Procesos de Decisión de Markov Descontados

Gabriel Zacarías Espinoza

Índice general

Capítulo 1

INTRODUCCIÓN

La presente tesis está relacionada con la teoría de Procesos de Decisión de Markov (PDM) a tiempo discreto (ver [2], [3], [10] y [12]) . Un PDM es utilizado para modelar un sistema que es observado de forma discreta en el tiempo. En este caso suponemos que el sistema presenta inseguridad en su movimiento, es decir, se encuentra influenciado por un ruido aleatorio en su transición. El análisis que hacemos es para sistemas que son observados en periodos de tiempo finito e infinito.

Los PDM son aplicados en áreas como Economía, Biología, Ingeniería, etc. En el caso de Economía se aplican para optimizar problemas de crecimiento económico y en teoría de inventarios, por mencionar algunas aplicaciones (ver [13]). Actualmente se han aplicado a inteligencia artificial para modelar movimiento de objetos (ver [11]).

De manera general, un PDM se encarga de modelar un sistema dinámico cuyos estados son observados de manera periódica por un controlador. El desarrollo de un PDM, a través del tiempo, está dado de acuerdo al siguiente procedimiento. En cada tiempo t, $t = 0, 1, ...$, el controlador decide el control que aplicará dependiendo del estado del sistema. Entonces, como consecuencia del estado y de haber aplicado el control, se paga un costo, y el sistema, mediante la ley de transición, se traslada a un nuevo estado en el instante de tiempo $t + 1$. Al ocurrir un estado en $t + 1$, se repite el procedimiento anteriormente descrito.

A la sucesión de controles aplicados en cada tiempo se le llama **política**. Para evaluar la calidad de cada política se cuenta con un criterio de rendimiento o función objetivo.

En esta tesis nos enfocamos al criterio de rendimiento de **costo total**

descontado.

El **problema de control óptimo** consiste en encontrar una política que optimice el criterio de rendimiento. La política que optimiza el criterio de rendimiento se le llama **política óptima** y, al criterio de rendimiento evaluado en la política óptima le llamamos **función de valores óptimos**. Un procedimiento de solución para PDM está basado en el principio de Bellman conocido como **Programación Dinámica** (ver [4]).

La tesis se organiza de la forma siguiente. En el Capítulo 2 se presentan los conceptos generales de la teoría de PDM. En el Capítulo 3 nos enfocamos a problemas de control con horizonte finito y con el criterio de rendimiento costo total descontado. En este capítulo se presentan dos ejemplos. El primero de ellos es referente a la teoría de inventarios, este ejemplo es clásico en la teoría de PDM (ver [10]) y por tanto su solución es perfectamente conocida. El trabajo que se realizo con este ejemplo fue en la forma en que se presenta su solución, la cual utiliza algunos resultados de funciones convexas. Además, se verifican de forma concisa todas las condiciones de PDM para el ejemplo. El segundo ejemplo es un problema de reemplazamiento de máquinas tomado de Bertsekas (ver [2]), en particular, se presenta un caso donde se da su solución de forma explícita. Finalmente, en el Capítulo 4 se presenta la teoría referente a la teoría de PDM con costo descontado y con horizonte infinito se proporcionan dos ejemplos con costo cuadrático y se resuelven.

Capítulo 2

PROCESOS DE CONTROL DE MARKOV

El objetivo principal en este capítulo es introducir formalmente el concepto de Proceso de Decisión de Markov (PDM), presentados en la introducción.

2.1. Procesos de Control de Markov

Definición 2.1.1 *Un Modelo de Control de Markov (MCM), estacionario, a tiempo discreto, consiste de una quíntupla:*

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

donde:

a. X es un espacio de Borel no vacío (ver Apéndice A), llamado el espacio de estados;

b. A es un espacio de Borel no vacío, llamado el conjunto de acciones o controles;

c. $\{A(x) \mid x \in X\}$ es una familia de subconjuntos medibles, no vacíos $A(x)$ de A, donde $A(x)$ denota el conjunto de controles admisibles cuando el sistema se encuentra en el estado $x \in X$. El conjunto $\mathbb{K}$ de parejas de estados acciones admisibles, está definido por

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

y se supone que es un conjunto medible del espacio producto $X \times A$;

d. Q es un kérnel estocástico (ver Apéndice C) definido en X dado $\mathbb{K}$, llamado la ley de transición, es decir, para cada $(x,a) \in \mathbb{K}$, $Q(\cdot\,|x,a)$ es una medida de probabilidad en X, y para cada $B \subset X$, medible, $Q(B\,|\cdot)$ es una función medible;

e. $c : \mathbb{K} \to \mathbb{R}$ es una función medible y se llama la función de costo de un paso.

Podemos pensar en un MCM estacionario a tiempo discreto como un sistema estocástico controlado que se observa de manera periódica en los tiempos $t = 0, 1, 2....$ La dinámica que describe a este sistema estocástico funciona de la forma siguiente: si el sistema al tiempo t se encuentra en el estado $x_t = x \in X$, y la acción $a_t = a \in A(x)$ es aplicada, entonces ocurren dos cosas:

a) se paga un costo $c(x,a)$; y

b) el sistema se traslada a un nuevo estado x_{t+1}, mediante la distribución de probabilidad $Q(\cdot\,|x,a)$ sobre X, es decir,

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

$B \in \mathcal{B}(X)$, donde $\mathcal{B}(X)$ denota la σ-álgebra de Borel de X.

Una vez hecha esta transición a un nuevo estado, se elige una nueva acción y la dinámica anteriormente descrita se repite.

Suposición 2.1.2 *Supondremos que $\mathbb{K}$ contiene la gráfica de una función medible de X a A, es decir, existe $f : X \to A$ medible, tal que $f(x) \in A(x)$, para toda $x \in X$. Al conjunto de estas funciones es denotado por $\mathbb{F}$ y sus elementos son llamados selectores de la multifunción $x \to A(x)$ (ver Apéndice D).*

Políticas de control. Para introducir el concepto de estrategia o política, considérese un MCM y defina $\mathbb{H}_t$, *el espacio de las historias observadas del proceso de control hasta el tiempo t*, como

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

para $t = 1, 2, ...$ Un elemento h_t de $\mathbb{H}_t$ llamado t-historia es un vector de la forma

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

donde $(x_i, a_i) \in \mathbb{K}$ para $i = 0, ..., t-1$ y $x_t \in X$.

Obsérvese que, para cada t, $\mathbb{H}_t$ es un subespacio de $\mathbf{H}_t := (X \times A)^t \times X$ y $\mathbf{H_0} := X$.

Definición 2.1.3 *Una política aleatorizada o simplemente política, es una sucesión $\pi = \{\pi_t, t = 0, 1, 2, ...\}$ de kérneles estocásticos definidas sobre A dado la historia del proceso $\mathbb{H}_t$ y satisface que: $\pi_t(\, A(x_t) |\, h_t) = 1$ para toda $h_t \in \mathbb{H}_t$ y $t = 0, 1, 2....$*

Al conjunto de todas las políticas se denotará por Π.

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

Denotemos a la familia de kérneles estocásticos sobre A dado X, como $P(\, A |\, X)$.

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

Observación 2.1.4 *Por la suposición 2.1.2 tenemos que $\mathbb{F} \subset \Phi$*

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

Markoviana Aleatorizada (Π_{RM})*. Si existe una sucesión $\{\varphi_t\}$ de kérneles estocásticos con $\varphi_t \in \Phi$ (definidas sobre A dado X), tales que, $\pi_t(\cdot |\, h_t) = \varphi_t(\cdot |\, x_t)$ para toda $h_t \in \mathbb{H}_t$ y $t = 0, 1, 2....$*

Markoviana Aleatorizada Estacionaria (Π_{RS})*. Si existe $\varphi \in \Phi$ kérnel estocástico, tal que: $\pi_t(\cdot |\, h_t) = \varphi(\cdot |\, x_t)$ para toda $h_t \in \mathbb{H}_t$ y $t = 0, 1, 2....$*

Determinista (Π_D)*. Si existe una sucesión $\{g_t\}$ de funciones medibles con $g_t : \mathbb{H}_t \to A$, tales que, para cada $h_t \in \mathbb{H}_t$ y $t = 0, 1, 2, ...$, se tiene que $g_t(h_t) \in A(x_t)$ y $\pi_t(\cdot |\, h_t)$ está concentrada en $g_t(h_t)$.*

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

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

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

Observación 2.1.6 *Notemos que, $\Pi_{RS} \subset \Pi_{RM} \subset \Pi$ y $\Pi_{DS} \subset \Pi_{DM} \subset \Pi_D \subset \Pi$.*

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

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

$$P_x^\pi(\, a_t \in C|\, h_t) = \pi_t(C|\, h_t), \tag{2.1}$$

$$P_x^\pi(\, x_{t+1} \in B|\, h_t, a_t) = Q(\, B|\, x_t, a_t). \tag{2.2}$$

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

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

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

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

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

El MCM descrito aquí, es llamado estacionario porque sus componentes no dependen del parámetro tiempo t, en caso de que esto suceda se dice ser un

modelo no estacionario, es decir, un modelo de la forma $(X_t, A_t, \{A_t(x) | x \in X_t\}, Q_t, C_t)$ para $t = 0, 1, 2, ...$

La suposición 2.1.2 asegura que $\mathbb{F}$ es no vacío y por lo tanto Π también. Esto se debe a que cada $f \in \mathbb{F}$ puede ser identificada por un kérnel estocástico φ, de la siguiente forma: $\varphi(C| x) = I_C(f(x))$ para toda $C \in \mathcal{B}(A)$ y $x \in X$.

Criterio de Rendimiento. Cada PDM estará dotado de una función real, llamada función objetivo o criterio de rendimiento, que medirá en algún sentido la calidad de cada política, a través de la sucesión de costos que genera.

Consideremos un modelo de control de Markov fijo y un conjunto de políticas Π. Definimos para cada $x \in X$ y $\pi \in \Pi$

$$V(\pi, x) := E_x^\pi \left[\sum_{t=o}^{N} c(x_t, a_t) \right].$$

$V(\pi, x)$ es conocido como el **Costo Total Acumulado.** Al entero positivo N se le conoce como horizonte del problema, el cual representa el número de etapas en el cual el sistema esta operando y puede ser finito o infinito.

Costo total descontado: para $\pi \in \Pi$ y $x \in X$ definimos

$$V(\pi, x) := E_x^\pi \left[\sum_{t=o}^{N} \alpha^t c(x_t, a_t) \right],$$

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

Definición 2.1.8 *Para cada $x \in X$ definimos*

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

V^* se le llama función de valores óptimos o valor óptimo.

Definición 2.1.9 *Una política $\pi^* \in \Pi$, es óptima, si*

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

$x \in X$.

El *problema de control óptimo* consiste en encontrar una política óptima, es decir, minimizar la función $\pi \to V(\pi, x)$ sobre Π, para toda $x \in X$.

Propiedad de Markov

A continuación recordamos la definición de un proceso de Markov a tiempo discreto y se mostrará que cuando usamos una política de Markov el resultado es un proceso de control de Markov. Definimos para $x \in X$ y $\varphi \in \Phi$

$$c(x, \varphi) := \int_A c(x, a)\varphi(\, da|\, x),$$

$$Q(\cdot|\, x, \varphi) := \int_A Q(\cdot|\, x, a)\varphi(\, da|\, x). \tag{2.3}$$

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

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

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

Definición 2.1.10 *Sea $\{R_t\}$ una sucesión de kérneles estocásticos en $P(X|\, X)$, y sea $\{x_t\}$ un proceso estocástico sobre X. Entonces a $\{x_t\}$ se le llama proceso de Markov no homogéneo con kérneles de transición $\{R_t\}$, si para cada $B \in \mathcal{B}(X)$ y $t = 0, 1, ...,$ se tiene que*

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

A la primera igualdad se le conoce como la propiedad de Markov

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

Proposición 2.1.11 *Sea v una distribución inicial arbitraria y $\pi = \{\varphi_t\} \in \Pi_{RM}$, entonces $\{x_t\}$ es un proceso de Markov no homogéneo con kérneles de transición $\{Q(\cdot|\, \cdot, \varphi_t)\}$, esto es, para cada $B \in \mathcal{B}(X)$ y $t = 0, 1, 2, ...,$*

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

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

Demostración. Primero demostraremos que para cualquier $\pi = \{\pi_t\} \in \Pi$ se tiene que

$$P_v^\pi(x_{t+1} \in B|\, h_t) = \int_A Q(B|\, x_t, a_t)\pi_t(da_t|\, h_t),$$

para toda $B \in \mathcal{B}(X)$ y $t = 0, 1, ...$

En efecto por las propiedades de Esperanza Condicional (ver Apéndice B.1(e)), tenemos que

$$\begin{aligned} P_v^\pi(x_{t+1} \in B|\, h_t) &= E_v^\pi\left[P_v^\pi(x_{t+1} \in B|\, h_t, a_t)|\, h_t\right], \\ &= E_v^\pi\left[Q(x_{t+1} \in B|\, x_t, a_t)|\, h_t\right], \text{ (por (2.2))} \\ &= \int_A Q(B|\, x_t, a_t)\pi_t(da_t|\, h_t), \text{ (por (2.1)).} \end{aligned}$$

En particular si $\pi = \{\varphi_t\} \in \Pi_{RM}$ entonces

$$\begin{aligned} P_v^\pi(x_{t+1} \in B|\, h_t) &= \int_A Q(B|\, x_t, a_t)\varphi_t(da_t|\, x_t), \\ &= Q(B|\, x_t, \varphi_t), (\text{por}(2.3)). \end{aligned}$$

Así, usando de nuevo a B.1.(e) y a B.1.(d) (ver Apéndice B), implica que

$$\begin{aligned} P_v^\pi(x_{t+1} \in B|\, x_0, x_1, ..., x_t) &= E_v^\pi\left[P_v^\pi(x_{t+1} \in B|\, h_t)|\, x_0, x_1, ..., x_t\right], \\ &= E_v^\pi\left[Q(B|\, x_t, \varphi_t)|\, x_0, x_1, .., x_t\right], \\ &= Q(B|\, x_t, \varphi_t). \end{aligned}$$

Similarmente,

$$\begin{aligned} P_v^\pi(x_{t+1} \in B|\, x_t) &= E_v^\pi\left[P_v^\pi(x_{t+1} \in B|\, h_t)|\, x_t\right], \\ &= E_v^\pi\left[Q(B|\, x_t, \varphi_t)|\, x_t\right], \\ &= Q(B|\, x_t, \varphi_t), \end{aligned}$$

y esto muestra la proposición. ■

Observación 2.1.12 *Si $Q(\cdot|\, \cdot, \varphi)$ es un kérnel de transición como en la proposición anterior, con $\varphi \in \Phi$. Denotamos las probabilidades de transición en n-etapas como $Q^n(\cdot|\, \cdot, \varphi)$, donde tenemos para $n \geq 1$*

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

Capítulo 3

PROBLEMAS CON HORIZONTE FINITO

Consideremos el criterio de rendimiento costo total acumulado con horizonte finito, es decir,

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

donde c_N es una función real valuada definida sobre X llamada la función de costo terminal.

El objetivo de este capitulo es proveer una técnica de solución para el problema de control óptimo, conocida como Programación Dinámica. Dicha técnica nos permite encontrar tanto a la función de valor óptimo V^*, como a la política óptima π^*.

El siguiente teorema impone una condición de medibilidad y la existencia de selectores (ver Apéndice D), mediante la Ecuación de Programación Dinámica (EPD).

3.1. Programación Dinámica

Teorema 3.1.1 *Sean $V_0, V_1, ..., V_N$ funciones sobre X definidas por*

$$V_N(x) := c_N(x), \tag{3.1}$$

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

$$V_t(x) := \min_{a \in A(x)} \left[c(x,a) + \int_X V_{t+1}(y) Q(\,dy|\, x, a) \right]. \tag{3.2}$$

Supongamos que estas funciones son medibles y que para cada $t = 0, 1, 2, ..., N-1$*, existe un selector* $f_t \in \mathbb{F}$ *con* $f_t(x) \in A(x)$*, tal que*

$$V_t(x) = c(x, f_t(x)) + \int_X V_{t+1}(y) Q(\,dy|\, x, f_t(x)).$$

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

La relación (3.2) es conocida como Ecuación de Programación Dinámica (EPD) junto con su condición inicial (3.1) y su nombre es debido a Bellman (ver [4]).

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

$$C_t(\pi, x) := E^{\pi} \left[\sum_{n=t}^{N-1} c(x_n, a_n) + c_N(x_N) \middle| x_t = x \right],$$

$$C_N(\pi, x) := c_N(x),$$

$t = 0, 1, ...N-1$. $C_t(\pi, x)$ es llamado el costo total del instante t a $N-1$ cuando usamos la política π y $x_t = x$. En particular notemos que

$$V(\pi, x) = C_0(\pi, x).$$

Para demostrar este teorema, tenemos que mostrar que para todo $x \in X$ y $t = 0, 1, ..., N-1$, se tiene que

$$C_t(\pi, x) \geq V_t(\pi, x),$$

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

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

En particular, si $t = 0$ tenemos que

$$C_0(\pi, x) = V(\pi, x) \geq V_0(x)$$

y si utilizamos a π^*:

$$C_0(\pi^*, x) = V(\pi^*, x) = V^*(x) = V_0(x).$$

Para probar las relaciones anteriores, primero observemos que

$$C_N(\pi, x) = V_N(x) = c_N(x).$$

Ahora, supongamos que para alguna $t \in \{0, 1, ...N-1\}$ se tiene que

$$C_{t+1}(\pi, x) \geq V_{t+1}(x).$$

Entonces por (2.1) y (2.2),

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

por el Apéndice B.1(e) se llega a que

$$E^\pi \left[\sum_{n=t+1}^{N-1} c(x_n, a_n) + c_N(x_N) \middle| \, x_t = x \right] =$$

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

Por otro lado, observemos que

$$E^\pi \left[\sum_{n=t+1}^{N-1} c(x_n, a_n) + c_N(x_N) | \, x_{t+1} = y \right] = C_{t+1}(\pi, x).$$

Así,

$$\begin{aligned} C_t(\pi,x) &= \int_A \left[c(x,a) + \int_X C_{t+1}(\pi,y) Q(\,dy|\,x,a) \right] \pi_t(\,da|\,x), \qquad (3.3) \\ &\geq \int_A \left[c(x,a) + \int_X V_{t+1}(y) Q(\,dy|\,x,a) \right] \pi_t(\,da|\,x), \\ &\geq \min_{a\in A(x)} \left[c(x,a) + \int_X V_{t+1}(y) Q(\,dy|\,x,a) \right] = V_t(x), \end{aligned}$$

y bajo la hipótesis de inducción se tiene que,

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

como se deseaba. ■

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

$$V_t(x) = \inf_{\pi\in\Pi} C_t(\pi,x),$$

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

Así, hemos calculado para cada tiempo t el costo óptimo de t en adelante, con esta interpretación de V_t es posible caracterizar a la EPD.

En efecto, sea $\pi = \{\pi_t, ..., \pi_{N-1}\}$ un política tal que $\pi_t = f \in \mathbb{F}$ es un selector arbitrario y $\{\pi_{t+1}, ..., \pi_{N-1}\}$ es una política óptima para el problema de $t+1$ hasta N, entonces por (3.3)

$$\begin{aligned} C_t(\pi,x) &= c(x,f) + \int V_{t+1}(y) Q(\,dy|\,x,f) \\ &\geq \min_{a\in A(x)} \left[c(x,a) + \int_X V_{t+1}(y) Q(\,dy|\,x,a) \right], \end{aligned}$$

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

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

para ver la desigualdad inversa nótese que

$$V_t(x) \leq c(x,f) + \int V_{t+1}(y) Q(\,dy|\,x,f),$$

lo cual implica que

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

lo cual prueba lo anteriormente afirmado.

3.2. Condición de Selección Medible

El teorema de Programación Dinámica tiene como suposición la existencia de selectores $f \in \mathbb{F}$, los cuales minimizan el lado derecho de la EPD en cada etapa. Esta suposición es referida como condición de selección medible.

Suposición 3.2.1 ***Condición de selección Medible***

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

$$u^*(x) := \inf_{a \in A(X)} \left\{ c(x,a) + \int_X u(y) Q(\,dy|\,x,a) \right\},$$

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

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

En conclusión, si esta suposición ocurre entonces se puede cambiar ínfimo por mínimo.

En muchos problemas la suposición anterior se puede verificar directamente, pero desde un punto de vista teórico, es conveniente tener condiciones generales, estas condiciones se obtienen generalmente de los teoremas de selección medible, dos resultados importantes se mencionan en el Apéndice D.

Definición 3.2.2 *Sean (X, d) un espacio métrico y $v : X \to \mathbb{R} \cup \{+\infty\}$ una función tal que $v(x) < \infty$ para al menos una $x \in X$, diremos que la función v es semicontinua inferiormente (l.s.c) en x, si*

$$\liminf_{n\to\infty} v(x_n) \geq v(x),$$

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

Si v es l.s.c para toda $x \in X$, diremos que es semicontinua inferiormente (l.s.c).

La función v es semicontinua superiormente (u.s.c), si $-v$ es l.s.c.

Obsérvese que v es continua, si y solo si, es l.s.c y u.s.c.

Definición 3.2.3 *Un kérnel estocástico $P \in P(X|\, A)$ es:*

a) Débilmente continuo, si la función $v'(a) := \int_X v(x)P(dx|\, a)$ es continua y acotada en A para cada v función continua y acotada sobre X

b) Fuertemente continuo, si la función $v'(a) := \int_X v(x)P(dx|\, a)$ es continua y acotada en A para cada v función medible y acotada sobre X

Acerca de la continuidad débil o fuerte de kérneles estocásticos, vea el Apéndice C.

Definición 3.2.4 *Una función $v : \mathbb{K} \to \mathbb{R}$ se llama inf-compacta sobre $\mathbb{K}$, si para toda $x \in X$ y $\lambda \in \mathbb{R}$, el conjunto $\{a \in A(x)|\, v(x, a) \leq \lambda\}$ es compacto.*

Condición 3.2.5 *a) La función de costo c es l.s.c, acotada inferiormente e inf-compacta sobre $\mathbb{K}$.*

b) La ley de transición Q es:

b1) Débilmente continua, es decir, $v'(x, a) := \int_X v(y)Q(dy|\, x, a)$ es continua y acotada en $\mathbb{K}$ para cada v función continua y acotada sobre X, ó

b2) Fuertemente continua, es decir, $v'(x, a) := \int_X v(y)Q(dy|\, x, a)$ es continua y acotada en $\mathbb{K}$ para cada v función medible y acotada sobre X.

Teorema 3.2.6 *Bajo la condición 3.2.5 se tiene que, para cualquier función $u : X \to \mathbb{R}$ medible y no negativa, la condición de selección medible se cumple.*

Demostración. Para una demostración consultar [10]. ■

Observación 3.2.7 *En el teorema anterior se supone que v es no negativa pero sólo es necesario suponer que v es acotada inferiormente.*

Existen otras condiciones para que la condición de selección medible ocurra, por ejemplo:

Condición 3.2.8 *a) $A(x)$ es compacto para todo $x \in X$;*

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

c) la función $v'(x,a) := \int_X v(y)Q(dy|\, x,a)$ sobre $\mathbb{K}$, satisface alguna de las siguientes condiciones

c1) $v'(x,\cdot)$ es (l.s.c) en $A(x)$ para cada $x \in X$ y cada función v continua y acotada sobre X, ó

c2) $v'(x,\cdot)$ es (l.s.c) en $A(x)$ para cada $x \in X$ y cada función v medible y acotada sobre X.

Condición 3.2.9 *a) $A(x)$ es compacto para toda $x \in X$;*

b) La función de costo c es l.s.c y acotada inferiormente;

c) La ley de transición Q es:

c1)Débilmente continua, es decir, $v'(x,a) := \int_X v(y)Q(dy|\, x,a)$ es continua y acotada en $\mathbb{K}$ para cada v función continua y acotada sobre X, ó

c2)Fuertemente continua, es decir, $v'(x,a) := \int_X v(y)Q(dy|\, x,a)$ es continua y acotada en $\mathbb{K}$ para cada v función medible y acotada sobre X.

Teorema 3.2.10 *Para cualquier función $u : X \to \mathbb{R}$ medible, no negativa, las condiciones 3.2.8 y 3.2.9 implican la condición de selección medible. Además, bajo 3.2.8(c1) y 3.2.9(c1) es suficiente tomar a u como no negativa y l.s.c, y bajo 3.2.9(a,b,c2) la función u^* es l.s.c.*

Demostración. Sea $u \geq 0$ una función medible sobre X, es suficiente considerar la condición 3.2.8 con los incisos a, b y c2. Por otro, lado la suma de dos funciones l.s.c es de nuevo una función l.s.c, (ver Apéndice A3). Si probamos que la función

$$a \to \int u(y)Q(dy|\, x,a),$$

es l.s.c sobre $A(X)$ para toda $x \in X$, entonces usando la Proposición 5.4.3 D3 del Apéndice D tendremos el resultado.

Para esto, sea $\{u_n\}$ una sucesión de funciones medibles y acotadas sobre X tales que $u_n \uparrow u$ y sea $\{a_k\}$ una sucesión convergente a $a \in A(x)$. Entonces, para toda n

$$\begin{aligned} \liminf_{k\to\infty} \int u(y)Q(\,dy|\,x,a^k) &\geq \liminf_{k\to\infty} \int u_n(y)Q(\,dy|\,x,a^k), \\ &\geq \int u_n(y)Q(\,dy|\,x,a), \end{aligned}$$

la última desigualdad es por (c2).

Ahora, si $n \to \infty$, por el Teorema de la convergencia monótona tenemos

$$\liminf_{k\to\infty} \int u(y)Q(\,dy|\,x,a^k) \geq \int u(y)Q(\,dy|\,x,a),$$

y esto prueba lo deseado.

Con un argumento similar, se prueba el teorema, si suponemos 3.2.8(c1) ó 3.2.9(c2).

Si $u \geq 0$ y l.s.c entonces existe una sucesión creciente de funciones continuas y acotadas (ver Apéndice A2) y se sigue el resultado como antes. Ahora si suponemos la condición 3.2.9(a,b,c2) entonces por la Proposición 5.4.3 D3 (b) del Apéndice D se tiene que u^* es l.s.c. ■

Observación 3.2.11 *Los dos teoremas anteriores proporcionan condiciones suficientes para que la condición de selección medible ocurra, y a su vez, se aplica en forma recursiva, en el algoritmo de Programación Dinámica.*

Un modelo de Markov que satisface la condición 3.2.8 o 3.2.9, se le conoce como modelo semicontinuo, y si el modelo satisface la condición 3.2.5, se le conoce como modelo semicontinuo-semicompacto.

3.3. Variantes de la Ecuación de Programación Dinámica (EPD)

A menudo es conveniente escribir la EPD en otras apropiadas y equivalentes formas.

3.3.1. Modelo de Ecuación de Diferencias

En algunas aplicaciones la ley de transición Q es inducida por una ecuación en diferencias estocásticas de la forma siguiente:

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

para $t = 1, 2, ...$, $x_0 = x$ conocido. Donde $\{\xi_t\}$ es una sucesión de variables aleatorias (v.a) independientes e idénticamente distribuidas (i.i.d) tomando valores en un espacio S con distribución común μ e independiente del estado x_0 y $F : \mathbb{K} \times S \to X$ es una función medible conocida. En este caso la ley de transición Q está dada por:

$$\begin{aligned} Q(B|\, x, a) &= \Pr(x_{t+1} \in B|\, x_t = x, a_t = a), \\ &= \Pr(F(x_t, a_t, \xi_t) \in B|\, x_t = x, a_t = a), \\ &= \Pr(F(x, a, \xi_t) \in B), \\ &= \mu\{s|\, F(x, a, s) \in B\}, \\ &= \int_S I_B\left(F(x, a, s)\right) \mu(ds), \\ &= E\left[I_B(F(x, a, \xi))\right]. \end{aligned}$$

$B \in \mathcal{B}(X)$, $(x, a) \in \mathbb{K}$, donde I_B es la función indicadora del conjunto B. Obsérvese que cuando la dinámica es determinista, es decir, $x_{t+1} = G(x_t, a_t)$ con $G : \mathbb{K} \to X$, la ley de transición es:

$$Q(B|\, x, a) = I_B(G(x, a)), B \in \mathbb{B}(X).$$

Si $\{\xi_t\}$ tiene densidad común Δ, se tiene que

$$Q(B|\, x, a) = \int I_B(L(x, a, s))\Delta(s)ds.$$

Por el teorema de cambio de variable, para funciones medibles sobre X,

tenemos que si v es una función medible sobre X, entonces

$$\begin{aligned} E\left[v(x_{t+1})|\, x_t = x, a_t = a\right] &= \int_X v(y)Q(\,dy|\, x, a), \\ &= \int_S v(F(x,a,s))\mu(ds), \\ &= \int v(F(x,a,s))\Delta(s)ds, \\ &= E\left[v(F(x,a,\xi)\right]. \end{aligned}$$

Así, la EPD queda como

$$V_N(x) := c_N(x),$$

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

$$\begin{aligned} V_t(x) &= \min_{a\in A(x)} \left\{ c(x,a) + \int V_{t+1}(y)Q(\,dy|\, x, a) \right\}, \\ &= \min_{a\in A(x)} \left\{ c(x,a) + E\left[V_{t+1}(F(x,a,\xi))\right] \right\}. \end{aligned}$$

Además, si el costo depende de forma explicita de la v.a ξ, es decir, $c(x,a,\xi)$, entonces

$$E[c(x,a,\xi_t) + V_{t+1}(F(x,a,\xi_t))] = \int_S \left[c(x,a,s) + V_{t+1}((F(x,a,s))\right]\mu(ds).$$

3.3.2. Forma Hacia Adelante de la EPD

Para $t = 0, ..., N$, definimos a $v_t := V_{N-t}$. Entonces escribimos la EPD hacia adelante como

$$v_0(x) = c_N(x),$$

y para $t = 1, ..., N$

$$v_t(x) = \min_{a\in A(x)} \left\{ c(x,a) + \int_X v_{t-1}(y)Q(\,dy|\, x, a) \right\},$$

Además, si $f_t \in \mathbb{F}$ es el selector que minimiza el lado derecho de la EPD, entonces $g_t := f_{N-t}$ para $t = 1, ..., N$, es un minimizador. Así, en términos

de v_t, el teorema de programación dinámica dice que $\widetilde{\pi} = \{g_N, g_{N-1}, ...g_1\}$ es una política óptima y la función de valor óptimo es

$$V^*(\cdot) = v_N(\cdot) = V(\widetilde{\pi}, \cdot),$$

es decir, para todo $x \in X$

$$v_N(x) = \inf_{\pi \in \Pi} V(\pi, x).$$

Las funciones v_t son llamadas funciones de iteración de valores (IV).

3.3.3. Costo Total Descontado

Recordemos que un modelo puede ser no estacionario, es decir, para $t = 0, 1, ...$

$$(X_t.A_t.\{\, A_t(x) |\, x \epsilon X_t\}, Q_t, C_t),$$

Consideremos el criterio de rendimiento de costo total descontado

$$V(\pi, x) := E_x^{\pi} \left[\sum_{t=o}^{N-1} \alpha^t c(x_t, a_t) + \alpha^N c_N(x_N) \right],$$

con $0 < \alpha < 1$. A c_N se le llama costo terminal.

Nótese, que si $C_t(x, a) := \alpha^t c(x, a)$. Entonces con X, A, y Q fijos y C_t como se definió anteriormente, tenemos un modelo no estacionario, cuya EPD es de la forma

$$V_N(x) := \alpha^N c_N(x),$$

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

$$V_t(x) := \min_{a \in A(x)} \left[\alpha^t c(x, a) + \int_X V_{t+1}(y) Q(\, dy |\, x, a) \right].$$

Si definimos a $J_t(\cdot) = \alpha^{-t} V_t(\cdot)$, con $t = 0, ...N$, entonces

$$J_N(x) := c_N(x),$$

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

$$J_t(x) := \min_{a \in A(x)} \left[c(x, a) + \alpha \int_X J_{t+1}(y) Q(\, dy |\, x, a) \right],$$

así el teorema de Programación Dinámica se cumple para la funciones J_t.

3.4. Ejemplos

Sistema de Inventarios

Consideramos un sistema de inventario de alguna producción en el cual la variable x_t es la cantidad de productos al principio del periodo t ($t = 0, 1, 2...$). La variable acción a_t es la cantidad pedida (u ordenada) e inmediatamente proporcionada al principio del período t, y la variable de perturbación (o ruido) ξ_t es la demanda durante ese período, supondremos que la sucesión $\{\xi_t\}$ de variables aleatorias son i.i.d.

En este caso la dinámica del sistema está dada por la siguiente ecuación en diferencias

$$x_{t+1} = x_t + a_t - \xi_t, \ t = 0, 1, 2, ...$$

La función de costo esta dada por

$$\begin{aligned} c(x_t, a_t, \xi_t) &= ba_t + h\max\{0, x_{t+1}\} + p\max\{0, -x_{t+1}\}, \\ &= ba_t + h\max\{0, x_t + a_t - \xi_t\} + p\max\{0, \xi_t - x_t - a_t\}, \end{aligned}$$

para $t = 0, 1, 2, ...$

Donde:

b es el costo de producción por unidad.
h es el costo por unidad de exceso del inventario (de almacenaje).
p es el costo por unidad de la demanda sin entregar.

Además b, h y p son constantes no negativas tales que $p \geq b$. Consideremos el espacio de estados X como el conjunto de los números reales, es decir, $X = \mathbb{R}$, el espacio de acciones $A = A(x) = [0, \infty)$, y que $\{\xi_t\}$ son v.a's no negativas con media finita definidas en $S = [0, \infty)$, bajo el supuesto de i.i.d, suponemos que μ y Δ es su distribución de probabilidad y su densidad común, respectivamente.

Deseamos minimizar el costo esperado descontado de operación

$$V(\pi, x) := E_x^{\pi}\left[\sum_{t=o}^{N-1} \alpha^t c(x_t, a_t, \xi_t)\right], \pi \in \Pi, x \in X.$$

Veamos que la función de costo c es, l.s.c, acotada inferiormente e inf-compacta sobre $\mathbb{K}$.

Es obvio que c esta acotada inferiormente por cero. Para ver que c es l.s.c, demostraremos que es continua, notemos que nuestro modelo esta determinada por una ecuación en diferencias y que el costo c depende de x, a, y ξ,

así podemos definir

$$\begin{aligned} c'(x,a) &= E(\, c(x_t, a_t, \xi_t)|\, x_t = x, a_t = a) \\ &= \int c(x,a,s)\mu(ds) \\ &= \int c(x,a,s)\Delta(s)ds, \end{aligned}$$

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

Así,

$$c'(x,a) = ba + hE[\max\{0, x+a-\xi\}] + pE[\max\{0, \xi - x - a\}],$$

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

Lema 3.4.1 *Bajo las condiciones antes mencionadas se satisfacen las condiciones 3.2.5.*

Demostración. Mostraremos que c' es continua en $\mathbb{K}$, para ello basta probar que

$$H(x,a) := hE[\max\{0, x+a-\xi\}] + pE[\max\{0, \xi - x - a\}],$$

es continua para cada $(x,a) \in \mathbb{K}$, dado que ba lo es. Para esto recordemos que

$$\max(x,y) = \frac{x+y+|x-y|}{2}, x, y \in \mathbb{R}.$$

Entonces,

$$\begin{aligned} H(x,a) &= hE\left[\frac{x+a-\xi+|x+a-\xi|}{2}\right] + pE\left[\frac{\xi-x-a+|x+a-\xi|}{2}\right], \\ &= \frac{h-p}{2}(x+a) + \frac{p-h}{2}E(\xi) + \frac{h+p}{2}E(|x+a-\xi|), (x,a) \in \mathbb{K}. \end{aligned}$$

Sea

$$h(x,a) = E(|x+a-\xi|) = \int |x+a-\xi|\, \Delta(s)ds,$$

veamos que h es continua en $\mathbb{K}$. Para ello sean $\{x_n\}$ y $\{a_n\}$ sucesiones en X y A, respectivamente, tales que, $x_n \to x'$ y $a_n \to a'$. Definimos a g_n y g como

$$\begin{aligned} g_n(s) &= |x_n + a_n - s|\, \Delta(s), \\ g(s) &= |x' + a' - s|\, \Delta(s). \end{aligned}$$

Así, tenemos que $g_n(s) \to g(s), s \in S$. Además,

$$\begin{aligned} g_n(s) &\leq (|x_n| + |a_n| + s)\Delta(s) \\ &\leq (M + s)\Delta(s) = f(s), \end{aligned}$$

esto es verdad, ya que $\{x_n\}$ y $\{a_n\}$ están acotadas, por ser convergentes, entonces,

$$\int g_n(s) \leq \int (M + s)\Delta(s)ds = M + E(\xi) < +\infty.$$

Por lo cual f es integrable y por el teorema de la convergencia dominada (ver Apéndice A) tenemos que $h(x_n, a_n) \to h(x', a')$, y concluimos que h es continua en $\mathbb{K}$ y por lo tanto c' es continua en $\mathbb{K}$.

Ahora veamos que c' es inf-compacta sobre $\mathbb{K}$, es decir, para todo $x \in X$ y $\lambda \in \mathbb{R}$, el conjunto $A_\lambda(x) = \{\, a \in A(x) |\, c'(x, a) \leq \lambda\}$ es compacto. Tenemos por la definición de c', que

$$\lim_{a\to\infty} c'(x, a) = \infty.$$

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

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

y esto implica que $\infty \leq \lambda$, lo cual es una contradicción. Ahora, sea $\{a'_n\}$ una sucesión en $A_\lambda(x)$, tal que $a'_n \to a \in A$, como $0 \leq c'(x, a_n) \leq \lambda$ y como c' es continua, tenemos que $0 \leq c(x, a) \leq \lambda$. Entonces, $a \in A_\lambda(x)$, por lo que concluimos que $A_\lambda(x)$ es cerrado. Aplicando el Teorema de Heine-Borel (ver [14]) tenemos que $A_\lambda(x)$ es compacto, y como λ y x fueron arbitrarios tenemos que c' es inf-compacto sobre $\mathbb{K}$.

Ahora, demostraremos que Q es fuertemente continua. Sea $\mu : X \to \mathbb{R}$ una función medible y acotada. Sea $(x_k, a_k) \in \mathbb{K}$ tal que $(x_k, a_k) \to (x, a) \in \mathbb{K}$, si $k \to \infty$. Así para cada $k = 1, 2, ...,$ usando el Teorema de Cambio de Variable, tenemos que

$$\begin{aligned} \int_X \mu(y)Q(\,dy|\, x, a) &= \int_{[0,+\infty)} \mu(x_k + a_k - s)\Delta(s)ds, \\ &= \int I_{(-\infty, x_k+a_k]}(l)\mu(l)\Delta(x_k + a_k - l)dl. \end{aligned}$$

Por otro lado,

$$\liminf(-\infty, x_k + a_k] \subset \limsup(-\infty, x_k + a_k] \subset (-\infty, x + a].$$

Así, $\{I_{(-\infty,x_k+a_k]}\}$ converge a $I_{(-\infty,x+a]}$ casi seguramente. ■

El lema anterior nos garantiza la existencia de selectores minimizadores, los cuales pueden ser obtenidos explícitamente mediante la EPD. Para resolver el ejemplo de inventarios usaremos los siguientes lemas auxiliares referentes a funciones convexas.

Lema 3.4.2 *Sea R una función real definida como*

$$R(x) = \min_{a \in A(x)} \{G(x, a)\} = G(x, f(x)),$$

con $f(x) \in A(x)$ y $x \in X$. Si $G(\cdot, \cdot)$ es una función real convexa definida en $\mathbb{K}$, y tanto X como $\mathbb{K}$, son conjuntos convexos, entonces R es una función convexa para cada $x \in X$.

Demostración. Sean $x, x' \in X$. Entonces para $0 < \alpha < 1$ tenemos

$$\begin{aligned} \alpha R(x) + (1-\alpha)R(x') &= \alpha G(x, f(x)) + (1-\alpha)G(x', f(x')) \\ &\geq G(\alpha x + (1-\alpha)x', \alpha f(x) + (1-\alpha)f(x')) \\ &= G(x'', y'') \\ &\geq G(x'', f(x'')) \\ &= R(x'') \end{aligned}$$

donde $x'' = \alpha x + (1-\alpha)x'$. ■

Lema 3.4.3 *Sea $I \subset \mathbb{R}$ un intervalo y $g : I \to \mathbb{R}$ una función convexa. Si $a < b < c$ en I, entonces*

$$\frac{g(b) - g(a)}{b - a} \leq \frac{g(c) - g(a)}{c - a} \leq \frac{g(c) - g(b)}{c - b}.$$

Demostración. Si $a < b < c$, entonces $b - a < c - a$, de lo cual tenemos si

$$\alpha = \frac{b - a}{c - a} < 1,$$

y

$$1 - \alpha = \frac{c - b}{c - a},$$

además

$$b = \frac{c-b}{c-a}a + \frac{b-a}{c-a}c,$$

por la convexidad de g tenemos que

$$g(b) \leq \frac{c-b}{c-a}g(a) + \frac{b-a}{c-a}g(c),$$

para obtener la primera desigualdad note lo siguiente

$$\begin{aligned} \frac{c-b}{c-a}g(a) + \frac{b-a}{c-a}g(c) &= \frac{c-a+a-b}{c-a}g(a) + \frac{b-a}{c-a}g(c), \\ &= g(a) + (b-a)\frac{g(c)-g(a)}{c-a}, \end{aligned}$$

y se deriva fácil.

De manera similar se obtiene la segunda desigualdad. ■

Lema 3.4.4 *Sea $I \subset \mathbb{R}$ un intervalo y $g : I \to \mathbb{R}$ una función convexa. Entonces f es monótona en I, o existe $p \in I$ tal que g es decreciente en $\{x \in I | \, x \leq p\}$ y g es creciente en $\{x \in I | \, p \leq x\}$.*

Demostración. Supongamos que g no es monótona en I. Entonces existen $x < y < z$ en I tales que, $g(x) > g(y)$ y $g(z) > g(y)$.

Por la convexidad de g, sabemos que g es continua en $[x, z]$, sabemos que existe $p \in [x, z]$ tal que, para cualquier $t \in [x, z]$ se tiene que

$$g(p) \leq g(t),$$

ahora,

si $b = x$ y $c = p$ en el lema anterior, tenemos que $g(a) \geq g(p)$ si $a < x$,
si $a = p$ y $b = z$ en el lema anterior, tenemos que $g(c) \geq g(p)$ si $c > z$,
así, concluimos que, para cualquier $t \in I$, se tiene que $g(t) \geq g(p)$.
Nuevamente el lema anterior muestra que

$$\begin{aligned} a &< b < p \text{ en } I \text{ implica que } g(a) \geq g(b), \\ p &< b < c \text{ en } I \text{ implica que } g(b) \leq g(c). \end{aligned}$$

■

Corolario 3.4.5 *Para cada $t < N$, V_t es una función convexa en X.*

Demostración. Para mostrar la convexidad de V_t, procedemos por inducción usando las funciones de iteración de valores. Inicialmente tenemos que $V_N(x) = 0, x \in X$. Entonces

$$\begin{aligned} V_{N-1}(x) &= \min_{A(x)}\{ba + L(x+a)\}, \\ &= \min_{y \geq x}\{by + L(y)\} - bx. \end{aligned}$$

Donde

$$L(y) = hE[\max\{0, y - \xi\}] + pE[\max\{0, \xi - y\}].$$

Como $G_{N-1}(y) := by + L(y), y \geq x$ es una función convexa, tenemos por el Lema 3.4.2 que V_{N-1} es una función convexa. Supongamos que V_{t+1} es convexo para $t \leq N-1$, entonces para t tenemos que

$$\begin{aligned} V_t(x) &= \min_{a \geq 0}\{ba + L(x+a) + \alpha E\left[V_{t+1}(x + a - \xi_t)\right]\} \\ &= \min_{y \geq x}\{by + L(y) + \alpha E\left[V_{t+1}(y - \xi_t)\right]\} - bx, \end{aligned}$$

donde $y = a + x$. Queremos ver que

$$G(y) = by + L(y) + \alpha E\left[V_{t+1}(y - \xi)\right],$$

$y \geq x, x \in X$ es convexa. Sabemos que $by + L(y)$ es una función convexa, así solo falta ver que $E\left[V_{t+1}(y - \xi)\right]$ también es una función convexa. Para ello sea

$$\begin{aligned} W(y) &= E\left[V_{t+1}(y - \xi)\right], \\ &= \int \left[V_{t+1}(y - s)\right] \Delta(s) ds. \end{aligned}$$

Sean y, $y' \in \mathbb{R}$. Entonces para $0 < \alpha < 1$

$$W(\alpha y + (1-\alpha)y') = \int \left[V_{t+1}(\alpha y + (1-\alpha)y' - s)\right] \Delta(s) ds,$$

como $s = \alpha s + (1-\alpha)s$, usando el hecho de que V_{t+1} es una función convexa

tenemos que

$$\begin{aligned} W\left(\alpha y+(1-\alpha)y'\right) &= \int \left[V_{t+1}(\alpha y+(1-\alpha)y'-(\alpha s+(1-\alpha)s))\right]\Delta(s)ds, \\ &= \int \left[V_{t+1}(\alpha(y-s)+(1-\alpha)(y'-s))\right]\Delta(s)ds, \\ &\leq \int \alpha\left[V_{t+1}(y-s)+(1-\alpha)V_{t+1}(y'-s)\right]\Delta(s)ds, \\ &= \alpha\int V_{t+1}(y-s)\Delta(s)ds+(1-\alpha)\int V_{t+1}(y'-s)\Delta(s)ds, \\ &= \alpha W(y)+(1-\alpha)W(y'), \end{aligned}$$

y esto prueba el corolario. ■

Con la ayuda de los resultados anteriores y el Teorema 3.1.1 encontraremos la política óptima del ejemplo de inventarios.

Así, para $t=N-1$, tenemos que

$$\begin{aligned} V_{N-1}(x) &= \min_{A(x)}\{ba+L(x+a)\}, \\ &= \min_{y\geq x}\{by+L(y)\}-bx. \end{aligned}$$

Sea $G_{N-1}(y):=by+L(y), y\geq x$. Derivando a G_{N-1} con respecto a y,

$$G'_{N-1}(y)=b+L'(y).$$

Por otro lado, notemos que

$$\begin{aligned} L(y) &= h\int_{-\infty}^{y}(y-s)\,\Delta(s)ds+p\int_{y}^{\infty}(s-y)\,\Delta(s)ds, \\ &= h\int_{-\infty}^{y}y\Delta(s)ds-h\int_{-\infty}^{y}s\Delta(s)ds+p\int_{y}^{\infty}s\Delta(s)ds-p\int_{y}^{\infty}y\Delta(s)ds, \\ &= hy\mu(y)-h\int_{-\infty}^{y}s\Delta(s)ds+pE(\xi)-p\int_{-\infty}^{y}s\Delta(s)ds-py(1-\mu(y)), \\ &= hy\mu(y)-(h+p)\int_{-\infty}^{y}s\Delta(s)ds+pE(\xi)-py(1-\mu(y)), \end{aligned}$$

Así, tenemos que

$$\begin{aligned} L'(y) &= h\mu(y) + hy\Delta(y) - (h+p)y\Delta(y) - p + p\mu(y) + py\Delta(y), \\ &= (h+p)\,\mu(y) - (h+p)y\Delta(y) + (h+p)\,y\Delta(y) - p, \\ &= (h+p)\,\mu(y) - p. \end{aligned}$$

Entonces,

$$G'_{N-1}(y) = b + (h+p)\,\mu(y) - p,$$

igualando a cero tenemos que

$$\begin{aligned} G'_{N-1}(y) &= 0, \\ (h+p)\,\mu(y) &= p - b, \\ \mu(y) &= \frac{p-b}{h+p}. \end{aligned}$$

Entonces, dado que μ es creciente, existe μ^{-1}, así, el punto

$$s_{N-1} = \mu^{-1}\left(\frac{p-b}{h+p}\right),$$

minimiza a G_{N-1} y tenemos que el minimizador de V_{N-1} es y^* donde

$$y^* = \begin{cases} x, & \text{si } x \geq s_{N-1}, \\ s_{N-1,} & \text{si } s_{N-1} > x. \end{cases}$$

Haciendo el cambio de variable tenemos que

$$f_{N-1}(x) = \begin{cases} 0, & \text{si } x \geq s_{N-1}, \\ s_{N-1} - x, & \text{si } s_{N-1} > x. \end{cases}$$

Así, sustituyendo en $V_{N-1}(x) = \min_{a \in A(x)} \{ba + L(x+a)\}$, obtenemos que

$$V_{N-1}(x) = \begin{cases} L(x), & \text{si } x \geq s_{N-1}, \\ b(s_{N-1} - x) + L(s_{N-1}), & \text{si } s_{N-1} > x. \end{cases}$$

$x \in X$.

Finalmente par concluir haremos uso del Lema 3.4.4 y del Corolario 3.4.5. La función $G_t(y) := by + L(y) + \alpha E\left[V_{t+1}(y - \xi_t)\right], y \geq x$ es convexa y tiene

un mínimo en un punto s_t, debido al Lema 3.4.4. Por lo tanto, al minimizar el lado derecho de la ecuación siguiente

$$V_t(x) = \min_{A(x)}\{ba + L(x+a) + \alpha E\left[V_{t+1}(x+a-\xi_t)\right]\},$$

tenemos que

$$f_t(x) = \begin{cases} 0, & \text{si } x \geq s_t, \\ s_t - x, & \text{si } s_t > x. \end{cases}$$

y

$$V_t(x) = \begin{cases} L(x) + \alpha E\left[V_{t+1}(x-\xi_t)\right], & \text{si } x \geq s_t, \\ b(s_t - x) + L(s_t) + \alpha E\left[V_{t+1}(s_t-\xi_t)\right], & \text{si } s_t > x. \end{cases}$$

Por el teorema de programación dinámica, los selectores f_t determinan una política óptima π^*.

Sistemas con Espacio de Estados Discretos

Existen sistemas en los cuales es factible considerar el espacio de estados discreto, tales sistemas son convenientes especificarlos en términos de las probabilidades de transición. Así, el espacio de estados X es finito o infinito numerable.

Denotaremos por $P_{ij}(a,t)$ a la probabilidad de estar en el estado j, dado que nos encontramos en el estado i y se escogió la acción a, en el tiempo t, es decir,

$$P_{ij}(a,t) = P\left[x_{t+1} = j \middle| x_t = i, a_t = a\right],$$

y tales sistemas están descritos en términos de la ecuación en diferencias de la forma siguiente

$$x_{t+1} = \xi_t.$$

Donde la distribución de ξ_t es

$$P\left[\xi_{t+1} = j \middle| x_t = i, a_t = a\right] = P_{ij}(a,t).$$

Ahora denotemos por $g(i,a)$ al costo de encontrarnos en el estado i cuando la acción a es aplicada. Nuestro criterio de rendimiento es el costo total acumulado

$$V(\pi, x) = E_x^{\pi}\left[\sum_{t=o}^{N-1} g_t(x_t, a_t) + g_N(x_N)\right].$$

En nuestro caso tenemos que la EPD

$$\begin{aligned} V_N(x) &= 0, \\ V_t(i) &= \min_{a\in A(i)} \{g(i,a) + E\left[V_{t+1}(\xi_t)\right]\}, \end{aligned}$$

y por la caracterización anterior tenemos que

$$V_t(i) = \min_{a\in A(i)} \{g(i,a) + \sum_j P_{ij}(a) V_{t+1}(j)\}.$$

Reemplazamiento de Maquinas.

Consideremos un problema de operación eficiente de una máquina en N periodos. El espacio de estados es finito, es decir, $S = \{1, 2, ..., n\}$. Suponga que encontrarse en el estado i es mejor que estar en el estado $i+1$, y el estado 1 es el de perfectas condiciones en la que puede operar la máquina.

Denotemos por $g(i)$ al costo de operación por periodo cuando estamos en el estado i y supongamos que

$$g(1) \leq g(2) \leq ... \leq g(n).$$

Durante un periodo de operación, el estado de la máquina puede empeorar o quedarse igual.

Consideremos la probabilidad de transición es

$$p_{ij} = \begin{cases} p(j\,|\,i), & j \geq i, \\ 0, & j < i. \end{cases}$$

La matriz de transición es

$$\begin{bmatrix} p_{11} & p_{12} & . & . & . & p_{1n} \\ 0 & p_{22} & . & . & . & p_{2n} \\ 0 & 0 & . & . & . & . \\ . & . & 0 & p_{ii} & . & p_{in} \\ . & . & . & . & . & . \\ 0 & . & . & . & 0 & p_{nn} \end{bmatrix}.$$

Supóngase también que al comenzar cada periodo conocemos el estado de la máquina y tenemos las opciones siguientes:

1. Que la máquina opere un periodo más, en el estado en que se encuentra.

2. Reparar la máquina hasta estar en perfectas condiciones, es decir, hasta estar en el estado 1, pagando un costo R.

Con esto tenemos que el espacio de acciones A es también finito y tiene dos elementos (operar ó reparar).

Suponemos que la máquina, una vez que esté reparada, está permanecerá en el estado 1 por al menos un período. En períodos subsecuentes, puede deteriorarse para los estados $j > 1$ según las probabilidades de transición p_{1j}.

Así el objetivo es decidir sobre el nivel de deterioro (el estado) en el cual vale la pena pagar el costo de reparación de la máquina, de tal modo obtener la ventaja de tener costos de operación más pequeños en el futuro. Notemos que la decisión también debe afectarse por el período en que nos encontramos, por ejemplo, estaríamos menos inclinados de reparar la máquina cuando hay pocos períodos a la izquierda.

Así,

$$V_t(i) = \min_{a \in A(i)} \{g(i,a) + E\left[V_{t+1}(\xi_t)\right]\}.$$

Ahora calculamos

$$\begin{aligned} E\left[V_{t+1}(\xi_t)\right] &= \sum_j p_{ij} V_{t+1}(j), \\ &= \sum_{j=i}^{n} p_{ij} V_{t+1}(j). \end{aligned}$$

Así tenemos que en general

$$V_k(i) = \min\{R + g(1) + V_{k+1}(1), g(i) + \sum_{j=i}^{n} P_{ij} V_{t+1}(j)\},$$

esto significa dejar un periodo más trabajando o reparar la máquina. Por ejemplo:

Supongamos que $X = \{1, 2, 3, 4\}$ y la máquina es revisada durante dos periodos. La matriz de transición es

$$\begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

el costo esta definido como $g(i,a) = i$, para $i = 1,2,3,4$. Y el costo de reparación $R = \frac{3}{2}$.

Utilicemos programación dinámica

$$V_2(x) = 0.$$

En la siguiente etapa

$$\begin{aligned} V_1(i) &= \min\{R + g(1) + V_2(1), g(i) + \sum_{j=i}^{4} P_{ij}V_2(j)\} \\ &= \min\{\frac{5}{2}, i\}. \end{aligned}$$

Así

$$V_1(i) = \begin{cases} i, & \text{si } i = 1,2, \\ \frac{5}{2}, & \text{si } i = 3,4. \end{cases}$$

es decir, para los dos primeros estados es mejor dejarlos donde están, para los siguientes es recomendable mandar a repara la máquina.

Para la siguiente etapa

$$\begin{aligned} V_0(i) &= \min\{R + g(1) + V_1(1), g(i) + \sum_{j=i}^{4} P_{ij}V_1(j)\} \\ &= \min\{\frac{5}{2} + 1, i + \sum_{j=i}^{4} P_{ij}V_2(j)\} \\ &= \min\{\frac{7}{2}, i + \sum_{j=i}^{4} P_{ij}V_2(j)\}. \end{aligned}$$

Ahora calculamos

$$\begin{aligned} V_0(1) &= \min\{\frac{7}{2}, 1 + \sum_{j=1}^{4} P_{1j}V_1(j)\} \\ &= \min\{\frac{7}{2}, 1 + \frac{1}{4}\left(1 + 2 + \frac{5}{2} + \frac{5}{2}\right)\} \\ &= \min\{\frac{7}{2}, 1 + \frac{1}{4}\left(1 + 2 + \frac{5}{2} + \frac{5}{2}\right)\} \\ &= \min\{\frac{7}{2}, 3\} \\ &= 3, \end{aligned}$$

$$\begin{aligned} V_0(2) &= \min\{\frac{7}{2}, 2 + \sum_{j=2}^{4} P_{2j} V_1(j)\} \\ &= \min\{\frac{7}{2}, 2 + \frac{1}{3}\left(2 + \frac{5}{2} + \frac{5}{2}\right)\} \\ &= \min\{\frac{7}{2}, 2 + \frac{1}{3}\left(2 + \frac{5}{2} + \frac{5}{2}\right)\} \\ &= \min\{\frac{7}{2}, \frac{13}{3}\} \\ &= \frac{7}{2}, \end{aligned}$$

$$\begin{aligned} V_0(3) &= \min\{\frac{7}{2}, 3 + \sum_{j=3}^{4} P_{3j} V_1(j)\} \\ &= \min\{\frac{7}{2}, 3 + \frac{1}{2}\left(\frac{5}{2} + \frac{5}{2}\right)\} \\ &= \min\{\frac{7}{2}, \frac{11}{2}\} \\ &= \frac{7}{2}, \end{aligned}$$

$$\begin{aligned} V_0(4) &= \min\{\frac{7}{2}, 4 + \frac{5}{2}\} \\ &= \frac{7}{2}, \end{aligned}$$

Así,

$$V_0(i) = \begin{cases} i, \text{si } i = 1, \\ \frac{7}{2}, \text{si } i = 2, 3, 4. \end{cases}.$$

Concluimos que solo en el primer estado es mejor dejarlo trabajar y en los restantes reparar la máquina.

Capítulo 4

PROBLEMAS CON HORIZONTE INFINITO

En ocasiones es conveniente considerar un PDM con horizonte infinito. El objetivo de este capítulo es nuevamente proveer una herramienta para el problema de control óptimo.

El criterio de rendimiento es el de Costo Total Descontado (o simplemente Costo Total) con horizonte infinito, es decir,

$$V(\pi, x) := E_x^{\pi}\left[\sum_{t=o}^{\infty} \alpha^t c(x_t, a_t)\right], \pi \in \Pi, x \in X.$$

Donde $\alpha \in (0, 1)$ es el factor de descuento. Nuevamente recordamos que el problema es determinar una política π^* óptima, es decir,

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

Supondremos en esta sección que el costo c es no negativo y definimos al n-ésimo costo descontado como

$$V_n(\pi, x) := E_x^{\pi}\left[\sum_{t=o}^{n-1} \alpha^t c(x_t, a_t)\right].$$

Así, por el teorema de la convergencia monótona (ver Apéndice A5), tenemos que

$$V(\pi, x) = \lim_{n \to \infty} V_n(\pi, x).$$

4.1. Ecuación Óptima para el Costo Descontado

Una función medible $v : X \to \mathbb{R}$ es una solución de la ecuación óptima para el costo descontado (EOCD) si satisface que, para cada $x \in X$

$$v(x) = \inf_{a \in A(X)} \left\{ c(x,a) + \alpha \int_X v(y) Q(dy \mid x, a) \right\},$$

Veremos que, la función de valores óptimos V^* satisface la EOCD, es decir, para cada $x \in X$

$$V^*(x) = \inf_{a \in A(X)} \left\{ c(x,a) + \alpha \int_X V^*(y) Q(dy \mid x, a) \right\}. \tag{4.1}$$

Para ello utilizaremos las funciones de iteración de valores (IV), definidas de la siguiente forma: para $x \in X$ y $n = 0, 1, 2, ...$

$$\begin{aligned} v_0(x) &= 0, \qquad (4.2) \\ v_n(x) &= \min_{A(x)} \left[c(x,a) + \alpha \int_X v_{n-1}(y) Q(dy \mid x, a) \right]. \end{aligned}$$

Observemos que, v_n es la función de valor óptimo del n-ésimo costo descontado V_n, con costo terminal cero, es decir,

$$v_n(x) = \inf_{\pi \in \Pi} V_n(\pi, x),$$

$x \in X$. Se mostrará que para cada $x \in X$

$$\lim_{n \to \infty} v_n(x) = V^*(x). \tag{4.3}$$

Tomando límite cuando $n \to \infty$ en 4.2 y usando 4.3, obtenemos 4.1, si se cumple el intercambio de límite con el mínimo. Este procedimiento es conocido como aproximaciones sucesivas, y requiere principalmente condiciones de selección medible, las cuales se darán a continuación.

Suposición 4.1.1 *a) El costo c es l.s.c, no negativo e inf-compacto en $\mathbb{K}$.*
b) Q es fuertemente continua.

Basta en a) que c sea acotada inferiormente en lugar de no negativa, debido a que, si $c \geq m$, entonces $c' = c - m \geq 0$.

Suposición 4.1.2 *Existe $\pi \in \Pi$ tal que, para cada $x \in X$ se tiene que $V(\pi, x) < +\infty$.*

La suposición anterior es de suma importancia para el caso horizonte infinito, pues garantiza que la función de valor óptimo es finita para cada $x \in X$. Claramente se cumple cuando el costo c es acotado, ya que, si $0 \leq c \leq M$ entonces para cada $x \in X$ y $\pi \in \Pi$

$$V(\pi, x) = E_x^{\pi}\left[\sum_{t=o}^{\infty} \alpha^t c(x_t, a_t)\right] \leq \sum_{t=o}^{\infty} \alpha^t M = \frac{M}{1-\alpha} < +\infty.$$

Teorema 4.1.3 *Bajo las suposiciones anteriores tenemos:*
a) La función de valor óptimo V^* es solución de la EOCD, es decir, para cada $x \in X$

$$V^*(x) = \inf_{a \in A(X)} \left\{ c(x,a) + \alpha \int_X V^*(y) Q(dy \mid x, a) \right\},$$

y si u es otra solución de la EOCD, entonces $u \geq V^*$;
b) Existe un selector $f_* \in \mathbb{F}$ tal que $f_*(x) \in A(x)$ y alcanza un mínimo en la EOCD, es decir, para cada $x \in X$

$$V^*(x) = c(x, f_*) + \alpha \int_X V^*(y) Q(dy \mid x, f_*),$$

c) Si π^* es una política tal que $V(\pi^*, \cdot)$ es una solución de la EOCD y satisface que para cada $x \in X$

$$\lim_{n \to \infty} \alpha^n E_x^{\pi} V(\pi^*, x_n) = 0,$$

entonces, $V(\pi^*, \cdot) = V^*(\cdot)$; de aquí, π^* es óptima, concluyendo que, si ocurre lo anterior, entonces π^* es una política óptima, si y sólo si, $V(\pi^*, \cdot)$ satisface la EOCD.
d) Si existe una política π óptima, entonces existe una que es determinista estacionaria.

Para la demostración del teorema anterior, se usarán los lemas siguientes.

Lema 4.1.4 *Sean u y u_n $(n = 1, 2, ...)$ funciones l.s.c, acotadas inferiormente e inf-compactas sobre $\mathbb{K}$. Si $u_n \uparrow u$, entonces para cada $x \in X$*

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

Demostración. Definamos para cada $x \in X$,

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

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

Dado que $u_n \uparrow u$, tenemos que $l(\cdot) \leq u^*(\cdot)$.

Para demostrar la desigualdad inversa, sea $x \in X$ fijo, y definamos para cada $n = 0, 1, 2, ...$

$$A_n := \{a \in A(x) \mid u_n(x, a) \leq u^*(x)\},$$

$$A_0 := \{a \in A(x) \mid u(x, a) = u^*(x)\}.$$

Notemos que para cada n, se tiene que A_n es compacto, debido a la inf-compacidad de u_n. Además, $A_n \downarrow A_0$,

$$\bigcap_{n=1}^{\infty} A_n = A_0.$$

Esto es verdad, ya que, si $a \in \bigcap_{n=1}^{\infty} A_n$ entonces para toda n,

$$u_n(x, a) \leq u^*(x),$$

luego,

$$\lim_{n\to\infty} u_n(x, a) = u(x, a) \leq u^*(x).$$

Además, por definición,

$$u^*(x) \leq u(x, a),$$

por lo que concluimos que $u(x, a) = u^*(x)$, de lo cual tenemos que $a \in A_0$.

Por lo anterior, A_0 es compacto.

Ahora, por el teorema de selección medible (ver apéndice D) se tiene que, para cada $n \geq 1$, existe $a_n \in A_n$ tal que

$$u_n(x, a_n) = \min_{A(x)} u_n(x, a).$$

Así, por la compacidad de A_0, existe una subsucesión $\{a_{n_i}\}$ de la sucesión $\{a_n\}$, tal que $a_{n_i} \to a_0 \in A_0$.

Entonces para toda $n_i \geq n$ se tiene que

$$u_{n_i}(x, a_{n_i}) \geq u_n(x, a_{n_i}).$$

Cuando $n_i \to \infty$ en la desigualdad anterior, obtenemos que

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

Si $n \to \infty$,

$$l(x) \geq u(x, a_0) \geq \min_{A(x)} u(x, a) = u^*(x),$$

por lo que concluimos que

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

■

Definición 4.1.5 *$M(X)^+$ denota el conjunto de funciones medibles y no negativas definidas sobre X, para cada $u \in M(X)^+$, definimos a la función Tu sobre X como*

$$Tu(x) := \min_{A(x)} \left[c(x, a) + \alpha \int_X u(y) Q(\, dy |\, x, a) \right].$$

Lema 4.1.6 *Bajo las suposición 4.1.1, T es un operador sobre $M(X)^+$, es decir, para cada $u \in M(X)^+$, tenemos que $Tu \in M(X)^+$. Además, existe $f \in \mathbb{F}$ tal que*

$$Tu(x) = c(x, f) + \alpha \int_X u(y) Q(\, dy |\, x, f).$$

Notemos que al usar al operador T, en el teorema anterior en a), tenemos que $V^* = TV^*$ y podemos escribir a las funciones de IV como

$$v_0 = 0,$$

$$v_n = Tv_{n-1},$$

para toda $n \geq 1$.

Mostraremos que la función de valor óptimo V^* es un punto fijo del operador T.

Lema 4.1.7 *Bajo las suposiciones 4.1.1 y 4.1.2 se tiene:*

a) Si $u \in M(X)^+$ y $u \geq Tu$, entonces, $u \geq V^*$;

b) Si $u : X \to \mathbb{R}$ es una función medible tal que Tu está bien definida y satisface que, para toda $\pi \in \Pi$ y $x \in X$

$$u \leq Tu,$$

y

$$\lim_{n\to\infty} \alpha^n E_x^\pi [u(x_n)] = 0,$$

entonces, $u \leq V^*$.

Demostración. *(a) Sea $u \in M(X)^+$, tal que $u \geq Tu$, por el lema anterior tenemos que para toda $x \in X$*

$$u(x) \geq c(x, f) + \alpha \int_X u(y) Q(\, dy |\, x, f),$$

iterando esta relación, obtenemos lo siguiente:

$$\begin{aligned} u(x) &\geq c(x,f) + \alpha \left[\int_X c(y,f) + \alpha \int_X u(z) Q(\, dz |\, y, f) \right] Q(\, dy |\, x, f), \\ &= c(x,f) + \alpha \int_X c(y,f) Q(dy \mid x, f) \\ &\quad + \alpha^2 \int_X \int_X u(z) Q(\, dz |\, y, f) Q(\, dy |\, x, f), \\ &= E_x^f \sum_{t=0}^{1} \alpha^t c(x_t, a_t) + \alpha^2 E_x^f [u(x_2)], \end{aligned}$$

continuando iterando llegamos a

$$u(x) \geq E_x^f \sum_{t=0}^{n-1} \alpha^t c(x_t, a_t) + \alpha^n E_x^f[u(x_n)],$$

donde

$$E_x^f[u(x_n)] = \int u(y) Q^n(\, dy|\, x, f).$$

Dado que u es no negativa se tiene que para toda n y x que

$$u(x) \geq E_x^f \sum_{t=0}^{n-1} \alpha^t c(x_t, a_t).$$

Así, si $n \to \infty$,

$$u(x) \geq V(f, x) \geq V^*(x),$$

para toda $x \in X$ y por lo tanto,

$$u \geq V^*,$$

de lo cual concluimos a).

(b) Ahora, sean $\pi \in \Pi$ y $x \in X$, y supongamos que $u \leq Tu$, entonces por la propiedad de Markov tenemos

$$\begin{aligned}
E_x^f \left[\alpha^{t+1} u(x_{t+1}) \middle|\, h_t, a_t\right] &= \alpha^{t+1} \int u(y) Q(\, dy|\, x_t, a_t), \\
&= \alpha^t [c(x_t, a_t) - c(x_t, a_t) + \\
&\quad \alpha \int u(y) Q(\, dy|\, x_t, a_t)], \\
&= \alpha^t \left[c(x_t, a_t) + \alpha \int u(y) Q(\, dy|\, x_t, a_t) \right] \\
&\quad -\alpha^t c(x_t, a_t), \\
&\geq \alpha^t \left[u(x_t) - c(x_t, a_t) \right],
\end{aligned}$$

lo que implica que

$$\alpha^t c(x_t, a_t) \geq E_x^f \left[\alpha^t u(x_{t+1}) - \alpha^{t+1} u(x_t) \middle|\, h_t, a_t\right].$$

Tomando esperanzas y sumando desde $t = 0, ..., n-1$, *en la última desigualdad, obtenemos*

$$\begin{aligned} E_x^f\left[\sum_{t=0}^{n-1}\alpha^t c(x_t,a_t)\right] &\geq E_x^f\sum_{t=0}^{n-1}\left[\alpha^t u(x_{t+1})-\alpha^{t+1}u(x_t)\middle| h_t,a_t\right], \\ &= E_x^f\left[u(x_0)\right]-\alpha^n E_x^f\left[u(x_n)\right], \\ &= u(x)-\alpha^n E_x^f\left[u(x_n)\right], \end{aligned}$$

para toda n, *así, cuando* $n\to\infty$, *tenemos que*

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

y esto implica que

$$V(\pi,x)\geq u(x,).$$

Como π *fue arbitraria, concluimos que*

$$V^*\geq u.$$

■

Lema 4.1.8 *Bajo las suposiciones 4.1.1 y 4.1.2 , se tiene que* $v_n\uparrow V^*$ *y* $V^*=TV^*$, *es decir,* V^* *es una solución de EOCD.*

Demostración. Para empezar notemos que para cada $x\in X$, $\pi\in\Pi$ y $n=1,2,...,$

$$v_n(x)\leq V_n(\pi,x)\leq V(\pi,x),$$

por lo tanto

$$v_n(x)\leq V^*(x).$$

Debido a que el operador T es monótono (i.e., si $u,u'\in M(X)^+$ entonces $Tu\leq Tu'$), se tiene que las funciones de iteración de valores:

$$v_0=0,$$

$$v_n=Tv_{n-1},$$

$n\geq 1$, forman una sucesión creciente en $M(X)^+$, lo cual implica que $v_n\uparrow v^*$, para alguna $v^*\in M(X)^+$.

Sean,

$$u_n(x,a) := c(x,a) + \alpha \int v_n(y) Q(\,dy|\,x,a),$$

$$u(x,a) := c(x,a) + \alpha \int v^*(y) Q(\,dy|\,x,a),$$

$(x,a) \in \mathbb{K}$. Entonces por el teorema de la convergencia monótona, tenemos que $u_n \uparrow u$.

Por otro lado, las funciones u_n y u, son l.s.c e inf-compactas en $\mathbb{K}$. Así por el Lema 4.1.4 tenemos que

$$v^* = \lim_{n\to\infty} v_n = \lim_{n\to\infty} T v_{n-1} = T v^*$$

esto es, v^* es una solución de la EOCD.

Para complementar la demostración, solo resta ver que $v^* = V^*$. Para esto por el Lema 4.1.7 (a), sabemos que $v^* = Tv^*$ lo cual implica que $v^* \geq V^*$. Además, ya que $v_n(x) \leq V^*(x)$, tenemos que $v^* \leq V^*$. Por lo tanto, $v^* = V^*$.

■

Ahora estamos listos para dar una demostración del Teorema 4.1.3.

Demostración. (Teorema 4.1.3)

Por el Lema 4.1.8, tenemos que V^* es solución de la EOCD, y por el Lema 4.1.7 a) V^* es la mínima solución de la EOCD, es decir, si $u \geq Tu$ entonces $u \geq V^*$.

Por el Lema 4.1.6 existe $f^* \in \mathbb{F}$, tal que para cada $x \in X$

$$V^*(x) = c(x, f^*) + \alpha \int V^*(y) Q(\,dy|\,x, f^*).$$

Iterando la relación anterior tenemos para cada $x \in X$ y $n \geq 1$

$$\begin{aligned} V^*(x) &= E_x^{f^*}\left[\sum_{t=0}^{n-1} \alpha^t c(x_t, f^*)\right] + \alpha^n E_x^{f^*}\left[V^*(x_n)\right], \\ &\geq E_x^{f^*}\left[\sum_{t=0}^{n-1} \alpha^t c(x_t, f^*)\right], \end{aligned}$$

lo cual implica, cuando $n \to \infty$, que

$$V^*(x) \geq V(f^*, x),$$

$x \in X$.

Por otro lado, sabemos que

$$V^*(x) \leq V(f^*, x),$$

$x \in X$. Así,

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

$x \in X$, es decir, f^* es óptima.

Recíprocamente, si $f^* \in \Pi_{DS}$ entonces

$$\begin{aligned} V(f_*, x) &= E_x^{f^*}\left[\sum_{t=0}^{\infty} \alpha^t c(x_t, f^*)\right], \\ &= E_x^{f^*}\left[c(x_0, f^*) + \sum_{t=1}^{n-1} \alpha^t c(x_t, f^*)\right], \\ &= c(x, f^*) + E_x^{f^*}\left[\sum_{t=1}^{\infty} \alpha^t c(x_t, f^*)\right]. \end{aligned}$$

Usando la Proposición B.1(c) del Apéndice B y la propiedad de Markov tenemos que

$$\begin{aligned} E_x^{f^*}\left[\sum_{t=1}^{\infty} \alpha^t c(x_t, f^*)\right] &= \int_X E_x^{f^*}\left[\sum_{t=1}^{\infty} \alpha^t c(x_t, f^*)\middle| x_1 = y\right] Q(\,dy|\, x, f^*), \\ &= \int_X V(f^*, y) Q(dy \mid x, f^*). \end{aligned}$$

En particular, si f^* es óptima entonces $V^*(x) = V(f^*, x)$.

Ahora, si π^* es una política tal que $V(\pi^*, \cdot)$ es una solución de la EOCD por el Lema 4.1.7 se tiene que $V(\pi^*, \cdot) = V^*(\cdot)$. Finalmente podemos concluir que, si una política óptima existe, entonces existe una política determinista estacionaria. ■

4.2. Ejemplos

4.2.1. Problema LQ o Lineal Cuadrático

Los problemas LQ consisten de un sistema lineal con un costo cuadrático y es uno de los problemas de control más usados en ingeniería, economía

y muchos otros campos. Consideremos un sistema definido por la siguiente ecuación en diferencias

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

para $t = 0, 1, ...$ Donde γ y β son constantes reales. La función de costo esta dada por

$$c(x, a) = qx^2 + ra^2,$$

donde q y r son constantes reales tales que $q \geq 0$ y $r > 0$. $\{\xi_t\}$ es una sucesión de v.a's i.i.d tomando valores en $S = \mathbb{R}$, con función de densidad continua Δ, independientes del estado inicial x_0, con media 0 y varianza finita σ^2, es decir,

$$\begin{aligned} E(\xi) &= 0, \\ E(\xi^2) &= \sigma^2 < +\infty. \end{aligned}$$

Sea $X = A = A(x) = \mathbb{R}$. Bajo las consideraciones anteriores deseamos encontrar una política que minimice el criterio de rendimiento

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

para $\pi \in \Pi, x \in X$.

Para resolverlo necesitamos verificar que cumpla con las suposiciones 4.1.1 y 4.1.2.

Lema 4.2.1 *El problema LQ satisface las suposiciones 4.1.1 y 4.1.2.*

Demostración. Notemos que el costo por ser cuadrático es continuo (y por tanto l.s.c y no negativo), solo falta probar que es inf-compacto. Para ello sean $\lambda \in \mathbb{R}$ y $x \in X$,

$$\begin{aligned} \{a|\, c(x, a) \leq \lambda\} &= \left\{a|\, qx^2 + ra^2 \leq \lambda\right\}, \\ &= \left\{a|\, a^2 \leq \lambda - qx^2\right\}, \\ &= \left\{a|\, a^2 \leq \frac{\lambda - qx^2}{r}\right\}, \\ &= \left\{a \,\middle|\, |a| \leq \sqrt{\frac{\lambda - qx^2}{r}}\right\}. \end{aligned}$$

Observemos que

$$\{a|\, c(x,a) \leq \lambda\} = \left[-\sqrt{\frac{\lambda - qx^2}{r}}, \sqrt{\frac{\lambda - qx^2}{r}}\right],$$

si $\sqrt{\frac{\lambda}{q}} \geq |x|$ y será vacío en otro caso. Por lo tanto, el conjunto $\{a|\, c(x,a) \leq \lambda\}$ es compacto para cualquier $\lambda \in \mathbb{R}$ y $x \in X$, es decir, el costo es inf-compacto.

Ahora veamos que Q es fuertemente continua, para ello recordemos del capítulo anterior que si la dinámica esta dada por una ecuación de diferencias tenemos que

$$\begin{aligned} Q(B|\, x,a) &= \Pr(x_{t+1} \in B|\, x_t = x, a_t = a), \\ &= \Pr(F(x_t, a_t, \xi_t) \in B|\, x_t = x, a_t = a), \\ &= \Pr(F(x,a,\xi) \in B), \\ &= \int_S I_B[F(x,a,s)]\mu(ds). \end{aligned}$$

Así, si Δ es la densidad de ξ, entonces

$$Q(B|\, x,a) = \int_S I_B[\gamma x + \beta a + s)]\Delta(s)ds,$$

con un cambio de variable tenemos que

$$Q(B|\, x,a) = \int I_C((u))\Delta(u - \gamma x - \beta a)du,$$

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

Ahora para verificar la condición 4.1.2, sea

$$f(x) = -\frac{\gamma x}{\beta},$$

$x \in X$. Entonces,

$$\begin{aligned} V(f,x) &= E_x^\pi \left[\sum_{t=0}^{\infty} \alpha^t \left(q \left(\gamma x_{t-1} - \beta \frac{\gamma x_{t-1}}{\beta} + \xi_{t-1} \right)^2 + r \frac{\gamma x_t{}^2}{\beta} \right) \right], \\ &= E_x^\pi \left[\sum_{t=0}^{\infty} \alpha^t \left(q\xi_{t-1}^2 + \frac{r\gamma}{\beta} x_t^2 \right) \right], \\ &= \sum_{t=0}^{\infty} \alpha^t \left(q\sigma^2 + \frac{r\gamma}{\beta} E_x^\pi \left(x_t^2 \right) \right). \end{aligned}$$

Ahora, dado que estamos aplicando la política f, obtenemos

$$\begin{aligned} x_{t+1} &= \gamma x_t + \beta f(x_t) + \xi_t, \\ &= \gamma x_t - \frac{\beta\gamma}{\beta} x_t + \xi_t, \\ &= \xi_t, \end{aligned}$$

de lo cual tenemos que

$$E_x^\pi \left(x_t^2 \right) = \sigma^2.$$

Por lo tanto,

$$\begin{aligned} V(f,x) &= \sum_{t=0}^{\infty} \alpha^t \left(\sigma^2 \left[q + \frac{r\gamma}{\beta} \right] \right), \\ &= \left(\sigma^2 \left[q + \frac{r\gamma}{\beta} \right] \right) \sum_{t=0}^{\infty} \alpha^t, \\ &= \left(\sigma^2 \left[q + \frac{r\gamma}{\beta} \right] \right) \frac{1}{1-\alpha} < +\infty, \end{aligned}$$

que es lo deseado. ■

Así podemos aplicar el algoritmo de programación dinámica, para ello inicialmente encontramos las funciones de iteración de valores.

$$\begin{aligned} v_0(x) &= 0, \\ v_n(x) &= \min_{A(x)} \left[c(x,a) + \alpha \int_X v_{n-1}(y) Q(dy \mid x, a) \right], \end{aligned}$$

$x \in X$.

Lema 4.2.2 *Para cada $n \geq 1$ las funciones de iteración de valores están dadas por*

$$v_n(x) = K_n x^2 + E_n,$$

$x \in X$, donde

$$K_n = \left(\frac{q\left(r + \alpha K_{n-1}\beta^2\right) + r\alpha K_{n-1}\gamma^2}{r + \alpha K_{n-1}\beta^2} \right),$$

y

$$E_n = K_{n-1}\alpha\sigma^2\left(1 + \alpha q\right).$$

para toda $x \in X$ y $n = 1, 2, ...$

Demostración. Para $n = 1$, tenemos

$$v_1(x) = \min_{A(x)} \left\{ qx^2 + ra^2 \right\} = qx^2,$$

con $f_1(x) = 0$. Ahora, para $n = 2$

$$\begin{aligned} v_2(x) &= \min_{A(x)} \left\{ qx^2 + ra^2 + \alpha E\left[v_1(\gamma x + \beta a + \xi)\right] \right\}, \\ &= \min_{A(x)} \left\{ qx^2 + ra^2 + \alpha q E\left[(\gamma x + \beta a)^2 + 2(\gamma x + \beta a)\xi + \xi^2\right] \right\}, \\ &= \min_{A(x)} \left\{ qx^2 + ra^2 + \alpha q(\gamma x + \beta a)^2 + \alpha q\sigma^2 \right\}. \end{aligned}$$

Sea

$$G_2(x, a) = qx^2 + ra^2 + \alpha q(\gamma x + \beta a)^2 + q\sigma^2,$$

$(x, a) \in \mathbb{K}$, derivando con respecto a a e igualando a cero tenemos que

$$2ra + 2\alpha q\beta(\gamma x + \beta a) = 0,$$

de lo cual obtenemos que

$$f_2(x) = -\frac{\alpha q\beta\gamma}{r + \alpha q\beta^2} x,$$

$x \in X$. Así, al aplicar esta política llegamos a

$$\begin{aligned} v_2(x) &= qx^2 + r\left(\frac{\alpha q\beta\gamma}{r+\alpha q\beta^2}\right)^2 x^2 + \alpha q(\gamma x - \beta\frac{\alpha q\beta\gamma}{r+\alpha q\beta^2}x)^2 + \alpha q\sigma^2, \\ &= qx^2 + r\left(\frac{\alpha q\beta\gamma}{r+\alpha q\beta^2}\right)^2 x^2 + \alpha q\left(\gamma\frac{r+\alpha q\beta^2}{r+\alpha q\beta^2}x - \frac{\alpha q\beta^2\gamma}{r+\alpha q\beta^2}x\right)^2 \\ &\quad +\alpha q\sigma^2, \\ &= \left(\frac{q\left(r+\alpha q\beta^2\right)^2 + r\left(\alpha q\beta\gamma\right)^2 + \alpha q\left(r\gamma\right)^2}{\left(r+\alpha q\beta^2\right)^2}\right)x^2 + \alpha q\sigma^2, \\ &= \left(\frac{q\left(r+\alpha q\beta^2\right) + r\alpha q\gamma^2}{r+\alpha q\beta^2}\right)x^2 + \alpha q\sigma^2, \\ &= K_2x^2 + \alpha q\sigma^2, \end{aligned}$$

$x \in X$, donde

$$K_2 = \left(\frac{q\left(r+\alpha q\beta^2\right) + r\alpha q\gamma^2}{r+\alpha q\beta^2}\right).$$

Para $n = 3$, tenemos que

$$\begin{aligned} v_3(x) &= \min_{A(x)}\left\{qx^2 + ra^2 + \alpha E\left[v_2(\gamma x + \beta a + \xi)\right]\right\}, \\ &= \min_{A(x)}\left\{qx^2 + ra^2 + \alpha E\left[K_2(\gamma x + \beta a + \xi)^2 + \alpha q\sigma^2\right]\right\}, \\ &= \min_{A(x)}\left\{qx^2 + ra^2 + \alpha K_2E[(\gamma x + \beta a + \xi)^2] + \alpha^2K_2q\sigma^2\right\}, \\ &= \min_{A(x)}\left\{qx^2 + ra^2 + \alpha K_2(\gamma x + \beta a)^2 + \alpha K_2\sigma^2 + \alpha^2K_2q\sigma^2\right\}. \end{aligned}$$

Si

$$G_3(x,a) = qx^2 + ra^2 + \alpha K_2(\gamma x + \beta a)^2 + \alpha K_2\sigma^2 + \alpha^2K_2q\sigma^2,$$

$(x,a) \in \mathbb{K}$, derivando con respecto a a e igualando a cero, obtenemos que

$$2ra + 2\alpha K_2\beta(\gamma x + \beta a) = 0,$$

entonces

$$f_3(x) = -\frac{\alpha K_2\beta\gamma}{r+\alpha K_2\beta^2}x,$$

$x \in X$. Sustituyendo f_3 en v_3, llegamos a

$$\begin{aligned} v_3(x) &= qx^2 + r\left(\frac{\alpha K_2\beta\gamma}{r+\alpha K_2\beta^2}x\right)^2 + \alpha K_2\left(\gamma x - \beta\frac{\alpha K_2\beta\gamma}{r+\alpha K_2\beta^2}x\right)^2 \\ &\quad +\alpha K_2\sigma^2 + \alpha^2 K_2 q\sigma^2, \end{aligned}$$

$x \in X$, simplificando, tenemos que

$$\begin{aligned} v_3(x) &= \left(\frac{q\left(r+\alpha K_2\beta^2\right) + r\alpha K_2\gamma^2}{r+\alpha K_2\beta^2}\right)x^2 + \alpha K_2\sigma^2 + \alpha^2 K_2 q\sigma^2, \\ &= K_3x^2 + E_3, \end{aligned}$$

$x \in X$, donde

$$\begin{aligned} K_3 &= \left(\frac{q\left(r+\alpha K_2\beta^2\right) + r\alpha K_2\gamma^2}{r+\alpha K_2\beta^2}\right), \\ E_3 &= \alpha K_2\sigma^2 + \alpha^2 K_2 q\sigma^2 = K_2\alpha\sigma^2\left(1+\alpha q\right). \end{aligned}$$

Continuando con el procedimiento encontramos que

$$v_n(x) = K_n x^2 + E_n,$$

$x \in X$, donde

$$\begin{aligned} K_n &= \left(\frac{q\left(r+\alpha K_{n-1}\beta^2\right) + r\alpha K_{n-1}\gamma^2}{r+\alpha K_{n-1}\beta^2}\right), \\ &= \frac{qr + K_{n-1}(\alpha q\beta^2 + r\alpha\gamma^2)}{r+\alpha K_{n-1}\beta^2} = \frac{P+K_{n-1}Q}{R+K_{n-1}S}, \\ E_n &= \alpha K_{n-1}\sigma^2 + \alpha^2 K_{n-1} q\sigma^2 = K_{n-1}\alpha\sigma^2\left(1+\alpha q\right). \end{aligned}$$

Con lo cual concluimos que el lema es valido. ■

Corolario 4.2.3 *La función de valor y la política óptima para el problema LQ están dadas por*

$$V^*(x) = Kx^2 + E,$$

y

$$f^*(x) = \frac{-\alpha\beta\gamma K}{r+\alpha\beta^2 K}x,$$

$x \in X$, *donde* $E = K\alpha\sigma^2\left(1+\alpha q\right)$ *y* $K = lim_{n\to\infty}K_n$, *respectivamente.*

Demostración. Debido a que $v_n(x) \to V^*(x), x \in X$, solo necesitamos probar que $K_n \to K$. Para ello procedemos de la forma siguiente, sea

$$g(z) = \frac{P + zQ}{R + zS},$$

$z \in \mathbb{C}$ y buscamos el punto fijo de g, es decir, $g(z) = z$,

$$\begin{aligned} \frac{P + zQ}{R + zS} &= z, \\ P + zQ &= zR + z^2 S, \end{aligned}$$

equivalentemente

$$z^2 S + z(R - Q) - P = 0. \tag{4.4}$$

Entonces las raíces son z_1 y z_2 dadas por

$$z_{1,2} = \frac{-(R - Q) \pm \sqrt{(R - Q)^2 + 4SP}}{2S},$$

además, es posible probar que $z_1 z_2 < 0$. Ahora, si

$$W = \frac{P + zQ}{R + zS},$$

y usando a z_j, $j = 1, 2$, tenemos que

$$\begin{aligned} W - z_j &= \frac{P + zQ}{R + zS} - z_j, \\ &= \frac{P + zQ - Rz_j - Szz_j}{R + zS}. \end{aligned}$$

También (4.4) implica que

$$Rz_j = z_j Q + P - Sz_j^2.$$

Luego

$$\begin{aligned} W - z_j &= \frac{Q(z - z_j) - Sz_j(z - z_j)}{R + zS}, \\ &= \frac{(Q - Sz_j)(z - z_j)}{R + zS}, \end{aligned}$$

entonces

$$\begin{aligned}\frac{W-z_1}{W-z_2} &= \frac{\dfrac{(Q-Sz_1)(z-z_1)}{R+zS}}{\dfrac{(Q-Sz_2)(z-z_2)}{R+zS}},\\ &= \frac{(Q-Sz_1)}{(Q-Sz_2)}\frac{(z-z_1)}{(z-z_2)},\\ &= \lambda\frac{z-z_1}{z-z_2},\end{aligned}$$

donde $\lambda = \dfrac{Q-Sz_1}{Q-Sz_2}$.

Observemos que $g(K_{n-1}) = K_n$, entonces para $n = 1, 2, ...,$

$$\frac{K_n - z_1}{K_n - z_2} = \lambda\frac{K_{n-1}-z_1}{K_{n-1}-z_2}.$$

Ahora, usando el hecho de que $K_0 = 0$, y desarrollando la relación anterior tenemos que

$$\begin{aligned}\frac{K_1-z_1}{K_1-z_2} &= \lambda\frac{z_1}{z_2},\\ \frac{K_2-z_1}{K_2-z_2} &= \lambda^2\frac{z_1}{z_2}.\end{aligned}$$

En general,

$$\frac{K_n-z_1}{K_n-z_2} = \lambda^n\frac{z_1}{z_2},$$

lo cual implica

$$\begin{aligned}K_n - z_1 &= \lambda^n\frac{z_1}{z_2}(K_n - z_2),\\ K_n &= \lambda^n\frac{z_1}{z_2}(K_n-z_2)+z_1,\end{aligned}$$

equivalentemente

$$\begin{aligned}K_n(1-\lambda^n\frac{z_1}{z_2}) &= (1-\lambda^n)z_1,\\ K_n(1-\lambda^n\frac{z_1}{z_2}) &= (1-\lambda^n)z_1.\end{aligned}$$

Finalmente

$$K_n = \frac{(1-\lambda^n)z_1}{(1-\lambda^n\frac{z_1}{z_2})}.$$

Notemos que $|\lambda| < 1$, es decir, $\left|\frac{Q-Sz_1}{Q-Sz_2}\right| < 1$, por lo que

$$\lim_{n\to\infty} K_n = \lim_{n\to\infty} \frac{(1-\lambda^n)z_1}{(1-\lambda^n\frac{z_1}{z_2})} = z_1,$$

de la cual concluimos que z_1 es la única raíz positiva y que $K_n \to K$, con $K = z_1$ cuando $n \to \infty$. Por lo tanto,

$$V^*(x) = \lim_{n\to\infty} v_n(x) = Kx^2 + E, \tag{4.5}$$

$x \in X$, donde

$$E = K\alpha\sigma^2\left(1+\alpha q\right).$$

Por último, sustituyendo (4.5) en la EPD se obtiene la política óptima f^*. ■

Ejemplo 2

Por ultimo, se presenta un ejemplo modificado del problema LQ. Consideremos la dinámica del sistema como

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

donde $x_t \in X = \mathbb{R}$ y la sucesión de v.a $\{\xi_t\}$ se supone i.i.d y con valores en $\mathbb{R}$. Suponemos también, que $E(\xi_t) = 0$ y $E(\xi_t^2) = \sigma^2$.

Consideremos el costo

$$c(x,a) = x^2 + sen(a) + 1,$$

donde $a \in A = A(x) = [-\frac{\pi}{2}, \frac{\pi}{2}]$. Notemos que c es no negativa, continua e inf-compacta.

Además, si suponemos que la densidad común de las v.a's ξ, Δ es continua, tenemos que la ley de transición Q es fuertemente continua.

Utilizando el método de aproximaciones sucesivas, tenemos que

$$v_0(x) = 0,$$

$$v_1(x) = \min_{a\in[-\frac{\pi}{2},\frac{\pi}{2}]}\{x^2 + sen(a) + 1\},$$

notemos que la función $g(a) := x^2 + sen(a) + 1$, tiene un mínimo en $f_1(x) = -\frac{\pi}{2}$, entonces

$$v_1(x) = x^2,$$

$x \in X$. Ahora calculamos a v_2

$$\begin{aligned} v_2(x) &= \min_{a\in[-\frac{\pi}{2},\frac{\pi}{2}]}\{x^2 + sen(a) + 1 + \alpha E(v_1(x+\xi))\}, \\ &= \min_{a\in[-\frac{\pi}{2},\frac{\pi}{2}]}\{x^2 + sen(a) + 1 + \alpha E((x+\xi)^2)\}, \\ &= \min_{a\in[-\frac{\pi}{2},\frac{\pi}{2}]}\{x^2 + sen(a) + 1 + \alpha E(x^2 + 2x\xi + \xi^2)\}, \\ &= \min_{a\in[-\frac{\pi}{2},\frac{\pi}{2}]}\{x^2 + sen(a) + 1 + \alpha x^2 + 2xE(\xi) + \alpha E(\xi^2)\}, \\ &= \min_{a\in[-\frac{\pi}{2},\frac{\pi}{2}]}\{x^2 + sen(a) + 1 + \alpha x^2 + \alpha\sigma^2\}, \end{aligned}$$

$x \in X$, y tomando a $f_2(x) = -\frac{\pi}{2}$, se tiene que

$$\begin{aligned} v_2(x) &= x^2 + \alpha x^2 + \alpha\sigma^2 \\ &= x^2(1+\alpha) + \alpha\sigma^2, \end{aligned}$$

$x \in X$. Veamos otra iteración

$$\begin{aligned} v_3(x) &= \min_{[-\frac{\pi}{2},\frac{\pi}{2}]}\{x^2 + sen(a) + 1 + \alpha E(v_2(x+\xi))\}, \\ &= \min_{[-\frac{\pi}{2},\frac{\pi}{2}]}\{x^2 + sen(a) + 1 + \alpha E[(x+\xi)^2(1+\alpha) + \alpha\sigma^2]\}, \\ &= \min_{[-\frac{\pi}{2},\frac{\pi}{2}]}\{x^2 + sen(a) + 1 + \alpha(1+\alpha)x^2 + \alpha(1+\alpha)E[\xi^2] + \alpha^2\sigma^2\}, \\ &= \min_{[-\frac{\pi}{2},\frac{\pi}{2}]}\{sen(a) + 1 + (1+\alpha+\alpha^2)x^2 + (\alpha+2\alpha^2)\sigma^2\}, \end{aligned}$$

$x \in X$, y nuevamente tomando a $f_3(x) = -\frac{\pi}{2}$, tenemos que

$$v_2(x) = (1+\alpha+\alpha^2)x^2 + (\alpha+2\alpha^2)\sigma^2.$$

En general tendremos

$$f_n(x) = -\frac{\pi}{2},$$

y

$$v_n(x) = x^2\sum_{k=0}^{n}\alpha^k + \sigma^2\sum_{k=1}^{n}k\alpha^k,$$

$x \in X$. Sabemos que $v_n(x) \to V^*(x), x \in X$, entonces

$$\begin{aligned} \lim_{n\to\infty} v_n(x) &= x^2 \lim_{n\to\infty} \sum_{k=0}^{n} \alpha^k + \sigma^2 \lim_{n\to\infty} \sum_{k=1}^{n} k\alpha^k, \\ &= \frac{x^2}{1-\alpha} + \sigma^2 \lim_{n\to\infty} \sum_{k=1}^{n} k\alpha^k, \end{aligned}$$

$x \in X$. Por lo tanto,

$$V^*(x) = \frac{x^2}{1-\alpha} + \frac{\sigma^2\alpha}{(1-\alpha)^2},$$

y $f^*(x) = -\pi/2, x \in X$.

4.3. Conclusiones

En la tesis se trató con la teoría de Procesos de Decisión de Markov (PDM). En ella se trabajaron problemas a tiempo discreto, con horizonte infinito. El criterio de rendimiento que se utilizó, para evaluar la calidad de las políticas admisibles fue el de costo total descontado. El análisis se presento para problemas con horizonte finito e infinito.

La teoría de PDM es presentada en el Capítulo 2, donde se dan de forma extendida los resultados principales. También se proporcionan las referencias para justificar los resultados principales de esta teoría. En el Capítulo 3 se procedió a dar la teoría referente a Procesos de Decisión de Markov con horizonte finito y una sección donde se ejemplifica la teoría con dos ejemplos clásicos de la teoría de PDM. El primero de ellos es referente a la teoría de inventarios. El trabajo realizado con este ejemplo fue con respecto a la solución del problema, ya que aunque su solución era conocida, el método que se uso para resolverlo en la tesis difiere a los conocidos en la literatura. El segundo ejemplo es un problema relacionado con reemplazamiento de máquinas, en este caso se presentó un ejemplo particular del cual se da la solución de forma explícita. El Capítulo 4 está relacionado con la teoría de PDM con horizonte infinito, y al igual que el Capítulo 3 se presentan dos ejemplos. Principalmente se resuelve un problema lineal cuadrático verificando todas las hipótesis de PDM y presentando la solución de forma detallada.

En resumen la tesis se enfoca a dar una primera introducción de la teoría de Procesos de Decisión de Markov para problemas con costo descontados y algunas de sus aplicaciones.

Los problemas que se consideran como consecuencia del trabajo, y se espera continuar su análisis, son los siguientes.

(a) Modelar usando PDM problemas de aplicación cotidiana y dar una solución. Como un antecedente se cuenta con la siguiente referencia donde se presenta una aplicación de PDM a movimiento de objetos controlados (por ejemplo robots que hacen uso de inteligencia artificial, ver [11]).

(b) Hacer un estudio de la diferenciabilidad en PDMs. En el siguiente sentido: la diferenciabilidad de la función de valor y de la política óptima. La razón de este problema es debido a que si sabemos que la solución del problema de control es diferenciable es posible presentar una solución más simple del problema de control, sin necesidad de realizar tantos cálculos como se hace ver en los ejemplos presentados, como referencia se tiene a [6] y [8].

(c) Estudiar otros métodos de solución a parte de Programación Dinámica como por ejemplo la Ecuación de Euler (ver [7]), para la cual sólo existen versiones (formales) para PDMs deterministas.

Capítulo 5

APÉNDICE

5.1. Apéndice A

5.1.1. Definiciones

Definición 5.1.1 *Sea (X, τ)un espacio topológico, la mínima σ-álgebra que contiene a τ es la σ-álgebra de Borel, es decir, la σ-álgebra generada por τ. La denotaremos por $\mathcal{B}(X)$.*

De aquí en adelante, cuando hablemos de conjuntos o funciones medibles, se entenderán como Borel medibles.

Definición 5.1.2 *X es un espacio de Borel, si X es un subconjunto de Borel de un espacio métrico, separable y completo.*

Por ejemplo, $\mathbb{R}^n$ con la topología usual, un espacio métrico compacto, por mencionar algunos.

Definición 5.1.3 *Un kérnel estocástico definido sobre X dado Y es una función $P(\cdot\,|\,\cdot)$ tal que:*

1. $P(\cdot\,|\,y)$ es una medida de probabilidad en X, para cada $y \in Y$
2. $P(B\,|\,\cdot)$ es una variable aleatoria (función medible) en Y para cada $B \in \mathcal{B}(X)$.

La familia de todos los kérneles estocásticos lo denotaremos por $P(X\,|\,Y)$.

5.1.2. Funciones Semicontinuas

Definición 5.1.4 *Sean (X, d) un espacio métrico y $v : X \to \mathbb{R} \cup \{+\infty\}$ una función tal que $v(x) < \infty$ para al menos una $x \in X$,. la función v se dice ser semicontinua inferiormente (l.s.c) en $x \in X$ si*

$$\liminf_{n\to\infty} v(x_n) \geq v(x),$$

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

Si v es l.s.c para toda $x \in X$, se llama inferiormente semicontinua (l.s.c).

La función v se dice ser superiormente semicontinua (u.s.c) si $-v$ es (l.s.c).

Proposición 5.1.5 ***A1*** *Las siguientes proposiciones son equivalentes:*

a) v es l.s.c;

b) El conjunto $epi(v) := \{(x, \lambda) \in X \times \mathbb{R} \mid v(x) \leq \lambda\}$, llamado el epígrafe de v, es cerrado;

c) Toda sección inferior $S_\lambda(v)$ es cerrado, donde

$$S_\lambda(v) := \{x \in X \mid v(x) \leq \lambda\}, \lambda \in \mathbb{R}.$$

$L(X)$ denota a la familia de todas las funciones v (l.s.c) y acotadas inferiormente sobre X.

Proposición 5.1.6 ***A2*** *$v \in L(X)$, si y sólo si, existe una sucesión $\{v_n\}$ de funciones continuas y acotadas sobre X, tales que $v_n \uparrow v$.*

Proposición 5.1.7 ***A3*** *Si $v, v_1, ..., v_n \in L(X)$, entonces*

a) $\alpha v, v_1 + ... + v_n$ y $\min_i v_i \in L(X)$ con $\alpha \geq 0$.

b) Si X es compacto, entonces v alcanza su mínimo, esto es, existe $x^* \in X$ tal que $v(x^*) = \min_X v(x)$.

Las demostraciones de A1-A3, pueden encontrarse en Ash [1] y Bertsekas and Shreve [3].

5.1.3. Teoremas Básicos de Integración

Supongamos que $(X, \mathcal{F}, \mu)$ es un espacio de medida fijo, y todas las funciones definidas en X son reales.

Teorema 5.1.8 ***A5 Teorema de la convergencia monótona***
Sea $\{f_n\}$ una sucesión de funciones medibles no negativas definidas en X, supongamos que convergen a una función medible f, entonces

$$\int f d\mu = \lim_{n\to\infty} \int f_n d\mu.$$

Nota: la sucesión $\{f_n\}$ puede converger a f casi donde quiera, y el teorema anterior sigue valiendo.

Teorema 5.1.9 ***A6 Teorema de la convergencia Dominada***
Sean $\{f_n\}$ una sucesión de funciones medibles y g una función medible y μ-integrable definidas en X, tales que $|f_n| \leq g$, para toda n. Si $f_n \to f$ casi en todas partes, entonces, f es μ-integrable y

$$\int f d\mu = \lim_{n\to\infty} \int f_n d\mu.$$

5.2. Apéndice B

5.2.1. Esperanza Condicional

Sean $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad, $\Im$ una σ-álgebra contenida en $\mathcal{F}$, y ξ una variable aleatoria. Si ξ es P-integrable, entonces definimos a la Esperanza Condicional de ξ dado $\Im$, como la variable aleatoria denotada por $E(\xi|\,\Im)$, y es tal que:

1. $E(\xi|\,\Im)$ es $\Im$-medible
2. Para todo $A \in \Im$, se tiene que $\int\limits_A E(\xi|\,\Im)dP = \int\limits_A \xi dP$.

Si $C \in \mathcal{F}$, entonces definimos a la probabilidad condicional de C dado $\Im$, como $P(C|\,\Im) := E(I_C|\,\Im)$.

Proposición 5.2.1 ***B1*** *Sean ξ y η variables aleatorias P-integrables sobre $(\Omega, \mathcal{F}, P)$, y sean $\Im$ y $\Im'$ σ-álgebras contenidas en $\mathcal{F}$, entonces*

a) Si ξ es una constante k, entonces $E(\xi|\,\Im) = k$,

b) $E(\xi + \eta|\,\Im) = E(\xi|\,\Im) + E(\eta|\,\Im)$,

c) $E[E(\xi|\,\Im)] = E(\xi)$,

d) Si ξ es $\Im$-medible, entonces $E(\xi\eta|\,\Im) = \xi E(\eta|\,\Im)$, en particular, $E(\xi|\,\Im) = \xi$,

e) Si $\Im \subset \Im'$, entonces $E(\xi|\,\Im) = E[\,E(\xi|\,\Im)|\,\Im'] = E[\,E(\xi|\,\Im')|\,\Im]$,

f) si $\xi_n \geq 0$, y $\xi_n \uparrow \xi$, entonces .$E(\xi_n|\,\Im) \uparrow E(\xi|\,\Im)$,

g) Si $\xi_n \geq 0$, entonces $E(\sum_{n=1}^{\infty} \xi_n \mid \Im) = \sum_{n=1}^{\infty} E(\xi_n|\,\Im)$.

5.3. Apéndice C

5.3.1. Kérneles Estocásticos

Sean X y A espacios de Borel.

Definición 5.3.1 ***C1*** *Un kérnel estocástico sobre X dado A, es una función $P(\cdot|\,\cdot)$ tal que:*

a) $P(\cdot|\,y)$ es una medida de probabilidad sobre X para cada $y \in Y$,

b) $P(B|\,\cdot)$ es una función medible sobre A para cada $B \in \mathcal{B}(X)$.

$P(X|\,A)$ denota al conjunto de todos los kérneles estocásticos sobre X dado Y.

$L(X)$ denota el conjunto de funciones (l.s.c) y acotadas sobre X.

$M(X)$ denota al conjunto de funciones medibles sobre X.

$M_b(X)$ denota al conjunto de funciones medibles y acotadas sobre X.

$C_b(X)$ denota al conjunto de funciones continuas y acotadas sobre X.

Proposición 5.3.2 ***C2*** *Si $v \in M_b(X \times Y)$, entonces la función*

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

Demostración. ver Bertsekas and Shreve [3]. ■

Definición 5.3.3 ***C3.*** *El kérnel estocástico $P \in P(X|\,A)$ es*

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

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

Es claro que fuertemente continuo implica débilmente continuo.

Proposición 5.3.4 ***C4*** *Las siguientes proposiciones son equivalentes:*

a) P es fuertemente continuo;

b) La función $y \to \int_X v(x)P(dx|\,y)$ es l.s.c, para cada $v \in M_b(X)$.

Proposición 5.3.5 *c) $P(B|\,\cdot)$ es continua sobre A, para todo $B \in \mathcal{B}(X)$.*

Además, las siguientes proposiciones son equivalentes

d) P es débilmente continuo;

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

La demostración de (d) implica (e) es directa de la definición C3(a) y la proposición A2, y la inversa, se sigue del hecho de que v es continua, si y sólo si, v es (l.s.c) y (u.s.c).

Proposición 5.3.6 ***C6*** *Si $P \in P(X|\,A)$ es débilmente continuo y $v \in C_b(X \times A)$ (respectivamente l.s.c y acotada inferiormente), entonces la función $y \to \int v(x)P(dx|\,y)$ es continua y acotada (resp. es l.s.c y acotada inferiormente) sobre A.*

Demostración. ver Bertsekas and Shreve [3]. ■

Proposición 5.3.7 ***C7 (Teorema de Ionesco-Tulcea)***

Sea $X_0, X_1, ...,$ una sucesión de espacios de Borel y, para $n = 0, 1, ...,$ definimos $Y_n := X_0 \times X_1 \times ... \times X_n$ y $Y := \prod_{n=0}^{\infty} X_n$. Sea v una medida de probabilidad arbitraria sobre X_0 y, para cada $n = 0, 1, ...,$ $P_n(dx_{n+1}|\,y_n)$ es un kérnel estocástico sobre X_{n+1} dado Y_n. Entonces existe una única medida

de probabilidad P_v sobre Y tal que, para cada rectángulo medible $B_0 \times ... \times B_n$ en Y_n,

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

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

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

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

Demostración. ver Ash [1] y Bertsekas and Shreve [3]. ■

5.4. Apéndice D

5.4.1. Multifunciones y Selectores.

Sean X y A espacios de Borel.

Una multifunción φ de X a A es una función tal que para toda $x \in X$ su imagen $\varphi(x)$ es un subconjunto no vacío de A, es decir, $\varphi : X \to \mathbf{P}(A)$, donde $\mathbf{P}$ denota al conjunto potencia de A. La grafica de φ es el subconjunto de $X \times A$ definido como

$$graf(\varphi) = \{\, (x, a) |\, x \in X, a \in \varphi(x)\}.$$

Definición 5.4.1 ***D1*** *Sea ψ una multifunción de X a A, ψ es*

a) Borel Medible, si $\psi^{-1}(B)$ es Borel medible en X para cada conjunto abierto B en A.

b) Semicontinua por arriba (u.s.c), $\psi^{-1}(C)$ si es cerrado en X para cada conjunto cerrado C en A.

c) Semicontinua por abajo (l.sc), $\psi^{-1}(B)$ si es abierto en X para cada conjunto abierto B en A.

d) Cerrada, si $\psi(x)$ es cerrado para toda $x \in X$.

e) Compacta, si $\psi(x)$ es compacta para toda $x \in X$.

Suponemos que la multifunción ψ es Borel medible, $v : graf(\psi) \to \mathbb{R}$ es una función medible y para cada $x \in X$.

$$v^*(x) := \inf_{a \in \psi(x)} v(x, a).$$

Además, si $v(x, \cdot)$ alcanza su mínimo en algún punto de $\psi(x)$, escribimos min en lugar de inf.

Definición 5.4.2 ***D2*** *La función $v : graf(\psi) \to \mathbb{R}$ se le llama inf-compacta sobre $graf(\psi)$, si para toda $x \in X$ y $\lambda \in \mathbb{R}$, el conjunto*

$$\{a \in \psi(x) |\, v(x, a) \leq \lambda\},$$

es compacto.

Proposición 5.4.3 D3 *Supongamos que ψ es compacta,*

a) si $v(x, \cdot)$ es (.s.c sobre $\psi(x)$ para cada $x \in X$, entonces existe un selector $f \in \mathbb{F}$ tal que para todo $x \in X$

$$v(x, f(x)) = v^*(x) = \min_{\psi(x)} v(x, a),$$

y v^* es medible.

b) si ψ es u.s.c y v es l.s.c y acotada inferiormente sobre $graf(\psi)$, entonces existe un selector $f \in \mathbb{F}$ tal que la relación anterior se cumple y v^* es l.s.c y acotada inferiormente en X.

Proposición 5.4.4 ***D4*** *Supongamos que $graf(\psi)$ es un subconjunto de Borel de $X \times A$, y que v es l.s.c, acotada inferiormente e inf-compacta sobre $graf(\psi)$, entonces*

a) existe un selector $f \in \mathbb{F}$ tal que

$$v(x, f(x)) = v^*(x) = \min_{\psi(x)} v(x, a).$$

b) Si agregamos que la multifunción

$$x \to \psi*(x) := \{a \in \psi(x) |\, v^*(x) = v(x, a)\}$$

es l.s.c, entonces v^* es l.s.c.

Bibliografía

[1] R. B. Ash and Doléans-Dade C. A., Probability and Measure Theory. Academic Press Elsevier, San Diego, 2005.

[2] D. P. Bertsekas, Dynamic Programming: Deterministic and Stochastic Models. Prentice-Hall, Inc., New Jersey, 1987.

[3] D. P. Bertsekas and Shreve S. E., Stochastic Optimal Control: The Discrete-Time Case. Athena Scientific, 1996.

[4] R. Bellman, Dynamic Programming. Dover, 2003.

[5] H. Cruz-Suárez, Análisis de problemas de control estocástico vía problemas de control determinista. Memorias de la Tercera Conferencia Iberoamericana en Sistemas, Cibernética e Informática CISCI'04, Orlando, Fl., V. 3, pp. 66-71, 2004.

[6] H. Cruz-Suárez, Procesos de decisión de Markov descontados: soluciones óptimas mediante problemas de control determinista diferenciables, Revista Iberoamericana de Sistemas Cibernética e Informática, V. 2 N. 1, 2005. (Disponible en la página web: http://www.iiisci.org/journal/risci/.)

[7] H. Cruz-Suárez and R. Montes-de-Oca, Discounted Markov control processes induced by deterministic systems. Kybernetika (Prague), V. 42 N. 6, pp. 647-664, 2006.

[8] H. Cruz-Suárez and R. Montes-de-Oca, An envelope theorem and some applications to discounted Markov decision processes. Por aparecer en Mathematical Methods of Operations Research, Springer Verlag (2006).

[9] B. Heer and A. Maussner, Dynamic General Equilibrium Modelling: Computational Method and Application, Springer-Verlag, Berlin, 2005.

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

[11] E. López, Barea R., Escudero M. S., Navegación topológica mediante POMDPs incorporando información visual. Departamento de Electrónica, Universidad de Alcalá, 2003.

[12] M. L. Puterman, Markov Decision Processes: Discrete Stochastic Dynamic Programming, Wiley, New York, 1994.

[13] N. L. Stokey and R. E. Lucas, Recursive Methods in Economic Dynamics. Harvard University Press Massachusetts, 1989.

[14] K. R. Stromberg, Introduction to Classical Real Analysis. Wadsworth International Group, Belmont, California, 1981.