,a) + \alpha \sum_{y \in X} V^*(y) P_{xy}(a) - V^*(x),$$

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

b) *Para una política arbitraria* $\pi \in \Pi$ *y un entero* $n \geq 0$, *definimos el costo descontado esperado de la etapa* n *en adelante, dado el estado inicial* $x_0 = x$, *como*

$$V_n(\pi, x) = E_x^{\pi} \left[\sum_{t=n}^{\infty} \alpha^{t-n} c(x_t, a_t) \right].$$

c) *Definimos, para* $n = 0$, $M_0 = V^*(x_0)$, *y para* $n = 1, 2, ...$

$$M_n = \sum_{t=0}^{n-1} \alpha^t c(x,a) + \alpha^n V^*(x_n).$$

Nótese que $V_0(\pi, \cdot) = V_\alpha(\pi, \cdot)$, es el costo descontado total esperado cuando se utiliza la política π.

A continuación veremos algunas propiedades de Φ, V_α y V_n así como sus relaciones:

Lema 3.1.2 *Para cada* $\pi \in \Pi$ *y* $x \in \mathbb{X}$,

$$V(\pi,x) = E_x^{\pi}[\sum_{t=0}^{n-1} \alpha^t c(x_t, a_t)] + \alpha^n V_n(\pi, x) = E_x^{\pi}(M_n) + \alpha^n [V_n(\pi, x) - E_x^{\pi} V^*(x_n)]. \tag{3.1}$$

Demostración.
La relación (3.1) se sigue de lo siguiente:

Para cada $\pi \in \Pi$, $x \in \mathbb{X}$ y $n \in \mathbb{N}$,

$$E_x^\pi(M_n) + \alpha^n[V_n(\pi,x) - E_x^\pi V^*(x_n)]$$
$$= E_x^\pi[\sum_{t=0}^{n-1} \alpha^t c(x_t,a_t)] + E_x^\pi[\alpha^n V^*(x_n)] + E_x^\pi[\sum_{t=n}^{\infty} \alpha^t c(x_t,a_t)] - E_x^\pi[\alpha^n V^*(x_n)]$$
$$= E_x^\pi[\sum_{t=0}^{n-1} \alpha^t c(x_t,a_t)] + \alpha^n V_n(\pi,x) = E_x^\pi[\sum_{t=0}^{\infty} \alpha^t c(x_t,a_t)] = V(\pi,x). \quad \blacksquare$$

Lema 3.1.3 *Supongamos que se cumple la Hipótesis 2.2.1. Entonces*

a) $\Phi(\cdot,\cdot)$ *es una función acotada y no negativa.*

b) $\min_{a \in A(x)} \Phi(x,a) = 0$ *para todo estado* $x \in \mathbb{X}$.

c) *Una política estacionaria* $f^* \in \mathbb{F}$ *es óptima si y sólo si* $\Phi(x, f^*(x)) = 0$.

Más aún, para cualquier política π *y* $x \in \mathbb{X}$

d)

$$\begin{aligned} \Phi(x_t,a_t) &= E_x^\pi[c(x_t,a_t) + V^*(x_{t+1}) - V^*(x_t)|h_t,a_t] \\ &= c(x_t,a_t) + \alpha \sum_{y \in \mathbb{X}} V^*(y) P_{x_t y}(a_t) - V^*(x_t) \qquad t \geq 0. \end{aligned} \tag{3.2}$$

e) $\sum_{t=n}^{\infty} \alpha^{t-n} E_x^\pi[\Phi(x_t,a_t)] = V_n(\pi,x) - E_x^\pi[V^*(x_n)], \;\; n \in \mathbb{N}_0$.

Demostración.

a) Que $\Phi(x,a)$ es acotada se sigue de la Hipótesis 2.2.1 (b) y (2.6). En efecto:

$$\begin{aligned} |\Phi(x,a)| &= |c(x,a) + \alpha \sum_{y \in \mathbb{X}} V^*(y) P_{xy}(a) - V^*(x)| \\ &\leq |c(x,a)| + |\alpha \sum_{y \in \mathbb{X}} V^*(y) P_{xy}(a)| + |V^*(x)| \\ &\leq M + |\frac{M\alpha}{1-\alpha} \sum_{y \in \mathbb{X}} P_{xy}(a)| + \frac{M}{1-\alpha} = \frac{2M}{1-\alpha} \quad \forall\ (x,a) \in \mathbb{K}. \end{aligned}$$

Además, $\Phi(x,a)$ es no negativa ya que:

$$\begin{aligned}\Phi(x,a) &= c(x,a) + \alpha \sum_{y\in X} V^*(y)P_{xy}(a) - V^*(x) \\ &= c(x,a) + \alpha \sum_{y\in X} V^*(y)P_{xy}(a) - \min_{a\in A(x)}\{c(x,a) + \alpha \sum_{y\in X} V^*(y)P_{xy}(a)\} \\ &\geq 0.\end{aligned}$$

b) Para todo $x \in \mathbb{X}$,

$$\begin{aligned}\min_{a\in A(x)} \Phi(x,a) &= \min_{a\in A(x)}\{c(x,a) + \alpha \sum_{y\in \mathbb{X}} V^*(y)P_{xy}(a) - V^*(x)\} \\ &= \min_{a\in A(x)}\{c(x,a) + \alpha \sum_{y\in \mathbb{X}} V^*(y)P_{xy}(a)\} - V^*(x) \\ &= V^*(x) - V^*(x) = 0.\end{aligned}$$

c) ($\Rightarrow$) Supongamos que $f^* \in \mathbb{F}$ es óptima, esto es, $V(f^*,x) = V^*(x)$. Entonces por el Teorema 2.4.2

$$V^*(x) = c(x, f^*(x)) + \alpha \sum_{y\in\mathbb{X}} V^*(y)P_{xy}(f^*(x)).$$

De aquí que,

$$\begin{aligned}\Phi(x, f^*(x)) &= c(x, f^*(x)) + \alpha \sum_{y\in\mathbb{X}} V^*(y)P_{xy}(f^*(x)) - V^*(x) \\ &= V^*(x) - V^*(x) = 0.\end{aligned}$$

($\Leftarrow$) Si $\Phi(x, f^*(x)) = 0$, entonces

$$c(x, f^*(x)) + \alpha \sum_{y\in\mathbb{X}} V^*(y)P_{xy}(f^*(x)) - V^*(x) = 0.$$

De aquí se tiene que,

$$V^*(x) = c(x, f^*(x)) + \alpha \sum_{y\in\mathbb{X}} V^*(y)P_{xy}(f^*(x)).$$

Es decir, f^* minimiza el lado derecho de la EO, y por el Teorema 2.4.2 se sigue que $\pi = f^*$ es una política α-óptima.

d) Esta parte es consecuencia de la relación (1.6) y las propiedades de la esperanza condicional. Es decir, para $t \geq 0$.

$$\begin{aligned} E_x^\pi[c(x_t,a_t)+\alpha V^*(x_{t+1})-V^*(x_t)|h_t,a_t] &= c(x_t,a_t)+\alpha E_x^\pi[V^*(x_{t+1})|h_t,a_t]-V^*(x_t) \\ &= c(x_t,a_t)+\alpha\sum_{y\in\mathbb{X}} V^*(y)P_{x_ty}(a_t)-V^*(x_t) \\ &= \Phi(x_t,a_t). \end{aligned}$$

e) Por el inciso (d),

$$\Phi(x_t,a_t) = E_x^\pi[c(x_t,a_t)+\alpha V^*(x_{t+1})-V^*(x_t)|h_t,a_t]$$

Usando las propiedades de la esperanza condicional del Apéndice A, tenemos lo siguiente,

$$\begin{aligned} E_x^\pi[\Phi(x_t,a_t)] &= E_x^\pi[E_x^\pi[c(x_t,a_t)+\alpha V^*(x_{t+1})-V^*(x_t)|h_t,a_t]] \\ &= E_x^\pi[c(x_t,a_t)+\alpha V^*(x_{t+1})-V^*(x_t)]. \end{aligned}$$

Así,

$$\begin{aligned} \sum_{t=n}^{\infty}\alpha^{t-n}E_x^\pi[\Phi(x_t,a_t)] &= \sum_{t=n}^{\infty}\alpha^{t-n}E_x^\pi[c(x_t,a_t)+\alpha V^*(x_{t+1})-V^*(x_t)] \\ &= \sum_{t=n}^{\infty}\alpha^{t-n}E_x^\pi[c(x_t,a_t)]+\sum_{t=n}^{\infty}\alpha^{t-n}E_x^\pi[\alpha V^*(x_{t+1})-V^*(x_t)] \\ &= E_x^\pi[\sum_{t=n}^{\infty}\alpha^{t-n}c(x_t,a_t)]-E_x^\pi[V^*(x_n)] \\ &= V_n(\pi,x)-E_x^\pi[V^*(x_n)]. \ \blacksquare \end{aligned}$$

Ahora veamos algunas propiedades del proceso M_n de la Definición 3.1.1 (c).

Lema 3.1.4 *Para cualquier política π, $\{M_n\}$ es una sub-martingala, es decir, para todo $x \in \mathbb{X}$ y $n \geq 0$,*

$$E_x^\pi(M_{n+1}|h_n) \geq M_n; \tag{3.3}$$

Entonces,

$$E_x^\pi(M_{n+1}) \geq E_x^\pi(M_n) \geq E_x^\pi(M_0) = V^*(x). \tag{3.4}$$

Demostración.

Por definición $M_0 = V^*(x)$ y $M_n = \sum_{t=0}^{n-1} \alpha^t c(x_t, a_t) + \alpha^n V^*(x_n) \quad \forall\, n = 1, 2, ...$

Así, podemos expresar

$$M_{n+1} = M_n + \alpha^n [c(x_n, a_n) + \alpha V^*(x_{n+1}) - V^*(x_n)]$$

Entonces, por el Lema 3.1.3 (d)

$$\begin{aligned} E_x^\pi(M_{n+1}|h_n) &= M_n + \alpha^n E_x^\pi[c(x_n, a_n) + \alpha V^*(x_{n+1}) - V^*(x_n)|h_n] \\ &= M_n + \alpha^n E_x^\pi[\Phi(x_n, a_n)|h_n] \geq 0, \end{aligned} \tag{3.5}$$

donde la última desigualdad se sigue del hecho de que $\Phi(\cdot, \cdot) \geq 0$. Por lo tanto

$$E_x^\pi(M_{n+1}|h_n) \geq M_n,$$

lo cual demuestra (3.3).

Calculando esperanza en ambos lados de esta desigualdad obtenemos

$$E_x^\pi(M_{n+1}) = E_x^\pi[E_x^\pi(M_{n+1}|h_t)] \geq E_x^\pi(M_n).$$

$$E_x^\pi(M_{n+1}) \geq E_x^\pi(M_n) \geq E_x^\pi(M_{n-1}) \geq ... \geq E_x^\pi(M_0) = V^*(x). \quad \blacksquare$$

3.2. Optimalidad asintótica

En esta sección introduciremos la noción de optimalidad asintótica. Supondremos que la Hipótesis 2.2.1 se cumple.

Definición 3.2.1

(a) *Una política π se dice ser asintóticamente óptima descontada (ADO) si, para todo $x \in \mathbb{X}$,*

$$V_n(\pi, x) - E_x^\pi[V^*(x_n)] \to 0 \quad \textit{cuando } n \to \infty. \tag{3.6}$$

(b) *Una política $\pi = \{f_n\}$ se dice ser puntualmente asintóticamente óptima descontada (ADOP) si, para cada $x \in \mathbb{X}$,*

$$\Phi(x, f_n(x)) \to 0 \quad \textit{cuando } n \to \infty. \tag{3.7}$$

Antes de presentar los resultados relacionados con optimalidad asintótica estableceremos la siguiente caracterización de la optimalidad estándar, la cual servirá como motivación.

Teorema 3.2.2 *Bajo la Hipótesis 2.2.1, los siguientes enunciados son equivalentes:*

(a) π *es una política óptima.*

(b) $V_n(\pi, x) = E_x^\pi[V^*(x_n)] \quad \forall\ n \in \mathbb{N}_0,\ x \in \mathbb{X}$

(c) $E_x^\pi[\Phi(x_n, a_n)] = 0 \quad \forall\ n \in \mathbb{N}_0\ ,\ x \in \mathbb{X}$

(d) $\{M_n\}$ *es una martingala (con respecto a* $P_x^\pi \ \forall\ x \in \mathbb{X}$*).*

Demostración.

(a)⇒(b) Como $\Phi \geq 0$, el Lema 3.1.3 (e) implica que para toda $\pi \in \Pi$, $x \in \mathbb{X}$ y $n \in \mathbb{N}_0$

$$0 \leq \sum_{t=n}^{\infty} \alpha^{t-n} E_x^\pi[\Phi(x_t, a_t)] = V_n(\pi, x) - E_x^\pi[V^*(x_n)]. \tag{3.8}$$

Por lo tanto,

$$V_n(\pi, x) \geq E_x^\pi[V^*(x_n)]. \tag{3.9}$$

Ahora supongamos que $\pi \in \Pi$ es óptima, es decir, $V^*(x) = V(\pi, x)$, $x \in \mathbb{X}$. Entonces, por el Lema 3.1.2 y (3.4)

$$\begin{aligned} V^*(x) &= E_x^\pi(M_n) + \alpha^n[V_n(\pi, x) - E_x^\pi V^*(x_n)] \\ &\geq V^*(x) + \alpha^n[V_n(\pi, x) - E_x^\pi V^*(x_n)]. \end{aligned}$$

Lo cual implica que

$$V_n(\pi, x) \leq E_x^\pi V^*(x_n) \tag{3.10}$$

De (3.9) y (3.10) se sigue que

$$V_n(\pi, x) = E_x^\pi V^*(x_n), \quad \forall\ n \in \mathbb{N}_0.$$

(b)⇒(c) Supongamos que (b) se cumple. Por el Lema 3.1.3 (e),

$$\sum_{t=n}^{\infty} \alpha^{t-n} E_x^\pi[\Phi(x_t, a_t)] = V_n(\pi, x) - E_x^\pi[V^*(x_n)] = 0, \ \ \forall\ n \in \mathbb{N}_0,\ x \in \mathbb{X}.$$

Como $\Phi \geq 0$, esto implica que,

$$E_x^\pi[\Phi(x_n, a_n)] = 0 \quad \forall\ n \in \mathbb{N}_0,\ x \in \mathbb{X}.$$

(c)⇒(d) Si (c) se cumple, entonces

$$E_x^\pi[\Phi(x_n, a_n)] = E_x^\pi[E_x^\pi[\Phi(x_n, a_n)|h_n]] = 0$$

Luego, como $\Phi \geq 0$ y por ecuación (3.5)

$$E_x^\pi[M_{n+1}|h_n] = M_n + \alpha^n E_x^\pi[\Phi(x_n, a_n)|h_n] = M_n,$$

es decir, $\{M_n\}$ es una martingala.

(d)⇒(a) Obsérvese que de (2.6)

$$|\alpha^n V^*(x_n)| \leq \frac{\alpha^n M}{1-\alpha} \quad \forall\ n \geq 0.$$

Esto implica que

$$\alpha^n E_x^\pi V^*(x_n) \to 0 \quad \text{cuando} \quad n \to \infty.$$

De aqui, por la Definición 3.1.1

$$E_x^\pi[M_n] = E_x^\pi[\sum_{t=0}^{n-1} \alpha^t c(x_t, a_t) + \alpha^n V^*(x_n)] \to V_\alpha(\pi, x) \quad \forall\, \pi \in \Pi,\ x \in \mathbb{X}.$$

Entonces, por la relación (3.4), $\{E_x^\pi(M_n)\}$ es una sucesión no-decreciente que converge a $V_\alpha(\pi, x)$, es decir,

$$V^*(x) \leq E_x^\pi[M_n] \leq E_x^\pi[M_{n+1}] \leq V(\pi, x) \quad \forall\ \pi \in \Pi,\ x \in \mathbb{X}. \quad (3.11)$$

Ahora, supongamos que $\{M_n\}$ es una martingala, es decir,

$$E_x^\pi[M_{n+1}|h_n] = M_n \quad \forall\ n \in \mathbb{N}_0,\ x \in \mathbb{X}.$$

para alguna política fija $\pi \in \Pi$. Entonces

$$E_x^\pi[M_{n+1}] = E_x^\pi[M_n] \quad \forall\ n \in \mathbb{N}_0.$$

Esto implica que la sucesión $\{E_x^\pi[M_n]\}$ es constante, y por (3.11) concluimos que

$$V^*(x) = V(\pi, x) \quad \forall\ x \in \mathbb{X}.$$

Por lo tanto $\pi \in \Pi$ es óptima. ■

Observe que del Teorema 3.2.2 (a)-(b), se sigue que si π es una política óptima, entonces π es ADO. En este sentido, la optimalidad asintótica es más debil que la optimalidad usual.

Como consecuencia del Teorema 3.2.2 tenemos la siguiente caracterización de la optimalidad asintótica.

Teorema 3.2.3 *Los siguientes enunciados son equivalentes:*

(a) π *es ADO.*

(b) *Para todo* $x \in \mathbb{X}$, $\sum_{t=n}^{\infty} \alpha^{t-n} E_x^{\pi}[\Phi(x_t, a_t)] \to 0$ *cuando* $n \to \infty$.

(c) *Para todo* $x \in \mathbb{X}$, $E_x^{\pi}[\Phi(x_t, a_t)] \to 0$ *cuando* $t \to \infty$.

(d) *Para todo* $x \in \mathbb{X}$, $\Phi(x_t, a_t) \to 0$ *en probabilidad -*P_x^{π} *cuando* $t \to \infty$.

Demostración.
La equivalencia de (a) y (b) se sigue de la ecuación (3.8).
La equivalencia de (b) y (c) es directa (ver Lema 3.1.3 (a)).
Se tiene que (c) implica (d), pues convergencia en media implica convergencia en probabilidad, por el Teorema A.3.2 (ver Apéndice A), e inversamente, (d) implica (c) por el Teorema de Convergencia Dominada C.0.15 (ver Apéndice C). ■

3.3. Optimalidad asintótica y algoritmos de aproximación

En esta sección analizaremos la optimalidad asintótica de las políticas que producen el algoritmo de iteración de valores y los algoritmos de sucesiones de costos.

3.3.1. Política Iteración de Valores

Recordemos que el algoritmo de IV se define como (ver (2.19), (2.20) y (2.21))

$$v_0 = 0$$

$$v_t(x) = Tv_{t-1}(x) = \min_{a \in A(x)} \left\{ c(x,a) + \sum_{y \in \mathbb{X}} v_{t-1}(y) P_{x,y}(a) \right\}, \quad t \in \mathbb{N},\ x \in \mathbb{X}.$$

Además, por (2.24),

$$\|v_t - V^*\| \to 0 \quad \text{cuando} \quad t \to \infty. \tag{3.12}$$

Definición 3.3.1 *Sea $\{v_t\}$ la sucesión de funciones de IV, y sea $\hat{\pi} = \{\hat{f}_t\}$ la política markoviana donde $\hat{f}_0 \in \mathbb{F}$ es arbitraria y para $t \geq 1$, $\hat{f}_t$ satisface*

$$v_t(x) = c(x, \hat{f}_t) + \alpha \sum_{y \in \mathbb{X}} v_{t-1}(y) P_{x,y}(\hat{f}_t(x)), \quad x \in \mathbb{X}.$$

Llamamos a $\hat{\pi}$ política de iteración de valores.

Teorema 3.3.2 *Bajo la Hipótesis 2.2.1, la política de IV $\hat{\pi}$ es ADO.*

Demostración.
Para cada $t \in \mathbb{N}_0$, definamos la función $\hat{\Phi}_t : \mathbb{K} \to \mathbb{R}$ como

$$\hat{\Phi}_t(x, a) = c(x, a) + \alpha \sum_{y \in \mathbb{X}} v_{t-1}(y) P_{xy}(a) - v_t(x).$$

Observe que por la definición de la política $\hat{\pi} = \{\hat{f}_t\}$ tenemos que $\hat{\Phi}_t(x, \hat{f}_t(x)) = 0$, para todo $x \in \mathbb{X}$. Entonces, para cada $t \in \mathbb{N}_0$,

$$\begin{aligned} \Phi(x_t, \hat{f}_t(x_t)) &\leq |\Phi(x_t, \hat{f}_t(x_t)) - \hat{\Phi}_t(x_t, \hat{f}_t(x_t))| \\ &\leq \sup_{(x,a) \in \mathbb{K}} |\Phi(x, a) - \hat{\Phi}_t(x, a)|. \end{aligned}$$

Por lo tanto, por el Teorema 3.2.3, para demostrar que $\hat{\pi}$ es ADO es suficiente demostrar

$$\sup_{(x,a) \in \mathbb{K}} |\Phi(x, a) - \hat{\Phi}_t(x, a)| \to 0 \quad \text{cuando} \quad t \to \infty. \tag{3.13}$$

Esto se deduce de lo siguiente:

$$\begin{aligned} |\Phi(x, a) - \hat{\Phi}_t(x, a)| &= |c(x, a) + \alpha \sum_{y \in \mathbb{X}} V^*(y) P_{x,y}(a) - V^*(x) \\ &\qquad - c(x, a) - \alpha \sum_{y \in \mathbb{X}} v_{t-1}(y) P_{x,y}(a) + v_t(x)| \\ &\leq \alpha \sum_{y \in \mathbb{X}} |V^*(y) - v_{t-1}(y)| P_{x,y}(a) + |V^*(x) - v_t(x)| \\ &\leq \alpha \|V^* - v_{t-1}\| + \|V^* - v_t\|, \quad t \in \mathbb{N}_0, \ (x, a) \in \mathbb{K}. \end{aligned}$$

Por lo tanto,

$$\sup_{(x,a)\in\mathbb{K}} |\Phi(x,a) - \hat{\Phi}_t(x,a)| \leq \alpha\|V^* - v_{t-1}\| + \|V^* - v_t\|.$$

Tomando límite cuando $t \to \infty$, la relación (3.12) implica (3.13), es decir, $\hat{\pi}$ es ADO. ■

3.3.2. Política bajo sucesiones de costos

Sea $\{c^n\}$ una sucesión de funciones de costo $c^n : \mathbb{K} \to \mathbb{R}$ tal que $c^n \nearrow c$ y U_n^* la función de valor óptimo correspondiente (ver (2.27), (2.28)). Recordemos que para cada $n \in \mathbb{N}_0$, U_n^* es la única solución de la EO:

$$U_n^*(x) = T_n U_n^*(x) = \min_{a\in A(x)} \left\{ c^n(x,a) + \alpha \sum_{y\in\mathbb{X}} U_n^*(y) P_{x,y}(a) \right\} \quad \forall\, x \in \mathbb{X},\ n \in \mathbb{N}_0. \tag{3.14}$$

Además, por la Proposición 2.5.3,

$$U_n^* \nearrow V^*. \tag{3.15}$$

Sea $\pi^* = \{f_n^*\}$ la política markoviana tal que $f_n^*(x) \in A(x)$ alcanza el mínimo en (3.14), es decir,

$$U_n^*(x) = c^n(x, f_n^*(x)) + \alpha \sum_{y\in\mathbb{X}} U_n^*(y) P_{x,y}(f_n^*(x)), \quad x \in \mathbb{X}.$$

Teorema 3.3.3 *Bajo la Hipótesis 2.2.1, la política π^* es puntualmente ADO.*

Demostración.
Para cada $n \in \mathbb{N}_0$, definimos la función

$$\Phi_n^*(x,a) = c^n(x,a) + \alpha \sum_{y\in\mathbb{X}} U_n^*(y) P_{x,y}(a) - U_n^*(x), \quad x \in \mathbb{X}.$$

Observe que $\Phi_n^*(x, f_n^*(x)) = 0$.

Por otro lado, para cada $(x,a) \in \mathbb{K}$ y $n \in \mathbb{N}_0$

$$
\begin{aligned}
&|\Phi(x,a) - \Phi_n^*(x,a)| \\
&= |c(x,a) + \alpha \sum_{y\in\mathbb{X}} V^*(y)P_{xy}(a) - V^*(x) - c^n(x,a) - \alpha \sum_{y\in\mathbb{X}} U_n^*(y)P_{xy}(a) + U_n^*(x)| \\
&\leq |c(x,a) - c^n(x,a)| + \alpha |\sum_{y\in\mathbb{X}} V^*(y)P_{xy}(a) - \alpha \sum_{y\in\mathbb{X}} U_n^*(y)P_{xy}(a)| + |V^*(x) - U_n^*(x)| \\
&\leq |c(x,a) - c^n(x,a)| + \alpha \sum_{y\in\mathbb{X}} |V^*(y) - U_n^*(y)|P_{xy}(a) + |V^*(x) - U_n^*(x)|.
\end{aligned}
$$

Ahora observemos que como $|V^*(x) - U_n^*(x)|$ es acotada, por el Teorema de la convergencia dominada C.0.15 (ver Apéndice C), el hecho de que $A(x)$ es finito y (3.15),

$$\sup_{a\in A(x)} \sum_{y\in\mathbb{X}} |V^*(y) - U_n^*(y)|P_{xy}(a) \to 0 \quad \text{cuando} \quad n \to \infty,$$

para cada $x \in \mathbb{X}$.
Como $c^n \nearrow c$ y $A(x)$ es finito,

$$\sup_{a\in A(x)} |c^n(x,a) - c(x,a)| \to 0 \quad \text{cuando} \quad n \to \infty.$$

De esta manera, llegamos a que para cada $x \in \mathbb{X}$,

$$\sup_{a\in A(x)} |\Phi(x,a) - \Phi_n^*(x,a)| \to 0 \quad \text{cuando} \quad n \to \infty. \tag{3.16}$$

Finalmente observemos que

$$\Phi(x, f_n^*(x)) = |\Phi(x, f_n^*(x)) - \Phi_n^*(x, f_n^*(x))| \leq \sup_{a\in A(x)} |\Phi(x,a) - \Phi_n^*(x,a)|, \quad x \in \mathbb{X}.$$

Por lo tanto (3.16) implica que π^* es ADO puntualmente. ■

3.3.3. Política bajo sucesión recursiva de costos

Consideremos de nuevo una sucesión $\{c^n\}$ de funciones de costo $c^n : \mathbb{K} \to \mathbb{R}$ tal que $c^n \nearrow c$. Sea $\{v_n'\}$ la sucesión de funciones definida como (ver (2.32))

$$v_0' := 0,$$

$$v'_n(x) := T_n v'_{n-1}(x) = \min_{a\in A(x)} \left[c^n(x,a) + \alpha \sum_{y\in\mathbb{X}} v'_{n-1}(y) P_{xy}(a) \right], \quad n \geq 1. \tag{3.17}$$

Por la Proposición 2.5.4 tenemos que

$$v'_n \nearrow V^* \quad \text{cuando } n \to \infty. \tag{3.18}$$

Sea $\pi' = \{f'_n\}$ la política markoviana tal que $f'_n(x) \in A(x)$ alcanza el mínimo en (3.17), es decir,

$$v'_n(x) = c^n(x, f'_n(x)) + \alpha \sum_{y\in\mathbb{X}} v'_{n-1}(y) P_{xy}(f'_n(x)). \tag{3.19}$$

Teorema 3.3.4 *Bajo la Hipótesis 2.2.1, la política π' es puntualmente ADO.*

Demostración.
La demostración de este Teorema es similar a la de los Teoremas anteriores, y por lo tanto suprimiremos algunos pasos.

Para cada $n \in \mathbb{N}_0$, definimos la función

$$\Phi'_n(x,a) := c^n(x,a) + \alpha \sum_{y\in\mathbb{X}} v'_{n-1}(y) P_{x,y}(a) - v'_n(x), \quad x \in \mathbb{X},$$

de tal forma que $\Phi'_n(x, f'_n(x)) = 0$ para toda $n \in \mathbb{N}_0$.

Además, es fácil ver que para cada $(x,a) \in \mathbb{K}$ y $n \in \mathbb{N}_0$

$$|\Phi(x,a) - \Phi'_n(x,a)| \leq |c(x,a) - c^n(x,a)| + \alpha \sum_{y\in\mathbb{X}} |V^*(x) - v'_{n-1}(y)| P_{x,y}(a) + |V^*(x) - v'_n(x)|.$$

Entonces, como $A(x)$ es finito, $c^n \nearrow c$ y $v'_n \nearrow V^*$ tenemos que

$$\sup_{a\in A(x)} |\Phi(x,a) - \Phi'_n(x,a)| \to 0 \quad \text{cuando } n \to \infty.$$

De aquí, para cada $x \in \mathbb{X}$, cuando $n \to \infty$,

$$\Phi(x, f'_n(x)) = |\Phi(x, f'_n(x)) - \Phi'_n(x, f'_n(x))| \leq \sup_{a\in A(x)} |\Phi(x,a) - \Phi'_n(x,a)| \to 0,$$

lo cual implica que π^* es puntualmente ADO. ■

3.4. Optimalidad en el límite

En esta sección estudiaremos el comportamiento limite de las sucesiones $\{\hat{f}_n\}$, $\{f_n^*\}$ y $\{f'_n\}$ que definen las políticas $\hat{\pi}$, π^* y π', respectivamente, estudiadas en la sección anterior. Específicamente demostraremos que el punto límite de estas sucesiones define una política óptima.

Antes de establecer el resultado principal, tenemos el siguiente resultado que garantiza la existencia de los puntos límite.

Lema 3.4.1 *Sea $\{f_n\}$ una sucesión de funciones en $\mathbb{F}$. Entonces, bajo la Hipótesis 2.2.1 (a), para cada $x \in \mathbb{X}$, existe una subsucesión $\{n_i(x)\} = \{n_i\} \subset \{n\}$ y $f_\infty \in \mathbb{F}$ tal que*

$$f_\infty(x) = \lim_{i\to\infty} f_{n_i}(x). \tag{3.20}$$

Demostración.
Sea $x \in \mathbb{X}$ fijo. Como $A(x)$ es un conjunto finito y $f_n(x) \in A(x)\ \forall\ n \in \mathbb{N}_0$, tenemos que $\{f_n(x)\}$ es una sucesión acotada, y por lo tanto contiene una subsucesión convergente. De aquí, existe una subsucesión $\{n_i(x)\} \subset \{n\}$ y $f_\infty \in \mathbb{F}$ tal que (3.20) se cumple. ■

A f_∞ se le conoce como *punto límite* o *punto de acumulación* de la sucesión $\{f_n\}$.

Sean $\hat{f}_\infty$, f_∞^* y f'_∞ los puntos límite de las sucesiones $\hat{\pi} = \{\hat{f}_n\}$, $\pi^* = \{f_n^*\}$ y $\pi' = \{f'_n\}$ respectivamente.

Teorema 3.4.2 *Bajo la Hipótesis 2.2.1, las políticas estacionarias $\hat{\pi}_\infty = \{\hat{f}_\infty\}$, $\pi^*_\infty = \{f^*_\infty\}$ y $\pi'_\infty = \{f'_\infty\}$ son óptimas.*

Demostración.
Las demostraciones de la optimalidad de cada una de las políticas son similares, por lo tanto solo demostraremos el caso de π'_∞ ya que sus argumentos cubren de cierta manera los usados en las otras demostraciones.

Observemos que de (3.19), para cada $x \in \mathbb{X}$,

$$v'_{n_i}(x) = c^{n_i}(x, f'_{n_i}(x)) + \alpha \sum_{y\in\mathbb{X}} v'_{n_i-1}(y) P_{xy}(f'_{n_i}(x)),\ i \in \mathbb{N}. \tag{3.21}$$

Supongamos por el momento que

$$\liminf_{i\to\infty} \sum_{y\in\mathbb{X}} v'_{n_i-1}(y)P_{xy}(f'_{n_i}(x)) \geq \sum_{y\in\mathbb{X}} V^*(y)P_{xy}(f'_\infty(x)). \tag{3.22}$$

Entonces, tomando lím inf cuando $i \to \infty$ en (3.21), y usando el hecho de que $c^{n_i} \nearrow c$ y $v'_{n_i} \nearrow V^*$, obtenemos

$$\begin{aligned} V^*(x) &= c(x, f'_\infty(x)) + \alpha \liminf_{i\to\infty} \sum_{y\in\mathbb{X}} v'_{n_i-1}(y)P_{x,y}(f'_{n_i}(x)) \\ &\geq c(x, f'_\infty(x)) + \alpha \sum_{y\in\mathbb{X}} V^*(y)P_{x,y}(f'_\infty(x)). \end{aligned}$$

Esto implica, por el Teorema 2.4.2, que

$$V^*(x) = c(x, f'_\infty(x)) + \alpha \sum_{y\in\mathbb{X}} V^*(y)P_{x,y}(f'_\infty(x)),$$

y por lo tanto $\pi'_\infty = \{f'_\infty\}$ es una política óptima.

Solo resta demostrar la desigualdad (3.22). Para esto, primero observemos

$$\begin{aligned} &|\sum_{y\in\mathbb{X}} v'_{n_i-1}(y)P_{x,y}(f'_{n_i}(x)) - \sum_{y\in\mathbb{X}} V^*(y)P_{x,y}(f'_{n_i}(x))| \\ &\quad\leq \sum_{y\in\mathbb{X}} |v'_{n_i-1}(y) - V^*(y)|P_{x,y}(f'_{n_i}(x)) \\ &\quad\leq \sup_{a\in A(x)} \sum_{y\in\mathbb{X}} |v'_{n_i-1}(y) - V^*(y)|P_{x,y}(a) \to 0 \quad \text{cuando } i \to \infty, \end{aligned} \tag{3.23}$$

donde la convergencia a cero se sigue del Teorema de Convergencia Dominada, porque $A(x)$ es finito y $v'_{n_i} \nearrow V^*$.

Ahora,

$$\begin{aligned} \sum_{y\in\mathbb{X}} v'_{n_i-1}(y)P_{x,y}(f'_{n_i}(x)) = &\sum_{y\in\mathbb{X}} v'_{n_i-1}(y)P_{x,y}(f'_{n_i}(x)) - \sum_{y\in\mathbb{X}} V^*(y)P_{x,y}(f'_{n_i}(x)) \\ &+ \sum_{y\in\mathbb{X}} V^*(y)P_{x,y}(f'_{n_i}(x)). \end{aligned}$$

Entonces, tomando lím inf cuando $i \to \infty$, por (3.23) obtenemos

$$\liminf_{i\to\infty} \sum_{y\in\mathbb{X}} v'_{n_i-1}(y)P_{x,y}(f'_{n_i}(x)) = \liminf_{i\to\infty} \sum_{y\in\mathbb{X}} V^*(y)P_{x,y}(f'_{n_i}(x)). \tag{3.24}$$

Además, por el Lema de Fatou C.0.14 (ver Apéndice C) y el hecho de que $A(x)$ es finito,

$$\liminf_{i\to\infty} \sum_{y\in\mathbb{X}} V^*(y) P_{x,y}(f'_{n_i}(x)) \geq \sum_{y\in\mathbb{X}} V^*(y) P_{x,y}(f'_{\infty}(x)). \tag{3.25}$$

Finalmente, combinando (3.24) y (3.25) demostramos (3.22). ■

Apéndice A

Variables Aleatorias Discretas

El material de este apéndice es muy estándar y puede ser consultado en muchos libros de probabilidad básica [ver [3]]. Para una fácil referencia en el trabajo, presentamos solo las definiciones y propiedades que hacen auto contenido al trabajo de tesis.

Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad. Denotaremos por $f_\xi(\cdot)$ a la función de probabilidad de una variable aleatoria (v.a.) discreta ξ:

$$f_\xi(t) = P[\xi = t] = P\{\omega \in \Omega : \xi(\omega) = t\}.$$

Un vector aleatorio $\bar{\xi}$ de dimensión n (n-vector) es de la forma $\bar{\xi} = (\xi_1, \xi_2, \ldots, \xi_n)$, donde ξ_i son v.a. definidas en $(\Omega, \mathcal{F}, P)$. Denotamos por $f_{\bar{\xi}} : \mathbb{R}^n \to [0, 1]$ a la función de probabilidad del n-vector $\bar{\xi}$:

$$f_{\bar{\xi}}(\mathbf{t}) = P[\xi_1 = t_1, \xi_2 = t_2, \ldots, \xi_n = t_n], \quad \mathbf{t} = (t_1, t_2, \ldots, t_n).$$

A.1. Esperanza de v.a.'s discretas

Definición A.1.1 *Sea ξ una v.a. discreta. Si se satisface al menos una de las condiciones siguientes:*

$$\sum_{t>0} t f_\xi(t) < \infty \quad o \quad \sum_{t<0} t f_\xi(t) > -\infty \tag{A.1}$$

entonces se define la esperanza (o valor esperado) de ξ como

$$E[\xi] := \sum_t t f_\xi(t). \tag{A.2}$$

Definición A.1.2 *Diremos que la v.a. ξ tiene esperanza finita si ambas condiciones en* (A.1) *se cumplen simultáneamente.*

Teorema A.1.3 *Sean $\bar{\xi}$ un n-vector aleatorio con función de probabilidad $f_{\bar{\xi}}$, y $h : \mathbb{R}^n \to \mathbb{R}$ una función arbitraria. Si la esperanza de la v.a. $Z = h(\bar{\xi})$ está bien definida, entonces*

$$E[Z] = \sum_{\mathbf{t}} h(\mathbf{t}) f_{\xi}(\mathbf{t}).$$

Teorema A.1.4 *Sean ξ_1 y ξ_2 dos v. a.'s con esperanza finita, y sea k una constante.*

(a) *Si $P[\xi_1 = k] = 1$, entonces $E[\xi_1] = k$.*

(b) *$E[k\xi_1] = kE[\xi_1] < \infty$.*

(c) *$E[\xi_1 + \xi_2] < \infty$ y además $E[\xi_1 + \xi_2] = E[\xi_1] + E[\xi_2]$.*

(d) *Si $P[\xi_1 \geq \xi_2] = 1$, entonces $E[\xi_1] \geq E[\xi_2]$.*

(e) *$|E[\xi_1]| \leq E[|\xi_1|]$.*

A.2. Esperanza condicional de v.a.'s discretas

Definición A.2.1 *Sean ξ_1 y ξ_2 dos v.a.'s discretas.*

(a) *Se define la función de probabilidad conjunta de ξ_1 y ξ_2, denotada por f_{ξ_1,ξ_2} ($f_{\xi_1,\xi_2} : \mathbb{R}^2 \to [0,1]$), como*

$$f_{\xi_1,\xi_2}(t,s) := P[\xi_1 = t, \xi_2 = s].$$

(b) *Se define la función de probabilidad condicional de ξ_2 dado ξ_1, denotada por $f_{\xi_2|\xi_1}$, como*

$$f_{\xi_2|\xi_1}(s|t) := P[\xi_2 = s|\xi_1 = t] = \frac{f_{(\xi_1,\xi_2)}(t,s)}{f_{\xi_1}(t)},$$

siempre que $f_{\xi_1}(t) > 0$.

Definición A.2.2 *Sean ξ_1 y ξ_2 dos v.a.'s discretas. Para $t \in \mathbb{R}$ (fijo) tal que $f_{\xi_1}(t) > 0$, se define la esperanza condicional de ξ_2 dado $\xi_1 = t$ por*

$$E[\xi_2|\xi_1 = t] := \sum_s s f_{\xi_2|\xi_1}(s|t).$$

Definición A.2.3 *La esperanza condicional de ξ_2 dado ξ_1 se define como*

$$E[\xi_2|\xi_1] := g(\xi_1),$$

donde

$$g(t) = E[\xi_2|\xi_1 = t].$$

Teorema A.2.4 *$E[\xi_2|\xi_1]$ tiene la propiedad de la doble esperanza, es decir,*

$$E[E[\xi_2|\xi_1]] = E[\xi_2].$$

A.3. Convergencia de v. a.'s discretas

Definición A.3.1 (Modos de convergencia) *Sean $X, X_1, X_2, \ldots$ variables aleatorias discretas en algún espacio de probabilidad (Ω, P). Decimos que:*

(a) *$X_n \to X$ **casi seguramente**, denotado por $X_n \xrightarrow{c.s.} X$, si*

$$\{\omega \in \Omega \ : \ X_n(\omega) \to X(\omega) \text{ cuando } n \to \infty\}$$

es un evento con probabilidad 1.

(b) *$X_n \to X$ **en r-ésima media**, denotado por $X_n \xrightarrow{r} X$, si $E[|X_n^r|] < \infty$ para toda $n \in \mathbb{N}$ y*

$$E[|X_n - X|^r] \to 0 \quad \text{cuando} \quad n \to \infty.$$

(c) *$X_n \to X$ **en probabilidad**, denotado por $X_n \xrightarrow{P} X$, si*

$$P(|X_n - X| > \epsilon) \to 0 \quad \text{cuando} \quad n \to \infty \quad \text{para toda} \quad \epsilon > 0.$$

(d) *$X_n \to X$ **en distribución**, denotado por $X_n \xrightarrow{D} X$, si*

$$P(X_n \leq x) \to P(X \leq x) \quad \text{cuando} \quad n \to \infty$$

para todo punto x en el cual la función $F_X(x) = P(X \leq x)$ es contínua.

Teorema A.3.2 *Las siguientes implicaciones se satisfacen:*

$$\begin{aligned}(X_n \xrightarrow{c.s.} X) &\Rightarrow (X_n \xrightarrow{P} X)\\ (X_n \xrightarrow{r} X) &\Rightarrow (X_n \xrightarrow{P} X)\\ (X_n \xrightarrow{P} X) &\Rightarrow (X_n \xrightarrow{D} X).\end{aligned}$$

para cualquier $r \geq 1$. *Además, si* $r > s \geq 1$ *entonces*

$$(X_n \xrightarrow{r} X) \Rightarrow (X_n \xrightarrow{s} X).$$

Ninguna otra implicación se cumple en general.

Apéndice B

Teorema del Punto Fijo

Definición B.0.3 *Un espacio normado X es un espacio vectorial con una norma definida en él. Una norma en un espacio vectorial (real o complejo) es una función $\|\cdot\| : X \to \mathbb{R}$, cuyo valor en $x \in X$ se denota por*

$$\|x\|,$$

y satisface las siguientes propiedades:

(i) $\|x\| \geq 0$

(ii) $\|x\| = 0$ *si y sólo si* $x = 0$

(iii) $\|\alpha x\| = |\alpha|\|x\|$

(iv) $\|x + y\| \leq \|x\| + \|y\|$.

donde $x, y, z \in X$ son vectores arbitrarios y α es cualquier escalar.

Definición B.0.4 *Un espacio métrico es una pareja (S, d), donde S es un conjunto no vacío, y $d : S \times S \to \mathbb{R}$ es una función tal que para $x, y, z \in S$ arbitrarios satisface las siguientes propiedades:*

(i) $d(x, x) = 0$

(ii) $d(x, y) > 0$ *si* $x \neq y$

(iii) $d(x, y) = d(y, x)$

(iv) $d(x, y) \leq d(x, z) + d(z, y)$ *(desigualdad del triángulo).*

A la función d se le conoce como métrica.

Teorema B.0.5 *Una norma en X define una métrica d en X dada por*

$$d(x,y) = \|x-y\| \qquad \forall\ x,y \in X$$

y es llamada la métrica inducida por la norma.

Demostración.
Veamos que satisface las 4 propiedades de una métrica.
Sean $x, y, z \in X$ arbitrarios.

(i) Por la propiedad (ii) en la Definición B.0.3

$$d(x,x) = \|x-x\| = \|0\| = 0$$

(ii) Si $x \neq y$ entonces $x - y \neq 0$, y por las propiedades (i) y (ii) en la Definición B.0.3

$$d(x,y) = \|x-y\| > 0.$$

(iii) De la propiedad (iii) en la Definición B.0.3 tenemos

$$\begin{aligned} d(x,y) &= \|x-y\| = \|(-1)(y-x)\| \\ &= |-1|\|y-x\| = \|y-x\| \\ &= d(y,x). \end{aligned}$$

(iv) Finalmente, de la propiedad (iv) en Definición B.0.3 se sigue que

$$\begin{aligned} d(x,y) &= \|x-y\| = \|(x-z)+(z-y)\| \\ &\leq \|x-z\| + \|z-y\| \\ &= d(x,z) + d(z,y). \quad \blacksquare \end{aligned}$$

Definición B.0.6 *Sea (S,d) un espacio métrico. Se dice que (S,d) es un espacio métrico completo si cualquier sucesión de Cauchy en S converge en S.*

Definición B.0.7 *Un espacio de Banach es un espacio normado completo con la métrica inducida por la norma.*

Teorema B.0.8 *El espacio de las funciones acotadas $B(X)$ es un espacio de Banach.*

Demostración.
Ver [14], Teorema 7.15, página 151.

Definición B.0.9 *Sea (S, d) un espacio métrico. Se dice que un operador*

$$T : S \to S$$

es de contracción módulo $\alpha \in (0, 1)$, si

$$d(Tx, Ty) \leq \alpha d(x, y) \quad \forall\ x, y \in S.$$

Teorema B.0.10 (Teorema del Punto Fijo de Banach) *Si (S,d) es un espacio métrico completo y $T : S \to S$ es un operador de contracción en S, entonces:*

(a) *Existe un único $x \in S$ tal que*

$$Tx = x.$$

(b) *Para cada $x_0 \in S$,*

$$\lim_{n\to\infty} T^n x_0 = x.$$

Demostración.

(a) El esquema de la demostración es el siguiente: primero construimos una sucesión $\{x_n\}$ en S y demostramos que es Cauchy, de esta manera converge en el espacio S; posteriormente demostramos que el límite x de la sucesión es un punto fijo de T, y finalmente demostramos que dicho punto es único.

Sea $x_0 \in S$ arbitrario, definimos la "sucesión iterativa" $\{x_n\}$ por

$$\begin{aligned} &x_0, \\ &x_1 = Tx_0 \\ &x_2 = Tx_1 = T^2 x_0 \\ &\vdots \\ &x_n = T^n x_0 \\ &\vdots \end{aligned} \tag{B.1}$$

La cual es la sucesión de imágenes de x_0 al aplicar el operador T repetidas veces. Veamos que es Cauchy.

Como T es un operador de contracción, existe $\alpha \in (0,1)$ tal que para todas $a, b \in S$

$$d(Ta, Tb) \leq \alpha d(a, b). \tag{B.2}$$

Luego, por (B.1) y (B.2)

$$\begin{aligned} d(x_{m+1}, x_m) &= d(Tx_m, T_{m-1}) \\ &\leq \alpha d(x_m, x_{m-1}) \\ &= d(Tx_{m-1}, Tx_{m-2}) \\ &\leq \alpha^2 d(x_{m-1}, x_{m-2}) \\ &\ldots \leq \alpha^m d(x_1, x_0). \end{aligned} \tag{B.3}$$

Entonces, por la desigualdad del triángulo y la fórmula para la suma de una sucesión geométrica, para $n > m$ obtenemos

$$\begin{aligned} d(x_m, x_n) &\leq d(x_m, x_{m+1}) + d(x_{m+1}, x_{m+2}) + \ldots + d(x_{n-1}, x_n) \\ &\leq (\alpha^m + \alpha^{m+1} + \ldots + \alpha^{n-1}) d(x_0, x_1) \\ &= \alpha^m \frac{1 - \alpha^{n-m}}{1 - \alpha} d(x_0, x_1). \end{aligned}$$

Como $\alpha \in (0,1)$, en el numerador tenemos $1 - \alpha^{n-m} < 1$. Así,

$$d(x_m, x_n) \leq \frac{\alpha^m}{1 - \alpha} d(x_0, x_1) \qquad (n > m). \tag{B.4}$$

En el lado derecho de la desigualdad (B.4) tenemos que $\alpha \in (0,1)$ y $d(x_0, x_1)$ son fijos, entonces podemos hacer el lado derecho de (B.4) tan pequeño como se quiera tomando m suficientemente grande (dado que $n > m$). Esto demuestra que $\{x_m\}$ es Cauchy, y como S es completo, se tiene que $\{x_m\}$ converge, digamos a x, es decir, $x_m \to x$.
Veamos que este límite x es un punto fijo de T.
Por la desigualdad del triángulo y (B.2) se tiene que

$$\begin{aligned} d(x, Tx) &\leq d(x, x_m) + d(x_m, Tx) \\ &\leq d(x, x_m) + \alpha d(x_{m-1}, x), \end{aligned}$$

y la última suma se puede hacer tan pequeña como cualquier $\epsilon > 0$ dado, pues $x_m \to x$. Entonces, concluimos que $d(x, Tx) = 0$, y de aquí que $Tx = x$. Esto demuestra que x es un punto fijo de T.
Además, x es el único punto fijo de T ya que si hubieran dos puntos fijos, $Tx = x$ y $T\bar{x} = \bar{x}$, por (B.2) obtenemos

$$d(x, \bar{x}) = d(Tx, T\bar{x}) \leq \alpha d(x, \bar{x}),$$

lo cual implica que $d(x, \bar{x}) = 0$ pues $\alpha \in (0,1)$, y así $x = \bar{x}$.

(b) En la demostración del inciso anterior probamos que para $x_0 \in S$ arbitrario, la sucesión formada por $\{x_n := T^n x_0\}$ converge al único punto fijo x. Por lo tanto $\lim_{n\to\infty} T^n y = x$. ■

Observación B.0.11 *Una consecuencia del teorema anterior es la siguiente:*

$$d(T^n x_0, x) \leq \alpha^n d(x_0, x), \quad \forall\ n \in \mathbb{N}. \tag{B.5}$$

En efecto, primero observemos que

$$d(Tx_0, x) = d(Tx_0, Tx) \leq \alpha d(x_0, x).$$

Ahora, supongamos que (B.5) se cumple para $n = k$, es decir,

$$d(T^k x_0, x) \leq \alpha^k d(x_0, x). \tag{B.6}$$

Entonces

$$\begin{aligned} d(T^{k+1} x_0, x) &= d(T(T^k x_0), Tx) \\ &\leq \alpha d(T^k x_0, Tx) \\ &\leq \alpha^{k+1} d(x_0, x), \end{aligned}$$

donde la última desigualdad es por (B.6). Esto demuestra (B.5).

Apéndice C

Teoremas de Convergencia

Supongamos que $\mathbb{Y}$ es un conjunto numerable.

Teorema C.0.12 *Sean $\{h_n\}$ una sucesión de funciones y B una función no-negativa definidas sobre $\mathbb{Y}$ que satisfacen*

$$|h_n(y)| \leq B(y), \quad y \in \mathbb{Y}. \tag{C.1}$$

Entonces, existe una subsucesión $\{n_k\}$, tal que la sucesión de funciones $\{h_{n_k}\}$ converge puntualmente, es decir,

$$\lim_{k\to\infty} h_{n_k}(x) \quad \textit{existe} \ \ \forall\ y \in \mathbb{Y}.$$

Demostración.
Suponga que $\mathbb{Y} = \{y_1, y_2, y_3, \ldots\}$. De (C.1), la sucesión de números reales $\{h_n(y_1)\}$ es acotada, y del Teorema de Bolzano-Weierstrass (TBW), sabemos que existe una subsucesión $\{m_k^1\}$ tal que $\{h_{m_k^1}(y_1)\}$ converge a un número real que denotamos por $h(y_1)$. Ahora considere la sucesión $\{h_{m_k^1}(y_2)\}$, la cual es acotada y de nuevo usando el TBW, existe una subsucesión $\{m_k^2\} \subset \{m_k^1\}$ tal que $\{h_{m_k^2}(y_2)\}$ resulta ser convergente a un real $h(y_2)$. Repitiendo este procedimiento se genera una familia de sucesiones "anidadas", es decir, $\{m_k^{t+1}\} \subset \{m_k^t\}$ con la característica

$$h_{m_k^t}(y_s) \to h(y_s) \quad \text{para} \ \ s = 1, 2, \ldots, t.$$

Finalmente, considere la sucesión formada por $n_k := m_k^k$ y note que $\{n_k\} \subset \{m_k^t\}$ para todo $t \in \mathbb{N}$. Por lo tanto, $\{h_{n_k}(y)\}$ es convergente para toda $y \in \mathbb{Y}$. ■

Teorema C.0.13 (Teorema de Convergencia Monótona) *Sea* $\{h_n\}$ *una sucesión de funciones sobre* $\mathbb{Y}$ *que satisfacen*

$$0 \leq h_n(\cdot) \leq h_{n+1}(\cdot) \quad \forall\ n \in \mathbb{N},$$

y π *una función no-negativa sobre* $\mathbb{Y}$*. Defina*

$$h(y) := \lim_{n\to\infty} h_n(y);$$

entonces,

$$\sum_{k=1}^{\infty} \pi(y_k)h(y_k) = \lim_{n\to\infty} \sum_{k=1}^{\infty} \pi(y_k)h_n(y_k)$$

Demostración.
Puesto que $h(\cdot) \geq h_n(\cdot)$, tenemos que

$$\sum_{k=1}^{\infty} \pi(y_k)h(y_k) \geq \sum_{k=1}^{\infty} \pi(y_k)h_n(y_k),$$

lo cual implica que

$$\sum_{k=1}^{\infty} \pi(y_k)h(y_k) \geq \lim_{n\to\infty} \sum_{k=1}^{\infty} \pi(y_k)h_n(y_k).$$

Por otra parte, como $h_n(\cdot) \geq 0$, note que

$$\begin{aligned}\lim_{n\to\infty} \sum_{k=1}^{\infty} \pi(y_k)h_n(y_k) &\geq \lim_{n\to\infty} \sum_{k=1}^{M} \pi(y_k)h_n(y_k) \\ &\geq \sum_{k=1}^{M} \pi(y_k)h(y_k),\end{aligned}$$

de donde se deduce que

$$\lim_{n\to\infty} \sum_{k=1}^{\infty} \pi(y_k)h_n(y_k) \geq \sum_{k=1}^{\infty} \pi(y_k)h(y_k).$$

Por lo tanto,

$$\sum_{k=1}^{\infty} \pi(y_k)h(y_k) = \lim_{n\to\infty} \sum_{k=1}^{\infty} \pi(y_k)h_n(y_k). \quad \blacksquare$$

Teorema C.0.14 (Lema de Fatou) *Sea* $\{h_n\}$ *una sucesión de funciones y* π *una función no-negativa definidas sobre* $\mathbb{Y}$. *Defina*

$$h(y) := \liminf_{n\to\infty} h_n(y) \quad y \in \mathbb{Y}.$$

(a) *Si existe una función* b *sobre* $\mathbb{Y}$ *tal que*

$$-\infty < \sum_{k=1}^{\infty} \pi(y_k) b(y_k)$$

y

$$-b(y) \leq h_n(y) \quad \forall\ y \in \mathbb{Y},\ n \geq 1,$$

entonces

$$\sum_{k=1}^{\infty} \pi(y_k) h(y_k) \leq \liminf_{n\to\infty} \sum_{k=1}^{\infty} \pi(y_k) h_n(y_k);$$

(b) *Si existe una función* B *sobre* $\mathbb{Y}$ *tal que*

$$\sum_{k=1}^{\infty} \pi(y_k) B(y_k) < \infty$$

y

$$h_n(y) \leq B(y) \quad \forall\ y \in \mathbb{Y},\ n \geq 1,$$

entonces

$$\limsup_{n\to\infty} \sum_{k=1}^{\infty} \pi(y_k) h_n(y_k) \leq \sum_{k=1}^{\infty} \pi(y_k) h(y_k).$$

Demostración.

(a) Defina

$$H_n(y) := \inf_{k\geq n} h_k(y) \quad y \in \mathbb{Y},$$

y observe que $H_n(\cdot)+b(\cdot)$ es no-negativa y converge crecientemente a la función $h(\cdot)+b(\cdot)$. Del Teorema de Convergencia Monótona (Teorema C.0.13), obtenemos

$$\sum_{k=1}^{\infty} \pi(y_k) h(y_k) = \lim_{n\to\infty} \sum_{k=1}^{\infty} \pi(y_k) H_n(y_k).$$

Por otro lado, puesto que $H_n(\cdot) \leq h_n(\cdot)\ \ \forall\ n$, concluimos que

$$\sum_{k=1}^{\infty} \pi(y_k) h(y_k) \leq \liminf_{n\to\infty} \sum_{k=1}^{\infty} \pi(y_k) h_n(y_k).$$

(b) Aplicando el resultado de la parte (a) a la sucesión $\{-h_n\}$ se concluye que

$$\sum_{k=1}^{\infty} \pi(y_k)[\liminf_{n\to\infty}(-h_n)(y_k)] \leq \liminf_{n\to\infty}[-\sum_{k=1}^{\infty} \pi(y_k)h_n(y_k)].$$

Usando el hecho de que para cualquier sucesión de números reales $\{a_n\}$ se cumple que

$$\liminf_{n\to\infty}(-a_n) = -\limsup_{n\to\infty} a_n,$$

se concluye que

$$\limsup_{n\to\infty} \sum_{k=1}^{\infty} \pi(y_k)h_n(y_k) \leq \sum_{k=1}^{\infty} \pi(y_k)h(y_k). \quad \blacksquare$$

Teorema C.0.15 (Teorema de la Convergencia Dominada) *Sea* $\{h_n\}$ *una sucesión de funciones,* h *y* M *funciones definidas sobre* $\mathbb{Y}$ *tales que*

$$h(y) = \lim_{n\to\infty} h(y) \quad \forall\ y \in \mathbb{Y},$$

$$|h_n(y)| \leq M(y) \quad \forall\ y \in \mathbb{Y},\ n \in \mathbb{N},$$

entonces,

$$\sum_{k=1}^{\infty} \pi(y_k)h(y_k) = \lim_{n\to\infty} \sum_{k=1}^{\infty} \pi(y_k)h_n(y_k).$$

Demostración.
Este resultado es una consecuencia inmediata del Lema de Fatou (Teorema C.0.14). $\blacksquare$

Bibliografía

[1] Bertsekas D.P.: Dynamic Programming and Stochastic Control. Academic Press, New York, 1976.

[2] Gordienko E.I., Minjárez-Sosa J.A.: Adaptive control for discrete-time Markov processes with unbounded costs: discounted criterion. Kybernetika 34 (1998), pp. 217-234.

[3] Grimmett G.R., Stirzaker D.R.: Probability and Random Processes. Oxford University Press Inc., New York, 2001.

[4] Hernández-Lerma O. : Adaptive Markov Control Processes. Springer-Verlag, New York, 1989.

[5] Hernández-Lerma O. : Lecture Notes on Discrete-Time Markov Processes. Departamento de Matemáticas, Centro de Investigación del IPN, México, D. F., 1990.

[6] Hernández-Lerma O., Lasserre J.B.: Discrete-Time Markov Control Processes: Basic Optimality Criteria. Springer-Verlag, New York, 1996.

[7] Hernández-Lerma O., Lasserre J.B.: Further Topics on Discrete-Time Markov Control Processes. Springer-Verlag, New York, 1999.

[8] Hilgert N., Minjárez-Sosa J.A.: Adaptive control of stochastic systems with unknown disturbance distribution: discounted criteria. Math. Methods Oper. Res., 63 (2006), pp. 443-460.

[9] Hilgert N., Minjárez-Sosa J.A.: Adaptive policies for time-varying stochastic systems under discounted criterion. Math. Methods Oper. Res., 54 (2001), pp. 491-505.

[10] Hoel P.G., Port S.C., Stone C.J.: Introduction to Probability Theory. Houghton Mifflin Company, Boston, 1971.

[11] Kreyszig E.: Introductory Functional Analysis with Applications. John Wiley & Sons. Inc., New York, 1978.

[12] Luque-Vásquez F., Minjárez-Sosa J.A., Vega-Amaya O.: Introducción a la Teoría de Control Estocástico. Departamento de Matemáticas, Universidad de Sonora, Hermosillo, Sonora, 1996.

[13] Minjárez-Sosa J.A.: Approximation and estimation in Markov control processes under discounted criterion. Kybernetika, 6, 40 (2004), pp. 681-690.

[14] Rudin W.: Principles of Mathematical Analysis. McGraw-Hill Inc., New York, 1976.