Benemérita Universidad Autónoma de Puebla

facultad de ciencias físico matemáticas
licenciatura en matemáticas aplicadas

El Problema de Control Óptimo de Markov Bajo el Criterio de Rendimiento Descontado

T E S I S

que para obtener el título de
Licenciado en Matemáticas Aplicadas

presenta
José Alberto Tepox Méndez

director de tesis
Dr. Hugo Adán Cruz Suárez

PUEBLA, PUE. ENERO DE 2017

A mis padres, Paula y Benjamín

Agradecimientos

A mis padres, Paula Méndez y Benjamín Tepox, por todo el apoyo, el cariño y la confianza que depositaron en mí; éste y cada uno de mis logros no habrían sido posibles sin ellos.

A mis hermanos, en orden de aparición, Rocío, Verónica, Raúl y Sandra, que a pesar de la distancia, sé que siempre puedo contar con ellos.

A todos los amigos y colegas con los tuve la oportunidad de intercambiar ideas, aprender y compartir experiencia durante mi vida universitaria. De manera particular a Rubén, Rocío, Isa, Erick, Marychel, Ovando, Reynaldo, Jeanille, Yasmín y Óscar.

A mi asesor de tesis y profesor en varios de mis cursos, Dr. Hugo A. Cruz Suárez, al cual respeto y admiro no sólo por su excelente calidad como profesor e investigador, sino por su gran calidad humana.

A mi tutor académico, Dra. Lidia A. Hernández, quien actúo como guía y consejera desde el primer día de la licenciatura.

A mis sinodales, Dr. Francisco S. Tajonar Sanabria, Dra. Hortensia J. Reyes Cervantes y Dr. Víctor Hugo Vázquez Guevara, por su trabajo en la revisión de esta tesis y sus aportaciones a la misma.

A los que olvidé en el momento de escribir estas líneas pero que estuvieron a largo de este viaje de 5 años.

A todos ustedes, muchas gracia.

Índice general

Introducción

En la presente tesis se abordan los Procesos de Decisión de Markov (PDM) que pertenecen al área de Teoría de Control Estocástico. Un PDM a tiempo discreto es un modelo para la toma secuencial de decisiones cuando el proceso de interés es observado de forma periódica considerando que presenta incertidumbre en sus transiciones, es decir, se encuentra influenciado por un ruido aleatorio. Este tipo de modelos aparecen en numerosas ciencias aplicadas como son Ingeniería, Economía, Finanzas, etc. En el caso de Economía, los PDM se aplican a problemas de control de población y optimización de recursos [18], mientras que en Ingeniería se desarrollan modelos de inteligencia artificial [14], por mencionar algunas aplicaciones.

De manera general, un PDM modela un sistema estocástico cuyos estados son observados de manera periódica por un controlador. La dinámica de un PDM puede ser descrita de la siguiente manera: en cada periodo de observación t, con $t = 0, 1, \ldots$, el controlador decide el control que aplicará dependiendo del estado actual del sistema, como consecuencia, se paga un costo que depende del estado del sistema y el control aplicado, posteriormente el sistema se traslada a un nuevo estado. En el siguiente periodo de observación, el sistema se encuentra en el nuevo estado y la dinámica anterior se repite. A la sucesión de controles aplicados en cada periodo se le conoce como política. Para evaluar de alguna manera la calidad de cada política se define el criterio de rendimiento o función objetivo. De esta manera, el problema de control óptimo consiste en encontrar una política que optimice el criterio de rendimiento; a dicha política se le denomina óptima. Un procedimiento para hallar la política óptima está basado en el principio de optimalidad de Bellman conocido como Programación Dinámica.

La teoría de los PDM y de solución del problema de control óptimo vía Programación Dinámica ha sido ampliamente estudiada por autores como Bertsekas [4], Hernández Lerma [11], [12], entre otros. Uno de los resultados principales de tales publicaciones son los teoremas que garantizan la existencia de la solución del problema de control óptimo. De esta forma, una de las aportaciones de esta tesis consiste en presentar demostraciones alternativas a las propuestas en los textos de estos autores. Para ello se desarrolla la teoría de los PDM, se abordan los problemas con horizonte finito e infinito considerando el criterio de costo descontado, posteriormente se enuncian y demuestran los teoremas de validación de programación dinámica para ambos casos.

Una aportación más de esta tesis está enfocada en ilustrar la teoría desarrollada presentando un par de ejemplos. En el caso de los problemas con horizonte finito se aborda el problema clásico conocido como Lineal Cuadrático (LQ), retomando la formulación multidimensional que se hace en [4], pero desarrollando cuidadosamente la verificación de condiciones que justifiquen la implementación de Programación Dinámica para hallar una solución analítica. En el caso de problemas con horizonte infinito se formula un modelo propio de esta tesis que representa un problema de consumo e inversión; al igual que en el caso anterior, se verifican las condiciones para la implementación de PD y se muestra que, a pesar de que se garantiza la existencia de la política óptima, su representación explícita sólo puede hacerse a través de métodos numéricos.

La tesis se organiza de la siguiente forma. En el primer capítulo se definen los Modelos de Decisión de Markov, se definen y clasifican las políticas de control, se hace un breve recordatorio de la propiedad de Markov para definir formalmente el Proceso de Decisión de Markov, finalmente se presentan los criterios de rendimiento y el horizonte del proceso para concluir con la definición del problema de control óptimo. En el segundo capítulo se plantea el problema de control con horizonte finito, se enuncia y demuestra el teorema que garantiza la existencia de una política óptima para dicho problema y que acredita el uso de programación dinámica para su solución; además, se presentan la condiciones de estructura necesarias que debe satisfacer el modelo de decisión de Markov; finalmente se desarrolla la teoría sobre el

problema Lineal Cuadrático así como un ejemplo numérico del mismo. Por último, en el tercer capítulo se presenta la teoría referente a la teoría de PDM con costo descontado y horizonte infinito, una vez más, se enuncia y prueba el teorema que garantiza la existencia de la política óptima y se presenta un problema de consumo e inversión.

Capítulo 1

Procesos de Decisión de Markov

En este capítulo se define formalmente un Proceso de Decisión de Markov a tiempo discreto. Para ello se presenta el Modelo de Decisión de Markov y se desarrollan conceptos relacionados al mismo, como son: criterios de rendimiento, políticas de control, entre otros. Luego se construye el espacio y la medida de probabilidad necesarios para definir el proceso y finalmente se concluye con la presentación del problema de control óptimo haciendo algunos comentarios adicionales sobre problemas con horizonte aleatorio.

1.1. Modelo de Decisión de Markov

Definición 1.1. Un *Modelo de Decisión de Markov (MDM)* estacionario, a tiempo discreto, consiste en una quíntupla de la forma:

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

donde

1. X es un espacio de Borel no vacío, llamado *espacio de estados*.
2. A es un espacio de Borel no vacío, llamado *conjunto de acciones* o *controles*.
3. $\{A(x) : x \in X\}$ es una familia de subconjuntos medibles de A, donde

$A(x)$ denota al *espacio de acciones o controles admisibles* cuando el sistema se encuentra en el estado $x \in X$. El conjunto $\mathbb{K}$ de parejas estado-acción admisibles, está definido por

$$\mathbb{K} \coloneqq \{(x, a) : x \in X,\ a \in A(x)\},$$

el cual, se supone, es un conjunto medible del espacio producto $X \times A$.

4. Q es un kérnel estocástico (ver Definición B.3) sobre X dado $\mathbb{K}$ llamado *ley de transición*.

5. $c : \mathbb{K} \to \mathbb{R}$ es una función medible llamada *función de costo de un paso*.

Un MDM estacionario a tiempo discreto representa a un sistema estocástico controlado que es observado de manera periódica en los instantes de tiempo $t = 0, 1, \ldots$. La dinámica que describe este sistema estocástico puede ser detallada de la siguiente manera: si el sistema se encuentra en el estado $x_t = x \in X$ al instante t y se aplica la acción (o control) $a_t = a \in A(x)$, entonces ocurren dos cosas:

1. Se paga un costo $c(x, a)$.

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

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

donde, $\mathcal{B}(X)$ denota a la *σ-álgebra de Borel* de X. Así, una vez hecha la transición a un nuevo estado, se elige una nueva acción y la dinámica anterior se repite.

Observación 1.1. El modelo de control de Markov de la Definición 1.1 es llamado *estacionario* debido a que sus componentes X, A, Q y c no dependen del parámetro del tiempo t. Si el modelo depende del tiempo, es decir, si el modelo es de la forma

$$(X_t, A_t, \{A_t(x) : x \in X_t\}, Q_t, c_t), \qquad t = 0, 1, \ldots,$$

entonces es llamado *no estacionario*.

Suposición 1.1. $\mathbb{K}$ contiene la gráfica de una función medible de X a A, es decir, existe $f : X \to A$ función medible, tal que $f(x) \in A(x)$, para cada $x \in X$. El conjunto de estas funciones será denotado por $\mathbb{F}$ y sus elementos serán llamados *selectores de la multifunción* $x \mapsto A(x)$.

1.2. Políticas

Para introducir las políticas de control, considere el MDM de la Definición 1.1 y, para cada $t = 0, 1, \ldots$, defina el *espacio de historias admisibles* hasta el tiempo t, $\mathbb{H}_t$, como $\mathbb{H}_0 \coloneqq X$ y

$$\begin{aligned} \mathbb{H}_t &\coloneqq \mathbb{K}^t \times X \\ &= \mathbb{K} \times \mathbb{H}_{t-1}, \quad \text{para } t = 1, 2, \ldots . \end{aligned}$$

Un elemento h_t de $\mathbb{H}_t$, llamado *t-historia admisible* o simplemente *t-historia*, es un vector de la forma

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

con $(x_i, a_i) \in \mathbb{K}$ para cada $i = 0, \ldots, t-1$, y $x_t \in X$. Observe que, para cada $t \geq 1$, $\mathbb{H}_t$ es un subespacio de $\overline{\mathbb{H}}_t \coloneqq (X \times A)^t \times X$ y $\overline{\mathbb{H}}_0 \coloneqq \mathbb{H}_0$.

Definición 1.2. Una *política de control aleatorizada* o simplemente *política de control*, es una sucesión $\pi = \{\pi_t : t = 0, 1, \ldots\}$ de kérneles estocásticos π_t sobre el conjunto de acciones A dado $\mathbb{H}_t$, que satisface

$$\pi_t(A(x_t)|h_t) = 1, \quad \forall\ h_t \in \mathbb{H}_t,\ t = 0, 1, \ldots .$$

El conjunto de todas las políticas es denotado por Π.

En términos generales, una política $\pi = \{\pi_t\}$ puede interpretarse como una sucesión $\{a_t\}$ de variables aleatorias sobre A, tales que, para cada t-historia y $t = 0, 1, \ldots$, 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 arbitratia, la acción que se elige al tiempo t del proceso, es una realización de la variable aleatoria a_t la cual depende de toda la historia h_t.

La familia de kérneles estocásticos sobre A dado X será denotada por $P(A|X)$. Por otro lado, se define Φ como el conjunto de todos los kérneles estocástico $\varphi \in P(A|X)$ tales que, para cada $x \in X$ se satisface que $\varphi(A(x)|x) = 1$.

Definición 1.3. Una política $\pi = \{\pi_t\} \in \Pi$ se dice:

- **Markoviana Aleatorizada**, si existe una sucesión de kérneles estocásticos $\{\varphi_t\} \subset \Phi$ tal que

$$\pi_t(\cdot|h_t) = \varphi_t(\cdot|x_t), \quad \forall\ h_t \in \mathbb{H}_t,\ t = 0, 1, \ldots.$$

  Al conjunto de las políticas markovianas aleatorizadas se denota por Π_{RM}.

- **Markoviana Aleatorizada Estacionaria**, si existe $\varphi \in \Phi$ kérnel estocástico, tal que

$$\pi_t(\cdot|h_t) = \varphi(\cdot|x_t), \quad \forall\ h_t \in \mathbb{H}_t,\ t = 0, 1, \ldots.$$

  El conjunto de estas políticas es denotado por Π_{RS}.

- **Determinista**, si existe una sucesión $\{g_t\}$ de funciones medibles $g_t : X \to A$, tales que, para cada $h_t \in \mathbb{H}_t$ y $t = 0, 1, \ldots$, se tiene que $g_t(h_t) \in A(x_t)$ y $\pi_t(\cdot|h_t)$ está *concentrada* en $g_t(h_t)$, es decir,

$$\pi_t(C|h_t) = I_C[g_t(h_t)], \quad \forall\ C \in \mathcal{B}(A).$$

  Al conjunto de las políticas deterministas se le denota por Π_D.

- **Determinista Markoviana**, si existe una sucesión $\{f_t\}$ de funciones $f_t \in \mathbb{F}$ tal que $\pi_t(\cdot|h_t)$ está concentrada en $f_t(x_t) \in A(x_t)$, para toda $h_t \in \mathbb{H}_t$ y $t = 0, 1, \ldots$.
  Este conjunto de políticas es denotado por Π_{DM}.

- **Determinista Estacionaria**, si existe una función $f \in \mathbb{F}$ tal que, $\pi_t(\cdot|h_t)$ está concentrada en $f(x_t) \in A(x_t)$, para toda $h_t \in \mathbb{H}_t$ y $t = 0, 1, \ldots$.
  De esta manera, Π_{DS} denota al conjunto de políticas deterministas estacionarias.

Observación 1.2. La Suposición 1.1 garantiza que $\mathbb{F}$ es no vacío, y por lo tanto, Π también lo es. Esto se debe a que cada $f \in \mathbb{F}$ puede ser identificada por un kérnel estocástico φ, de la siguiente manera:

$$\varphi(C|x) \coloneqq I_C(f(x)),$$

para cada $C \in \mathcal{B}(A)$ y $x \in X$.

Por otro lado,

$$\Pi_{RS} \subset \Pi_{RM} \subset \Pi,$$

y a su vez

$$\Pi_{DS} \subset \Pi_{DM} \subset \Pi_D \subset \Pi.$$

1.3. Proceso de Decisión de Markov

Antes de construir el Proceso de Decisión de Markov, se define el *Proceso de Markov* y con él la propiedad del mismo nombre, la cual de manera general señala que, dado el estado actual del sistema, el pasado no tiene influencia en el estado futuro.

Definición 1.4. Sea $\{R_t\}$ una sucesión de kérneles estocásticos en $P(X|X)$ y considere $\{x_t\}$ un proceso estocástico sobre X. Se dice que $\{x_t\}$ es un *Proceso de Markov no homogéneo* con kérneles de transición $\{R_t\}$, si para cada $B \in \mathcal{B}(X)$, $i_0, \ldots, i_{t-1}, x \in X$ y $t = 0, 1, \ldots$

$$P(x_{t+1} \in B | x_0 = i_0, \ldots, x_t = x) = P(x_{t+1} \in B | x_t = x) = R_t(B|x_t = x).$$

La primera igualdad es conocida como *Propiedad de Markov*.

Observación 1.1. En la definición anterior, si $\{R_t\}$ es invariante en el tiempo, es decir, si R_t es igual a R para toda $t = 0, 1, \ldots$, con $R \in P(X|X)$, entonces $\{x_t\}$ se dice un *Proceso de Markov homogéneo* con kérnel de transición R.

1.3.1. Construcción

Sea $(\Omega, \mathcal{F})$ un espacio medible conformado por el espacio muestral canónico $\Omega \coloneqq \overline{\mathbb{H}}_\infty = (X \times A)^\infty$ y $\mathcal{F}$ su correspondiente σ-álgebra producto. Los elementos de Ω son de la forma $\omega = (x_0, a_0, x_1, a_1, \ldots)$, con $x_t \in X$ y $a_t \in A$

para toda $t = 0, 1, \ldots$; las proyecciones $x_t,\ a_t$ de Ω sobre los conjuntos X y A son llamadas variables de estado y acción, respectivamente.

Note que $\mathbb{H}_\infty = \mathbb{K}^\infty \subset \Omega$, donde $\mathbb{H}_\infty$ es llamado espacio de historias admisibles, y para cada $(x_0, a_0, x_1, a_1, \ldots) \in \mathbb{H}_\infty$ se tiene que $(x_t, a_t) \in \mathbb{K}$, para cada $t = 0, 1, \ldots$.

Sean $\pi \in \Pi$ una política de control arbitraria y $x_0 = x \in X$ un *estado inicial*. Entonces, por el Teorema de Ionescu-Tulcea (ver Teorema B.1, Apéndice B), existe una única medida de probabilidad P_x^π sobre $(\Omega, \mathcal{F})$, más aún, para toda $B \in \mathcal{B}(X)$, $C \in \mathcal{B}(A)$, $x \in X$, $a \in A$ y $h_t \in \mathbb{H}_t$

$$P_x^\pi(a_t \in C | h_t) = \pi_t(C|h_t), \tag{1.1}$$

$$P_x^\pi(x_{t+1} \in B | h_t, a_t = a) = Q(B|x_t = x, a_t = a). \tag{1.2}$$

El operador esperanza inducido por la medida de probabilidad P_x^π es denotado por E_x^π.

Definición 1.5. El proceso estocástico $(\Omega, \mathcal{F}, P_x^\pi, \{x_t\})$ es llamado *Proceso de Decisión de Markov* (PDM) a tiempo discreto.

Observación 1.3. En general, en lugar de un estado inicial $x \in X$, se puede dar una medida de probabilidad ν sobre X, llamada *distribución inicial*, la cual, para cada $B \in \mathcal{B}(X)$ satisface que

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

La ecuación (1.2) es una condición similar a la propiedad de Markov, ya que, como se observa en el lado derecho de la ecuación, la probabilidad de transición no depende de la historia completa (h_t) del proceso, sino unicamente del estado x y la acción a, correspondientes al instante anterior. Sin embargo, en general, el proceso $\{x_t\}$ no es Markoviano en el sentido usual. A pesar de ello, si la elección de la política π se limita al conjunto Π_{RM} (o Π_{DM}), entonces $\{x_t\}$ resulta ser un proceso de Markov. En la siguiente proposición se prueba esta afirmación.

Proposición 1.1. Considere un proceso de decisión de Markov y ν una distribución inicial. Si $\pi = \{\varphi_t\} \in \Pi_{RM}$, entonces $\{x_t\}$ es un Proceso de

Markov no homogéneo con kérneles de transición $\{Q(\cdot|\cdot,\varphi_t)\}$, es decir, para cada $B \in \mathcal{B}(X)$, $i_0,\ldots,i_{t-1},x \in X$ y $t=0,1,\ldots,$

$$\begin{aligned} P_\nu^\pi(x_{t+1}\in B|x_0=i_0,\ldots,x_t=x) &= P_\nu^\pi(x_{t+1}\in B|x_t=x) \\ &= Q(B|x_t,\varphi_t), \end{aligned} \tag{1.3}$$

en donde,

$$Q(\cdot|x,\varphi) \coloneqq \int_A Q(\cdot|x,a)\varphi(da|x). \tag{1.4}$$

En particular, si $\pi=\{f_t\}\in\Pi_{DM}$, la igualdad de (1.3) se conserva para los kérneles de transición $Q(\cdot|\cdot,f_t)$. Más aún, para políticas estacionarias $\varphi^\infty\in\Pi_{RS}$ y $f^\infty\in\Pi_{DS}$, $\{x_t\}$ es un Proceso de Markov homogéneo con kérnel de transición $Q(\cdot|\cdot,\varphi)$ y $Q(\cdot|\cdot,f)$, respectivamente.

Demostración. Sea $\pi=\{\pi_t\}\in\Pi$ una política arbitraria, entonces se satisface que

$$P_\nu^\pi(x_{t+1}\in B|h_t)=\int_A Q(B|x_t=x,a_t=a)\pi_t(da|h_t),$$

para cada $B\in\mathcal{B}(X)$, $h_t=(i_0,j_0,\ldots,i_{t-1},j_{t-1},x)\in\mathbb{H}_t$ y $t=0,1,\ldots$.

En efecto, por las propiedades de Esperanza Condicional (Proposición D.1), se tiene que

$$\begin{aligned} P_\nu^\pi(x_{t+1}\in B|h_t) &= E_\nu^\pi[P_\nu^\pi(x_{t+1}\in B|h_t,a_t=a)|h_t] \\ &= E_\nu^\pi[Q(x_{t+1}\in B|x_t=x,a_t=a)|h_t] \\ &= \int_A Q(B|x_t=x,a_t=a)\pi_t(da|h_t). \end{aligned}$$

En particular, si $\pi=\{\varphi_t\}\in\Pi_{RM}$, por la ecuación (1.4), se verifica que

$$\begin{aligned} P_\nu^\pi(x_{t+1}\in B|h_t) &= \int_A Q(B|x_t=x,a_t=a)\varphi_t(da|h_t) \\ &= Q(B|x_t=x,\varphi_t). \end{aligned}$$

Así, ocupando nuevamente la Proposición D.1, se concluye que

$$\begin{aligned} P_\nu^\pi(x_{t+1} \in B | x_0 = i_0, \ldots, x_t = x) &= E_\nu^\pi[P_\nu^\pi(x_{t+1} \in B | h_t) | x_0 = i_0, \ldots, x_t = x] \\ &= Q(B | x_t = x, \varphi_t). \end{aligned}$$

De manera similar

$$\begin{aligned} P_\nu^\pi(x_{t+1} \in B | x_t = x) &= E_\nu^\pi[P_\nu^\pi(x_{t+1} \in B | h_t) | x_t = x] \\ &= Q(B | x_t = x, \varphi_t), \end{aligned}$$

concluyendo con ello la demostración.

■

1.4. El Problema de Control Óptimo

Una vez que se ha construido el proceso de decisión de Markov, se prosigue a completar la descripción del *problema de control óptimo*; para ello se requiere de una *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.

Sin embargo, antes de definir estos criterios, se debe considerar el *horizonte de planeación* u *horizonte del problema*, es decir, el periodo de tiempo en el cual se observará el proceso.

En general, se pueden definir dos tipos de horizonte, el caso *finito* y el caso *infinito*. El caso finito es utilizado cuando el interés principal es conocer el comportamiento del proceso durante un intervalo de tiempo determinado, mientras que el caso infinito se emplea cuando no se puede precisar una cota a priori al término de la planeación.

Los ejemplos con horizonte finito abundan en los procesos asociados al ámbito financiero, ya que en ellos se consideran periodos de planeación estrictos donde se obervan, por ejemplo, el nivel de inventario de una empresa, las utilidades que percibe un inversionista al final de un contrato, o incluso, problemas de un sólo paso donde se decide entre ejercer o no el derecho que concede una opción Europea sobre algún bien subyacente [3]. Por otro

lado, existen procesos que, aunque no se "observan" de manera infinita, se pueden considerar de horizonte infinito debido al gran número de periodos que se estudian, por ejemplo, al considerar redes de sensores inalámbricos en donde el periodo entre observaciones es de fracciones de segundo [1], o bien, el nivel de combustible en cada hora en una red de gasolineras [17].

De la misma manera, puede considerarse que el horizonte de planeación es aleatorio, sin embargo, como se desarrollará más adelante, bajo las condiciones necesarias el problema con horizonte aleatorio puede asociarse con un problema de horizonte finito (o infinito) equivalente.

A continuación se definen los Criterios de Costo Total Acumulado y de Costo Total Descontado, los cuales serán utilizados para el desarrollo de esta tesis.

Definición 1.6. Considere un PDM fijo, un conjunto de políticas Π y $N \in \mathbb{N} \cup \{+\infty\}$ arbitrario. Se dice que la función $V : \Pi \times X \to \mathbb{R}$, es un *criterio de rendimiento* o *función objetivo* con horizonte de planeación N, de tipo

- **Costo Total Acumulado**, si para cada $x \in X$ y $\pi \in \Pi$,

$$V(\pi, x) \coloneqq E_x^{\pi}\left[\sum_{t=0}^{N-1} c(x_t, a_t)\right];$$

  o es de tipo

- **Costo Total Descontado**, si para cada $x \in X$ y $\pi \in \Pi$

$$V(\pi, x) \coloneqq E_x^{\pi}\left[\sum_{t=0}^{N-1} \alpha^t c(x_t, a_t)\right],$$

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

Observe que el criterio de costo total acumulado no es más que la suma de los costos que se van generando en cada etapa del proceso, mientras que el costo total descontado introduce un factor de descuento constante α para "traer los costos a valor presente". En ambos casos se toma la esperanza inducida por la medida de probabilidad P_x^{π}, debido a que se trata de variables aleatorias.

Observación 1.2. Cada uno de estos criterios tiene una función equivalente cuando el MDM es no estacionario, en el sentido de que c depende del tiempo, y cuando el factor de descuento no es constante. Estas consideraciones en la definición se realizarán cuando sean necesarias para el desarrollo de este trabajo.

Definición 1.7. La *función de valor óptimo*, o simplemente *función de valor*, se define, para cada $x \in X$, como:

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

De esta manera, el *Problema de Control Óptimo* consiste en hallar la política $\pi^* \in \Pi$ que satisfaga:

$$V^*(x) = V(\pi^*, x), \quad \forall\ x \in X. \tag{1.5}$$

La política $\pi^* \in \Pi$ que satisface (1.5) es conocida como *política óptima*.

1.4.1. Horizonte Aleatorio

Como se mencionó anteriormente, para la formulación del problema de control óptimo debe definirse el horizonte de planeación. Sin embargo, en algunas aplicaciones es adecuado considerar un horizonte dado por una variable aleatoria que dependa de algún evento de interés relacionado con el proceso o independiente del mismo. Numerosos ejemplos de estos problemas son encontrados en aplicaciones de finanzas, en los denominados Problemas de Paro Óptimo (ver [3]), mientras que la teoría es ampliamente desarrollada en trabajos como [8], [13] y [19], por mencionar algunos.

De esta manera, el problema con horizonte aleatorio puede plantearse como sigue: considere el proceso de decisión de Markov de la Definición 1.5 con horizonte de planeación τ, donde τ es una variable aleatoria (v.a.) sobre $(\Omega', \mathcal{F}')$ con soporte un subconjunto de $\mathbb{N} \cup \{+\infty\}$. Se define el criterio de rendimiento como

$$V^\tau(\pi, x) \coloneqq E\left[\sum_{t=0}^{\tau} c(x_t, a_t)\right], \tag{1.6}$$

para cada $\pi \in \Pi$ y $x \in X$, donde E representa la esperanza respecto a la

distribución conjunta del proceso $\{(x_t, a_t)\}$ y τ.

A continuación se muestra cómo la función definida en (1.6) es equivalente a otro criterio de rendimiento con horizonte infinito (o finito), dependiente de la relación de τ con el PDM.

Horizonte independiente del proceso

Considere a τ como una v.a. con soporte $\{0, 1, \ldots, T\}$, donde $T \in \mathbb{N} \cup \{\infty\}$, y defina $\rho_n \coloneqq P(\tau = n)$, además, se asume que para cada $x \in X$ y $\pi \in \Pi$, el proceso inducido $\{(x_t, a_t)\}$ es independiente de τ. Note que, bajo esta suposición se satisface lo siguiente:

$$\begin{aligned}
E\left[\sum_{t=0}^{\tau} c(x_t, a_t)\right] &= E\left[E\left[\sum_{t=0}^{\tau} c(x_t, a_t)\middle| \tau\right]\right] \\
&= \sum_{n=0}^{T} \rho_n E_x^{\pi}\left[\sum_{t=0}^{\tau} c(x_t, a_t)\middle| \tau = n\right] \\
&= \sum_{n=0}^{T} \rho_n \sum_{t=0}^{n} E_x^{\pi}[c(x_t, a_t)] \\
&= \sum_{t=0}^{T}\sum_{n=t}^{T} \rho_n E_x^{\pi}[c(x_t, a_t)] \\
&= E_x^{\pi}\left[\sum_{t=0}^{T} P_t c(x_t, a_t)\right],
\end{aligned}$$

donde $P_t \coloneqq \sum_{n=t}^{T} \rho_n = P(\tau \geq t),\ t = 0, 1, 2, \ldots, T$.

Con esto, el problema de control óptimo con horizonte aleatorio τ es equivalente al problema de control con horizonte de planeación T y costo no homogéneo $P_t c(\cdot)$. Así mismo, el problema resultante será de horizonte finito o infinito dependiendo si $T < \infty$ o $T = \infty$, respectivamente. El caso cuando $T < \infty$ puede consultarse en [8].

Horizonte dependiente del proceso

Para este caso, se considera a $K \in \mathcal{B}(X)$ y se define $\tau \coloneqq \text{ínf}\{i \in \mathbb{N} : x_i \in K\}$. Bajo esta definición, τ es un *tiempo de paro* (ver Definición D.4, Apéndice D) respecto a la filtración natural $\{\mathcal{F}_t\}$ generada por el proceso $\{(x_t, a_t)\}$. Así, la función objetivo para el problema de control óptimo, considerando el criterio de costo total, está dada por:

$$V^{\tau}(\pi, x) \coloneqq E\left[\sum_{t=0}^{\tau-1} c(x_t, a_t)\right], \tag{1.7}$$

para cada $x \in X$ y $\pi \in \Pi$.

Observe que, dada la función de costo en un paso $c : \mathbb{K} \to \mathbb{R}$, se puede definir la función

$$c'_t(x, a) \coloneqq c(x, a) I_{\{t<\tau\}},$$

para cada $(x, a) \in \mathbb{K}$, de manera que la función objetivo dada por la ecuación (1.7), es equivalente a la función

$$\begin{aligned} V_1^{\tau}(\pi, x) &\coloneqq E\left[\sum_{t=0}^{\infty} c'_t(x_t, a_t)\right] \\ &= E\left[\sum_{t=0}^{\infty} c(x_t, a_t) I_{\{i<\tau\}}\right]. \end{aligned}$$

Con esto, el problema de control con horizonte aleatorio, se ha convertido en un problema de horizonte infinito con función de costos c'. A pesar de ello, este nuevo problema no es menos complejo, por lo que no se desarrolla en lo posterior, sin embargo, en [7] puede consultarse la solución cuando el problema considera el criterio de costo total descontado.

En esta tesis se presentan los resultados para los problemas de control con horizonte determinista, sin embargo, el material que se desarrolla en el siguiente capítulo puede ser aplicado para la solución del problema con horizonte aleatorio independiente del proceso, al menos cuando el soporte de la variable aleatoria τ es finito.

Capítulo 2

Problemas con Horizonte Finito

Considere un Modelo de Decisión de Markov a tiempo discreto no estacionario, dado por:

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

así como, el criterio de rendimiento de costo total acumulado con horizonte finito N, definido por:

$$V(\pi, x) \coloneqq E_x^{\pi}\left[\sum_{t=0}^{N-1} c_t(x_t, a_t)\right].$$

El hecho de considerar un MDM no estacionario está motivado por la última sección del capítulo anterior, en donde se señala que algunos problemas de control óptimo con horizonte aleatorio pueden ser equivalentes a problemas con función de costos no homogénea.

El objetivo principal de este capítulo es presentar cómo la programación dinámica puede ser utilizada para hallar la solución del problema de control óptimo. Este método permitirá encontrar tanto a la función de valor óptimo V^*, como a la política óptima π^*. Sin embargo, para hallar π^* se requiere de condiciones generales para la función de costos de manera que se pueda garantizar la solución de problema de control óptimo; estas condiciones se

presentan en la Sección 2.1. Posteriormente se enuncia y se demuestra el Teorema de Programación Dinámica para PDM; se dan algunas formas alternativas de la ecuación de programación dinámica y finalmente se presenta un ejemplo clásico en la teoría de los PDM conocido como Problema Lineal Cuadrático. Este problema se resuelve utilizando programación dinámica y se da un ejemplo numérico para visualizar una de sus características más importantes.

2.1. Selección Medible

Suposición 2.1. (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

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

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,

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

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

En la mayoría de los problemas aplicados, la Suposición de Selección Médible puede ser verificada directamente, sin embargo, desde un punto de vista teórico, es conveniente tener condiciones generales bajo las cuales esta hipótesis es verdadera. Estas condiciones son obtenidas usualmente a partir de los teoremas de selección medible (Apéndice C). De la misma forma, se puede probar que bajo cualquiera de las Condiciones 2.1, 2.2 o 2.3, la Suposición de Selección Medible se verifica.

Condición 2.1. a) La función de costos c es *semicontinua inferiormente* (l.s.c., ver Apéndice A), acotada inferiormente e *inf-compacta* (ver Apéndice C) sobre $\mathbb{K}$.

b) La ley de transición Q es:

i) *Débilmente continua*, es decir, la función

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

es continua y acotada en $\mathbb{K}$ para cada función v continua y acotada sobre X, ó

ii) *Fuertemente continua*, es decir, la función

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

es continua y acotada en $\mathbb{K}$ para cada función v medible y acotada sobre X.

Teorema 2.1. Bajo la Condición 2.1 se tiene que, para cualquier función $u : X \to \mathbb{R}$ medible y no negativa, la Suposición de Selección Medible se satisface.

Demostración. Para una demostración detallada consultar [12]. ■

Además de la Condición 2.1, existen otras que permiten que la Suposición de Selección Médible se cumpla, por ejemplo:

Condición 2.2.

a) El conjunto de controles admisibles $A(x)$ es compacto para cada $x \in X$.

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 una de las siguientes condiciones:

   i) v' es l.s.c. sobre $A(x)$ para cada $x \in X$ y cada función v continua y acotada sobre X.

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

Condición 2.3.

a) $A(x)$ es compacto para cada $x \in X$, y la multifunción $x \mapsto A(x)$ es u.s.c. (ver Definición C.1, Apéndice C)

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

c) La ley de transición Q es

   i) Débilmente continua, ó

   ii) Fuertemente continua.

Teorema 2.2. Para cualquier función $u : X \to \mathbb{R}$ medible y no negativa, las Condiciones 2.2 y 2.3 implican la condición de selección medible. Más aún, bajo 2.2 (I) y 2.3 (I) es suficiente tomar a u como no negativa y l.s.c.; y bajo 2.3 (II) la función u^* es l.s.c.

Demostración. La demostración puede encontrarse en [12]. ■

2.2. Teorema de Programación Dinámica

El siguiente teorema presenta un método de solución para el problema de control óptimo con horizonte finito garantizando la existencia de una política óptima determinista de Markov. Una de las aportaciones de esta tesis sobre este teorema es la demostración, la cual es distinta a la presentada en referencias como [12] y [20], ya que se basa en la construcción de un proceso martingala.

Teorema 2.3. Sean $V_0, V_1, ..., V_N$ funciones sobre X definida por

$$V_N(x) := 0, \tag{2.1}$$

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

$$V_t(x) := \min_{x \in A(x)} \left\{ c_t(x,a) + \int_X V_{t+1}(y) Q(dy|x,a) \right\}. \tag{2.2}$$

Bajo la Suposición de Selección Medible (Suposición 2.1), estas funciones son medibles y para cada $t = 0, 1, \ldots, N-1$, existen selectores $f_t \in \mathbb{F}$ con $f_t(x) \in A(x)$, tal que

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

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 $x \in X$ se verifica que $V^*(x) = V(\pi^*, x) = V_0(x)$.

La relación (2.2) es conocida como la *Ecuación de Programación Dinámica* (EPD) junto con su condición inicial (2.1).

Demostración. Sea $\pi = \{\pi_t\}$ una política arbitraria y $x \in X$ un estado inicial cualquiera.

Considere el proceso $\{V_t(x_t) : t = 0, 1, \ldots\}$, con $V_t(x_t) = 0$ para cada $t = N, N+1, \ldots$. Claramente este nuevo proceso es adaptado a la filtración natural generada por el proceso original, la cual se denota por $\{\mathcal{F}_t\}$.

Ahora, defina $Y_0 \coloneqq V_0(x) - E_x^\pi [V_0(x)]$, y para cada $n = 1, 2, \ldots$, sea

$$Y_n \coloneqq \sum_{t=1}^{n} \{V_t(x_t) - E_x^\pi[V_t(x_t)|\mathcal{F}_{t-1}]\}.$$

Por el Ejemplo D.1 del Apéndice D, $\{Y_n, n \geq 0\}$ es una martingala, además, $E_x^\pi[Y_n] = 0$ para cada $n = 0, 1, \ldots$.

Por lo anterior, se satisface lo siguiente:

$$\begin{aligned} E_x^\pi[Y_{N-1}] &= E_x^\pi \left[\sum_{t=1}^{N-1} \{V_t(x_t) - E_x^\pi[V_t(x_t)|\mathcal{F}_{t-1}]\} \right] \\ &= \sum_{t=1}^{N-1} \{E_x^\pi[V_t(x_t)] - E_x^\pi[E_x^\pi[V_t(x_t)|\mathcal{F}_{t-1}]]\} \\ &= \sum_{t=1}^{N-1} \{E_x^\pi[V_t(x_t)] - E_x^\pi[V_t(x_t)|x_t, a_t]\}. \end{aligned} \tag{2.3}$$

Note que, para cada $t = 0, 1, \ldots, N-1$ y cada $z \in X$, se tiene:

$$V_t(z) \leq c_t(z, a) + \int_X V_{t+1}(y) Q(dy|z, a), \quad \forall a \in A(z), \tag{2.4}$$

también

$$E_x^\pi[V_{t+1}(x_{t+1})] = \int_X V_{t+1}(y) Q(dy|x_t, a_t).$$

Por lo anterior, partiendo de (2.3), se satisfacen las siguientes relaciones:

$$\begin{aligned} E_x^{\pi}[Y_{N-1}] &\leq \sum_{t=1}^{N-1} \{E_x^{\pi}[c_t(x_t,a_t) + E_x^{\pi}[V_{t+1}(x_{t+1})|x_t,a_t]] - E_x^{\pi}[V_t(x_t)]\} \quad (2.5) \\ &= \sum_{t=1}^{N-1} E_x^{\pi}[c_t(x_t,a_t)] + \sum_{t=1}^{N-1} \{E_x^{\pi}[V_{t+1}(x_{t+1})] - E_x^{\pi}[V_t(x_t)]\} \\ &= V(\pi,x) - E_x^{\pi}[c_0(x,a_0)] + E_x^{\pi}[V_N(x_N)] - E_x^{\pi}[V_1(x_1)] \\ &= V(\pi,x) - E_x^{\pi}[c_0(x,a_0) + E_x^{\pi}[V_1(x_1)]]. \quad (2.6) \end{aligned}$$

Además, como $E_x^{\pi}[Y_{N-1}] = 0$, por (2.4) y (2.6), es válido que

$$V_0(x) \leq E_x^{\pi}\left[c_0(x,a_0) + \int_X V_1(y)Q(dy|x,a_0)\right] \leq V(\pi,x).$$

Es decir, dado que π es una política arbitraria y x es un estado inicial cualquiera, se concluye que

$$V_0(x) \leq V(\pi,x), \quad \forall\ \pi \in \Pi,\ \forall\ x \in X.$$

Ahora se probará que la política π^* es óptima. Por el resultado anterior, bastará probar que $V(\pi^*,x) = V_0(x)$, para cada $x \in X$.

Considere la política determinista de Markov $\pi^* = \{f_0, f_1, ..., f_{N-1}\}$, donde, para cada $t = 0, 1, \ldots, N-1$, $f_t(x) \in A(x)$ es tal que

$$V_t(x) = c_t(x, f_t(x)) + \int_X V_{t+1}(y)Q(dy|x, f_t(x)). \quad (2.7)$$

para cada $x \in X$. Luego, sea $x \in X$ arbitrario, a partir de las ecuaciones (2.3) y (2.7), se satisface lo siguiente:

$$\begin{aligned} E_x^{\pi^*}[Y_{N-1}] &= \sum_{t=1}^{N-1} \left\{E_x^{\pi^*}[V_t(x_t)] - E_x^{\pi^*}[V_t(x_t)]\right\} \\ &= \sum_{t=1}^{N-1} \left\{E_x^{\pi^*}[c_t(x_t, f_t(x_t)) + E_x^{\pi^*}[V_{t+1}(x_{t+1})|x_t, f(x_t)]] - E_x^{\pi^*}[V_t(x_t)]\right\} \\ &= V(\pi^*,x) - E_x^{\pi^*}[c_0(x, f_0(x)) + E_x^{\pi^*}[V_1(x_1)]] \\ &= V(\pi^*,x) - V_0(x). \end{aligned}$$

Una vez más, dado que $E_x^{\pi^*}[Y_{N-1}] = 0$ y $x \in X$ es arbitrario, se concluye que

$$V_0(x) = V(\pi^*, x), \qquad \forall\ x \in X.$$

Es decir, la política π^* es óptima. ■

Observación 2.1. Como se mencionó anteriormente, con el Teorema 2.3 se puede garantizar la existencia de la solución al problema con horizonte aleatorio independiente del proceso, ya que, como se probó en la Subsección 1.4.1, el problema mencionado es equivalente al problema con criterio de rendimiento dado por

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

donde $P_t := \sum_{n=t}^{T} \rho_n = P(\tau \geq t),\ t = 0, 1, 2, \ldots, T.$

Cabe señalar, que las hipótesis del Teorema 2.3 se satisfacen siempre que:

1. $T \in \mathbb{N}$;
2. c es l.s.c., acotada inferiormente e inf-compacta sobre $\mathbb{K}$, y;
3. Q es débilmente continua ó fuertemente continua.

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

A menudo es conveniente reescribir la EPD en otras formas convenientes y apropiadas. En esta sección se presentan algunas de las formas alternativas más frecuentes y que serán de utilidad en los ejemplos que se desarrollan al final del capítulo.

2.3.1. Modelo de Ecuaciones en Diferencia

En algunas aplicaciones la ley de transición Q es inducida por una ecuación en diferencias estocásticas dada de manera general por:

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

para $t = 0, 1, \ldots$, y $x_0 = x$ conocido. Donde $\{\xi_t\}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas (i.i.d.) que toman valores en un espacio de Borel S, con función de distribución común μ e independientes del estado inicial x_0. La función que describe la dinámica del sistema, $F : \mathbb{K} \times S \to X$, es una función medible conocida. En este caso, para cada $B \in \mathcal{B}(X)$ y $(x, a) \in \mathbb{K}$, la ley de transición Q está dada por:

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

donde E denota la esperanza inducida por la distribución μ. Luego, por el teorema de cambio de variable, para cualquier función v medible sobre X, se tiene

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

en el sentido que, si una de las integrales existe, entonces las otras existen y son iguales.

Observación 2.1. Si ξ_t tiene función de densidad común Δ para cada $t = 0, 1, \ldots$, se satisface que

$$Q(B|x,a) = \int_S I_B(F(x,a,s))\Delta(s)ds,$$

y también

$$E[v(x_{t+1})|x_t = x, a_t = a] = \int v(F(x,a,s))\Delta(s)ds.$$

Por lo anterior, la EPD puede reescribirse, para cada $x \in X$, como

$$V_N(x) = 0,$$

y para cada $t = 0, 1, \ldots, N-1$,

$$\begin{aligned} V_t(x) &= \min_{a \in A(x)} \left\{ c(x,a) + \int_S V_{t+1}(F(x,a,s))\mu(ds) \right\} \\ &= \min_{a \in A(x)} \left\{ c(x,a) + E[V_{t+1}(F(x,a,\xi))] \right\}. \end{aligned}$$

Además, si la función de costo depende de manera explícita de la variable aleatoria ξ, es decir, $c(x,a,\xi)$, entonces

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

2.3.2. Forma Hacia Adelante de la EPD

En las funciones definidas por la ecuación (2.2) del Teorema 2.3, se verifica que V_t depende de V_{t+1} para cada $t = 0, 1, \ldots, N-1$, sin embargo, en algunas ocasiones es conveniente trabajar "hacia adelante".

Para ello, definimos a $v_t \coloneqq V_{N-t}$, para cada $t = 0, 1, \ldots, N$ y escribimos la EPD "hacia adelante" como

$$v_0(x) = 0,$$

y para $t = 1, \ldots, N$,

$$v_t(x) = \min_{a \in A(x)} \left\{ c(x,a) + \int_X v_{t-1}(y)Q(dy|x,a) \right\}. \tag{2.8}$$

Más aún, si $f_t \in \mathbb{F}$ es el selector que minimiza el lado derecho de la EPD, entonces $g_t \coloneqq f_{N-t}$, para $t = 1, \ldots, N$, minimiza el lado derecho de (2.8). Así, en términos de las funciones v_t, la conclusión del Teorema 2.3 puede ser replanteada de la siguiente manera: $\tilde{\pi} = \{g_N, g_{N-1}, \ldots, g_1\}$ es una política

óptima y la función de valor óptimo es

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

es decir, para toda $x \in X$

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

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

2.3.3. Criterio de Costo Descontado

Considere ahora el criterio de rendimiento de Costo Total Descontado para cada $\pi \in \Pi$ y $x \in X$ definido por

$$V(\pi, x) \coloneqq E_x^{\pi} \left[\sum_{t=0}^{N-1} \alpha^t c(x_t, a_t) \right],$$

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

Note que, si $c_t(x, a) \coloneqq \alpha^t c(x, a)$, entonces se tiene el mismo problema para el cual se probó el Teorema 2.3. De esta manera, la EPD es de la forma

$$V_N(x) \coloneqq 0,$$

y para cada $t = 0, 1, \ldots, N-1$,

$$\begin{aligned} V_t(x) &\coloneqq \min_{a \in A(x)} \left\{ \alpha^t c(x, a) + \int_X V_{t+1}(y) Q(dy|x, a) \right\} \\ &= \min_{a \in A(x)} \left\{ c_t(x, a) + \int_X V_{t+1}(y) Q(dy|x, a) \right\} \end{aligned}$$

para cada $x \in X$.

Sin embargo, en algunos textos es común que se haga una transformación en la EPD con el fin de desarrollar la teoría a partir de una función de costos que no dependa de t. Para ello, se define $J_t(\cdot) \coloneqq \alpha^{-1} V_t(\cdot)$, para $t = 0, 1, \ldots, N$, luego, para cada $x \in X$,

$$J_N(x) \coloneqq 0,$$

y para cada $t = 0, 1, \ldots, N-1$,

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

así, el Teorema de Programación Dinámica se sigue cumpliendo para las funciones J_t, pero ahora con un MDM homogéneo.

2.4. Ejemplo: Lineal Cuadrático (LQ)

En esta sección se considera un problema ampliamente estudiado dentro de los PDM, el modelo es conocido como Lineal Cuadrático (LQ, por sus siglas en inglés), el cual lleva ese nombre porque la dinámica que sigue es de tipo lineal, mientras que la función de costos involucra el cuadrado del estado y el cuadrado de la acción.

El modelo LQ es una formulación popular de un problema de regulación, donde el objetivo principal es mantener al sistema cerca del origen de coordenadas. Tales problemas son comunes en la teoría de control automático.

El uso de la función cuadrática es razonable debido a que induce una penalización alta para grandes desviaciones del estado del sistema respecto del origen, mientras que el costo es pequeño si el sistema se encuentra cerca del origen. Cabe señalar que la función cuadrática es usada frecuentemente, incluso cuando ello no es completamente justificado, debido a que conduce hacia una solución analítica simple.

Para dar paso a la formulación del problema, se presenta la dinámica del sistema, la cual está dada por la siguiente ecuación en diferencias

$$x_{t+1} = \gamma_t x_t + \beta_t a_t + \xi_t, \qquad t = 0, 1, \ldots, N-1. \tag{2.9}$$

Mientras que las funciones de costo están dadas por

$$c_t(x_t, a_t) = x_t' q_t x_t + a_t' r_t a_t, \qquad t = 0, 1, \ldots, N-1, \tag{2.10}$$

donde, para cada $t = 0, 1, \ldots, N-1$, x_t y a_t son vectores con entradas reales

de dimensión n y m, respectivamente, es decir, $X = \mathbb{R}^n$ y $A = \mathbb{R}^m$. Las matrices γ_t, β_t, q_t y r_t son conocidas y de dimensión apropiada. Por otro lado, $\{\xi_t\}$ son vectores aleatorios que toman valores en $S = \mathbb{R}^n$. Así mismo, el vector o matriz seguido por un apóstrofe ($'$) corresponde a la transpuesta de dicho elemento.

Con lo anterior, el problema que se desea resolver es minimizar el costo total acumulado, por lo que la función objetivo está dada por:

$$\begin{aligned} V(\pi, x) &:= E_x^\pi \left[\sum_{t=0}^{N-1} c_t(x_t, a_t) \right] \\ &= E_x^\pi \left[\sum_{t=0}^{N-1} (x_t' q_t x_t + a_t' r_t a_t) \right], \end{aligned}$$

para cada $\pi \in \Pi$ y $x \in X$.

Sin embargo, para el desarrollo del problema se presentan las siguientes suposiciones de estructura.

Suposición 2.2. La sucesión $\{\xi_t\}$ está compuesta de vectores aleatorios i.i.d. con función de densidad de probabilidad Δ conocida, más aún, la sucesión $\{\xi_t\}$ es independiente del proceso $\{x_t\}$ y de $\{a_t\}$ para cada $t = 0, 1, \ldots, N-1$; por otro lado, cada ξ_t tiene media cero y segundo momento finito. Además, para cada $t = 0, 1, \ldots, N-1$,

1. El vector a_t no tiene restricciones, es decir, $A(x_t) = A$.
2. Las matrices γ_t y β_t son no singulares.
3. q_t es una matriz simétrica y semidefinida positiva.
4. La matriz r_t es simétrica y definida positiva.

Note que bajo estas suposiciones, se considera que las varibles aleatorias ξ_t se distribuyen según una función de densidad general Δ, sin embargo, en numerosas aplicaciones se supone que cada ξ_t tiene distribución *Gaussiana*, estos problemas son ampliamente conocidos como Problemas Lineal Cuadrático Gaussianos LQG y ejemplos de ellos son mencionados en [9] y [16].

Lema 2.1. Bajo la Suposición 2.2, se satisfacen las Condiciones 2.1, es decir,

1. Para cada $t = 0, 1, \ldots,$ c_t es l.s.c., acotada inferiormente e inf-compacta sobre $\mathbb{K}$.

2. La ley de transición Q, inducida por la ecuación en diferencias (2.9), es fuertemente continua.

Demostración. Para la primera parte, considere $t \in \{0, 1, \ldots, N-1\}$, de esta manera, c_t es acotada inferiormente por cero, ya que q_t es semidefinida positiva y r_t es definida positiva, es decir, para cada $x \in X$, $x'q_tx \geq 0$, y para cada $a \in A\backslash\{0\}$, $a'r_ta > 0$. Por otro lado, es claro que c_t es una función continua sobre $\mathbb{K}$.

Resta verificar que c_t es inf-compacta sobre $\mathbb{K}$, es decir, para todo $x \in X$ y $\lambda \in \mathbb{R}$, el conjunto $A_{t,\lambda}(x) = \{a \in A(x) : c_t(x, a) \leq \lambda\}$ es compacto. Para ello, como $A_{t,\lambda}(x) \subset \mathbb{R}^m$, por el Teorema de Heine-Borel, bastará probar que $A_{t,\lambda}(x)$ es cerrado y acotado.

Sean $x \in X$ y $\lambda \in \mathbb{R}$, por el contrario, suponga que $A_{t,\lambda}(x)$ es no acotado, entonces existe una sucesión $\{a_n\} \subset A_{t,\lambda}(x)$, tal que $\|a_n\| \to \infty$. Luego, por la definición de c_t

$$\lim_{n\to\infty} c_t(x, a_n) = \infty,$$

así, existe $K \in \mathbb{N}$ tal que, para cada $m \geq K$, $c_t(x, a_m) > \lambda$, lo cual es una contradicción, por lo que $A_{t,\lambda}(x)$ es acotado. Ahora, sea $\{a_n\}$ una sucesión en $A_{t,\lambda}(x)$, como éste es un conjunto acotado, existe una subsucesión $\{\widehat{a}_n\}$ tal que $\widehat{a}_n \to a \in A$, entonces, como $0 \leq c_t(x, \widehat{a}_n) \leq \lambda$ y dado que c_t es continua, se tiene que $0 \leq c_t(x, a) \leq \lambda$, es decir, $a \in A_{t,\lambda}(x)$, por lo que se concluye que $A_{t,\lambda}(x)$ es cerrado.

Por lo anterior, como x y λ fueron arbitrarios, se tiene que c_t es inf-compacto sobre $\mathbb{K}$.

Ahora, se probará que Q es fuertemente continua, para ello recordemos que si la dinámica está dada por una ecuación en diferencias, se tiene que

$$Q(B|x, a) = \int_S I_B[F(x, a, s)]\mu(ds).$$

Así, si Δ es la función densidad de ξ, se verifica que

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

con un cambio de variable se tiene

$$Q(B|x,a) = \int_S I_C[u]\Delta(u - \gamma_t x - \beta_t a)du,$$

de ello, como Δ es continua, se garantiza que Q es fuertemente continua. ■

De acuerdo al Lema 2.1, se puede aplicar el algoritmo de Programación Dinámica; para ello definimos las siguientes funciones, para cada $x \in X$,

$$V_N(x) = 0,$$

y para cada $t = 0, 1, \ldots, N-1$,

$$\begin{aligned} V_t(x) &= \min_{a\in A(x)} \left\{ x'q_t x + a'r_t a + \int V_{t+1}(y)Q(dy|x,a) \right\} \\ &= \min_{a\in A(x)} \left\{ x'q_t x + a'r_t a + E[V_{t+1}(\gamma_t x + \beta_t a + \xi_t)] \right\}. \end{aligned} \quad (2.11)$$

Luego, para $t = N-1$, la ecuación (2.11) se escribe como

$$\begin{aligned} V_{N-1}(x) &= \min_{a\in A(x)} \{x'q_{N-1}x + a'r_{N-1}a \\ &\qquad + E[V_N(\gamma_{N-1}x + \beta_{N-1}a + \xi_{N-1})]\} \\ &= \min_{a\in A(x)} \{x'q_{N-1}x + a'r_{N-1}a\} \\ &= x'q_{N-1}x + \min_{a\in A(x)} \{a'r_{N-1}a\}. \end{aligned}$$

Como r_{N-1} es una matriz definida positiva, entonces

$$\min_{a\in A(x)} \{a'r_{N-1}a\} = 0,$$

con $a = 0_m$, donde, 0_m es el vector cero de $\mathbb{R}^m$. Por lo tanto,

$$V_{N-1}(x) = x'q_{N-1}x,$$

con $f^*_{N-1}(x) = 0_m$.

Con lo anterior se puede enunciar y probar la siguiente proposición para caracterizar a los controles óptimos f^*_t y a las funciones V_t de manera general.

Proposición 2.1. Para cada $x \in X$ y $t \in \{0, 1, \ldots, N-2\}$, las funciones $V_t(x)$ están dadas por

$$V_t(x) = x' K_t x + \sum_{j=t}^{N-2} E[\xi_j' K_{j+1} \xi_j],$$

mientras que el control óptimo para la etapa t, está dado por

$$f^*_t(x) = a^*_t = L_t x_t,$$

con $L_t = -(\beta_t' K_{t+1} \beta_t + r_t)^{-1} \beta_t K_{t+1} \gamma_t$. Donde las matrices simétricas K_t están definidas recursivamente como

$$K_{N-1} = q_{N-1},$$

$$K_t = \gamma_t'(K_{t+1} - K_{t+1}\beta_t(\beta_t' K_{t+1}\beta_t + r_t)^{-1}\beta_t' K_{t+1})\gamma_t + q_t.$$

Esta última relación es conocida como la *Ecuación de Riccati a tiempo discreto.*

Demostración. La prueba se hará de manera recursiva hacia atrás sobre t. De está manera, para $t = N-2$ la ecuación (2.11) se escribe de la siguiente forma:

$$\begin{aligned} V_{N-2}(x) &= \min_{a \in A(x)} \{x' q_{N-2} x + a' r_{N-2} a \\ &\qquad + E[V_{N-1}(\gamma_{N-2} x + \beta_{N-2} a + \xi_{N-2})]\} \\ &= \min_{a \in A(x)} \{x' q_{N-2} x + a' r_{N-2} a \\ &\quad + E[(\gamma_{N-2} x + \beta_{N-2} a + \xi_{N-2})' q_{N-1} (\gamma_{N-2} x + \beta_{N-2} a + \xi_{N-2})]\}. \end{aligned}$$

Expandiendo el término dentro de la esperanza, se tiene

$$\begin{aligned} V_{N-2}(x) &= x'q_{N-2}x + \min_{a\in A(x)}\{a'r_{N-2}a \\ &\quad + E[x'\gamma'_{N-2}q_{N-1}\gamma_{N-2}x + 2x'\gamma'_{N-2}q_{N-1}\beta_{N-2}a \\ &\qquad + a'\beta'_{N-2}q_{N-1}\beta_{N-2}a + (x'\gamma'_{N-2}q_{N-1} + a'\beta'_{N-2}q_{N-1})\xi_{N-2} \\ &\qquad + \xi'q_{N-1}(\gamma_{N-2}x + \beta_{N-2}a) + \xi'q_{N-1}\xi]\}. \end{aligned}$$

Después, usando la linealidad de la esperanza y el hecho de que $E[\xi] = 0$, se pueden eliminar los términos $\xi'q_{N-1}(\gamma_{N-2}x + \beta_{N-2}a)$ y $(x'\gamma'_{N-2}q_{N-1} + a'\beta'q_{N-1})\xi$, con ello se obtiene

$$\begin{aligned} V_{N-2}(x) &= x'q_{N-2}x + x'\gamma'_{N-2}q_{N-1}\gamma_{N-2}x + E[\xi'q_{N-1}\xi] \\ &\quad + \min_{a\in A(x)}\{a'r_{N-2}a + 2x'\gamma'_{N-2}q_{N-1}\beta_{N-2}a + a'\beta'q_{N-1}\beta_{N-2}a\}. \end{aligned}$$

De esta manera, derivando el término entre llaves con respecto de a e igualando a cero, se tiene

$$(r_{N-2} + \beta'_{N-2}q_{N-1}\beta_{N-2})a = -\beta'_{N-2}q_{N-1}\gamma_{N-2}x.$$

Dado que β_{N-2} es una matriz no singular y q_{N-1} es semidefinita positiva, entonces $\beta'_{N-2}q_{N-1}\beta_{N-2}$ es semidefinita positiva, más aún, como r_{N-2} es definida positiva, entonces $r_{N-2} + \beta'_{N-2}q_{N-1}\beta_{N-2}$ es definida positiva y por lo tanto, es invertible. Se concluye que el control que minimiza la función está dado por

$$f^*_{N-2}(x) = a^*_{N-2} = L_{N-2}x,$$

con

$$L_{N-2} = -(r_{N-2} + \beta'_{N-2}q_{N-1}\beta_{N-2})^{-1}\beta'_{N-2}q_{N-1}\gamma_{N-2}.$$

Sustituyendo a^*_{N-2} en la expresión que se tiene de V_{N-2}, se verifica que

$$V_{N-2}(x) = x'K_{N-2}x + E[\xi'q_{N-1}\xi],$$

donde la matriz K_{N-2} se define como

$$\begin{aligned} K_{N-2} &= \gamma'_{N-2}(q_{N-1} - q_{N-1}\beta_{N-2}(\beta'_{N-2}q_{N-1}\beta_{N-2} + r_{N-2})^{-1}\beta'_{N-2}q_{N-1})\gamma_{N-2} \\ &\quad + q_{N-2}. \end{aligned}$$

Claramente la matriz K_{N-2} es simétrica, más aún, es semidefinida positiva. En efecto, note que del procedimiento para calcular K_{N-2}, para cada $x \in X$, se tiene

$$\begin{aligned} x'K_{N-2}x &= \min_{a\in A(x)} \{x'q_{N-2}x + a'r_{N-2}a \\ &\qquad + (\gamma'_{N-2}x + \beta_{N-2}a)q_{N-1}(\gamma_{N-2}x + \beta_{N-2}a)\}, \end{aligned}$$

y, dado que q_{N-2}, r_{N-2}, y q_{N-1} son matrices semidefinidas positivas, la expresión entre llaves es no negativa para toda $a \in A$, entonces $x'K_{N-2}x \geq 0$, para toda $x \in X$. Por lo que, K_{N-2} es semidefinida positiva.
Entonces, V_{N-2} es una función cuadrática con una matriz semidefinida positiva más un término constante.

Ahora, suponga que para algún $t \in \{N-3, \ldots, 1\}$, se satisface

$$V_t(x) = x'K_t x + \sum_{j=t}^{N-2} E[\xi'_j K_{j+1}\xi_j]. \tag{2.12}$$

Probemos que la afirmación es válida para $n = t-1$.

Por la ecuación (2.11) se tiene que

$$V_n(x) = \min_{a\in A(x)} \left\{x'q_n x + a'r_n a + E[V_{n+1}(\gamma_n x + \beta_n a + \xi_n)]\right\}.$$

Usando la ecuación (2.12), se verifica que

$$\begin{aligned} V_n(x) &= \min_{a\in A(x)} \{x'q_n x + a'r_n a \\ &\qquad + E[(\gamma_n x + \beta_n a + \xi_n)'K_{n+1}(\gamma_n x + \beta_n a + \xi_n) + \sum_{j=n+1}^{N-2} E[\xi'_j K_{j+1}\xi_j]\}. \end{aligned}$$

Note que la constante $\sum_{j=n+1}^{N-2} E[\xi'_j K_{j+1}\xi_j]$ es irrelevante al momento de minimizar la función entre llaves, por lo que el procedimiento realizado para hallar a_n^* es análogo al caso cuando $t = N-2$ debido a que K_{n+1} es también una matriz simétrica y semidefinida positiva. En conclusión, se tiene que

$$V_n(x) = x'K_n x + E[\xi'_{n+1}K_{n+1}\xi_{n+1}] + \sum_{j=n+1}^{N-2} E[\xi'_j K_{j+1}\xi_j],$$

ó bien

$$V_{t-1}(x) = x'K_{t-1}x + \sum_{j=t-1}^{N-2} E[\xi_j' K_{j+1}\xi_j],$$

y además,

$$f_{t-1}^*(x) = a_{t-1}^* = L_{t-1}x,$$

con K_{t-1} y L_{t-1} como se definieron anteriormente. ■

Aplicando el Teorema 2.3 se puede concluir que la función de valor, para cada $x \in X$, está dada por

$$V(\pi^*, x) = V_0(x) = x'K_0x + \sum_{t=0}^{N-2} E[\xi_t' K_{t+1}\xi_t], \tag{2.13}$$

con $\pi^* = \{f_0^*, f_1^*, \ldots, f_{N-1}^*\}$.

2.4.1. Ejemplo Numérico

Como se ha mencionado, el modelo LQ es una formulación de un problema de regulación, donde se pretende mantener al sistema cerca del origen de coordenadas o de otro punto de interés. Bajo esta premisa, se plantea un modelo LQG que ilustra la afirmación anterior.

Considere el espacio de estados $X = \mathbb{R}^2$, el espacio de acciones $A = \mathbb{R}^2$ y una sucesión $\{\xi_t\}$ de variables aleatorias i.i.d., donde, para cada $t = 0, 1, \ldots$

$$\xi_t \sim \mathcal{N}\left((0,0), \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right), \tag{2.14}$$

es decir, ξ_t sigue una distribución normal bivariada con vector de medias $(0,0)$ y matriz de covarianza $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

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

$$x_{t+1} = \mathbb{I}_2 x_t + \mathbb{I}_2 a_t + \xi_t, \quad t = 0, 1, \ldots, \tag{2.15}$$

mientras que la función de costos está dada por

$$c(x_t, a_t) = x_t' \mathbb{I}_2 x_t + a_t' \mathbb{I}_2 a_t, \quad t = 0, 1, \ldots,$$

donde, $\mathbb{I}_2$ es la matriz identidad de dimensión dos. Con ello, las matrices del problema LQ original, se han considerado homogéneas en el tiempo e idénticas a la matriz $\mathbb{I}_2$. Por otro lado, se dirá que el proceso definido en (2.15) es *no controlado* cuando $a_t = 0$, para cada $t = 0, 1, \ldots$

El problema presentado satisface la Suposición 2.2.

Para definir el problema de control óptimo, se considera un horizonte $N = 51$ y al criterio de rendimiento definido, para cada $\pi \in \Pi$ y $x \in X$, por

$$\begin{aligned} V(\pi, x) &:= E_x^\pi \left[\sum_{t=0}^{N-1} c_t(x_t, a_t) \right] \\ &= E_x^\pi \left[\sum_{t=0}^{N-1} x_t' \mathbb{I}_2 x_t + a_t' \mathbb{I}_2 a_t \right]. \end{aligned}$$

Ahora, se define el *proceso óptimo* como aquel cuya dinámica está dada por

$$x_{t+1} = \mathbb{I}_2 x_t + \mathbb{I}_2 a_t^* + \xi_t, \quad t = 0, 1, \ldots, N-2, \tag{2.16}$$

donde, a_t^* es el control óptimo al tiempo t dado en la Proposición 2.1, es decir,

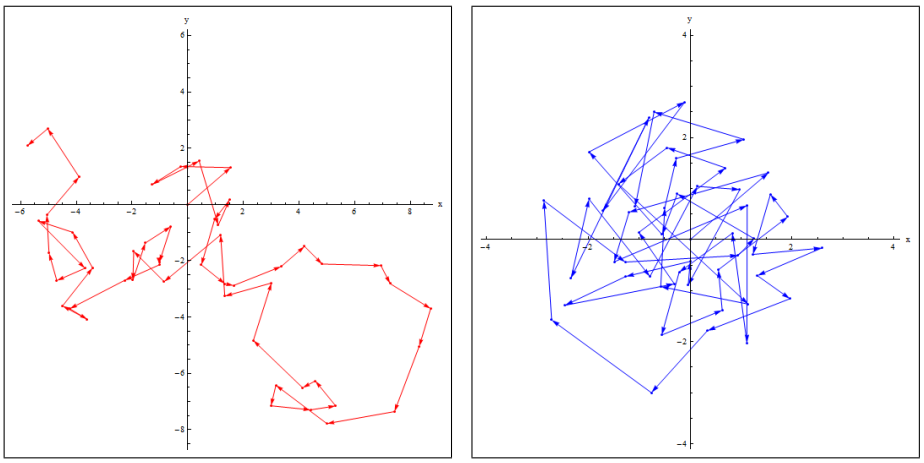
$$a_t^* = L_t x_t,$$

con $L_t = -(K_{t+1} + \mathbb{I}_2)^{-1} K_{t+1}$, y las matrices simétricas K_t definidas recursivamente como

$$K_{N-1} = \mathbb{I}_2,$$

$$K_t = (K_{t+1} - K_{t+1}(K_{t+1} + \mathbb{I}_2)^{-1} K_{t+1}) + \mathbb{I}_2.$$

Con lo anterior, se puede simular el proceso no controlado y el proceso óptimo. Por otro lado, dado que el espacio de estados es $X = \mathbb{R}^2$, se puede graficar una trayectoria del proceso en $\mathbb{R}^3$ y proyectarla sobre el plano. En la Figura 2.1 se observan las proyeccione de la trayectorias simuladas de los procesos considerando como punto inicial $x_0 = (0, 0)$. En el resto de esta sección y para los gráficos que se presentan, se asocia el color rojo al proceso sin control, mientras que al proceso óptimo se le asocia el color azul.



(a) No controlado (b) Controlado

Figura 2.1: Trayectorias de los procesos

En la Figura 2.2 se aprecia cómo el proceso óptimo permanece próximo al origen de coordenadas, mientras que el proceso no controlado presenta una mayor dispersión.

Para dar una mejor perspectiva a este resultado, se presenta el siguiente procedimiento para comparar el proceso óptimo con el proceso no controlado: Se considera como punto inicial a $x_0 = (0, 0)$ y se renombra al proceso óptimo por $\{x_t^*\}$, posteriormente se efectua lo siguiente:

1. Se realiza una simulación del proceso $\{\xi_t\}$ hasta $t = N - 1$.
2. Se evalúan los estados del proceso no controlado, es decir, se obtiene x_t, para $t = 0, 1, \ldots, N - 1$.
3. Se evalúan los estados del proceso óptimo, es decir, se obtiene x_t^*, para $t = 0, 1, \ldots, N - 1$.
4. Se observan x_{N-1} y x_{N-1}^*, y se calcula su distancia respecto al origen.

El procedimiento anterior se repite en k ocasiones, para algún $k \in \mathbb{N}$, de manera que se obtiene una muestras de tamano k de x_{N-1} y otra del mismo tamaño de x_{N-1}^*. Con lo anterior, se calcula la distancia promedio al origen de los k puntos correspondientes a la muestra de x_{N-1}, y se hace lo mismo

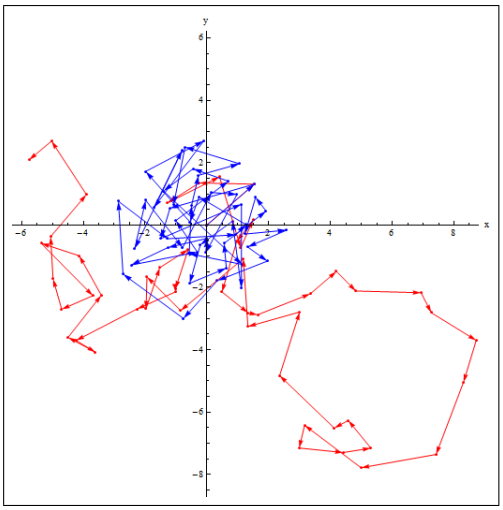


Figura 2.2: Comparación

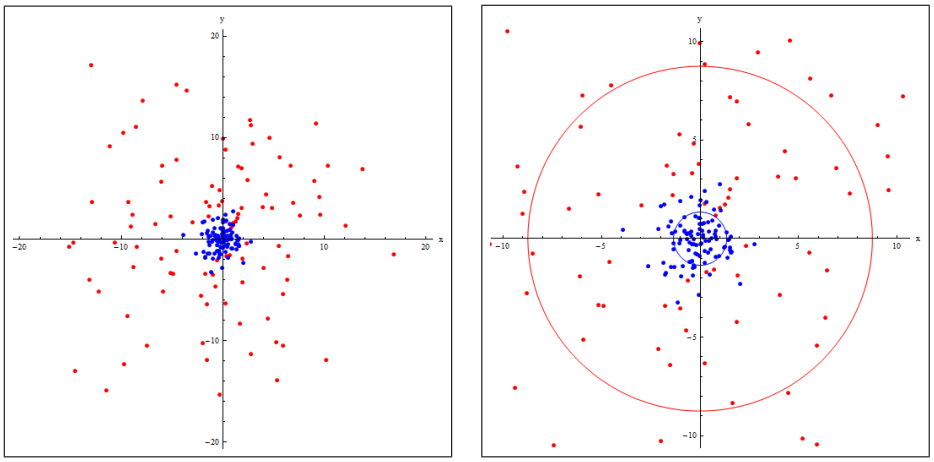
con la muestra de x^*_{N-1}.

En el Cuadro 2.1 con los resultados obtenidos a partir de la programación del algoritmo anterior en el software Mathematica 9, considerando el horizonte de planeación $N = 51$ y para distintos valores de k.

Tamaño de muestra	Distancia promedio	
k	x_{50}	x^*_{50}
50	9.95204	1.49118
100	8.744	1.34925
500	8.84497	1.43045
1000	8.53397	1.38624

Cuadro 2.1: Resultados

En la Figura 2.3 (a) se han graficado en color rojo los puntos correspondientes a una muestra de tamaño 100 de x_{50} y los puntos de una muestra del mismo tamaño de x^*_{50} en color azul. Por otro lado, en la Figura 2.3 (b) se ha realizado un acercamiento a la Figura 2.3 (b) y se ha añadido una circunferencia en color rojo de radio la distancia promedio respecto al origen



(a) (b)

Figura 2.3: Resultados para k=100

de la muestra de x_{50} y una circunferencia en color azul de radio la distancia promedio respecto al origen de la muestra obtenida de x^*_{50}.

Los resultados presentados sustentan la idea de que el proceso óptimo mantiene a los estados del sistema más cerca del origen de coordenadas comparado con el proceso no controlado.

Capítulo 3

Problemas con Horizonte Infinito

Como se mencionó en la Sección 1.4, en ocasiones es conveniente considerar un PDM con horizonte infinito, por ello, el objetivo de este capítulo es proveer una herramienta para garantizar la existencia de la solución del problema de control óptimo.

En lo subsecuente, se considera el criterio de rendimiento de costo total descontado con horizonte infinito como se presentó en la Sección 1.4, es decir,

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

para cada $\pi \in \Pi$ y $x \in X$, donde $\alpha \in (0, 1)$ es el factor de descuento.

Nuevamente, el problema consiste en determinar una política $\pi^* \in \Pi$ óptima, es decir, π^* satisface

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

para cada $x \in X$, donde V^* es la función de valores óptimos.

3.1. Función de Costos Acotada

Definición 3.1. Se dice que una función $v : X \to \mathbb{R}$ es una solución de la Ecuación Óptima para el Costo Descontado (EOCD) si satisface:

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

para cada $x \in X$.

En el Teorema 3.1 se prueba que la función de valores óptimos V^* es una solución de la EOCD, es decir, satisface que

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

para cada $x \in X$. Para ello, se requiere de las siguientes condiciones de estructura.

Condición 3.1.

(a) Para cada estado $x \in X$, el conjunto $A(x)$ es un subconjunto compacto no vacío de A.

(b) Existe una constante $M > 0$, tal que $|c(x,a)| \leq M$ para cada $(x,a) \in \mathbb{K}$ y, además, para cada $x \in X$, $c(x,a)$ es l.s.c. en $\mathbb{K}$.

(c) La ley de transición Q es fuertemente continua.

Notación 1. Se denota por $B(X)$ al espacio de Banach de funciones reales sobre X medibles y acotadas bajo la norma del supremo ($\|v\| \coloneqq \sup_x |v(x)|$). Si además se considera que las funciones son no negativas, el espacio será denotado por $B(X)^+$. No confundir con la σ-álgebra de Borel $\mathcal{B}(X)$.

A continuación se enuncia el teorema que garantiza que V^* es la única solución de la EOCD y que existe una política óptima.

Teorema 3.1. Bajo la Condición 3.1,

a) La función de valores óptimos V^* es la única solución en $B(X)$ de la ecuación óptima para el costo descontado, es decir, V^* satisface,

$$V^*(x) = \min_{a \in A(x)} \left\{ c(x,a) + \alpha \int_X V^*(y) Q(dy|x,a) \right\}, \tag{3.1}$$

para cada $x \in X$.

La ecuación (3.1) también es conocida como la ecuación de programación dinámica para el costo total descontado.

b) Una política $f^* \in \mathbb{F}$ es óptima si y sólo si $f^*(x)$ minimiza el lado derecho de la ecuación (3.1) para el costo descontado para toda $x \in X$, esto es,

$$V^*(x) = c(x, f^*(x)) + \alpha \int_X V^*(y) Q(dy|x, f^*(x)), \tag{3.2}$$

para cada $x \in X$.

Para probar el Teorema 3.1 se deben probar algunos resultados preliminares.

Sea T el operador sobre $B(X)$ definido por

$$Tv(x) \coloneqq \min_{a \in A(x)} \left\{ c(x,a) + \alpha \int_X v(y) Q(dy|x,a) \right\},$$

para cada $v \in B(X)$ y $x \in X$. El operador T es conocido como el *operador de programación dinámica*. Usando la Proposición C.2 en el Apéndice C, se puede mostrar que $Tv \in B(X)$, para todo $v \in B(x)$. Con lo anterior, note que la EPD puede escribirse como

$$V^* = TV^*.$$

Ahora, para cada política estacionaria $g \in \mathbb{F}$, se define el operador T_g sobre $B(X)$ como

$$T_g v(x) \coloneqq c(x, g(x)) + \alpha \int_X v(y) Q(dy|x, g(x)),$$

donde $v \in B(X)$ y $x \in X$. Así, la ecuación (3.2) en el Teorema 3.1 b) se puede reescribir

$$V^* = T_{f^*} V^*.$$

Lema 3.1. Para cada $g \in \mathbb{F}$, T y T_g son *operadores contracción* (Definición E.1, Apéndice E) con módulo α; entonces, por el Teorema del Punto Fijo de Banach (Proposición E.1, Apéndice E), existe una única función $u^* \in B(X)$

y una única función $v_g \in B(X)$, tal que

$$Tu^* = u^* \qquad \text{y} \qquad T_g v_g = v_g,$$

además, para cada función $v \in B(X)$,

$$\|T^n v - u^*\| \to 0 \quad \text{y} \quad \|T^n v - v_g\| \to 0 \quad \text{siempre que} \quad n \to \infty.$$

Demostración. Probemos que T es un operador monótono. Sean $u, v \in B(X)$, tales que, $u \leq v$, entonces

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

para toda $(x,a) \in \mathbb{K}$, luego, dado $\alpha \in [0,1)$, se tiene

$$\min_{a\in A(x)} \left\{ c(x,a) + \alpha \int_X u(y)Q(dy|x,a) \right\} \leq \min_{a\in A(x)} \left\{ c(x,a) + \alpha \int_X v(y)Q(dy|x,a) \right\}$$

para cada $x \in X$, es decir, $Tu \leq Tv$.

Por otro lado, para cualquier constante k y para cada $x \in X$, se cumple que

$$\begin{aligned} T(v(x)+k) &= \min_{a\in A(x)} \left\{ c(x,a) + \alpha \int_X (v(y)+k)Q(dy|x,a) \right\} \\ &= \min_{a\in A(x)} \left\{ c(x,a) + \alpha \int_X v(y)Q(dy|x,a) + \alpha k \int_X Q(dy|x,a) \right\} \\ &= Tv(x) + \alpha k, \end{aligned}$$

es decir, $T(v+k) = Tv + \alpha k$, para cada $v \in B(X)$ y cualquier constante k.

Entonces, por la Proposición E.2 (Apéndice E), T es un operador contracción con módulo α. Por lo tanto, por el Teorema del Punto Fijo de Banach, existe un único punto fijo $u^* \in B(X)$.

Ahora, sea $n \in \mathbb{N}$, una vez más, por el Teorema del punto fijo, se tiene que

$$\|T^n u - u^*\| \leq \alpha^n \|u - u^*\|,$$

por lo tanto, $\|T^n u - u^*\| \to 0$ siempre que $n \to \infty$.

La prueba para T_g es evidente a partir de lo anterior. ■

A continuación se pretende relacionar los "puntos fijos" u^* y v_g del Lema 3.1 con la función de valor óptimo V^* y con la función $V(g,x)$ cuando se usa la política estacionaria g. Para comenzar con ello, en el siguiente Lema se prueba que $v_g = V_g$, con $V_g(x) = V(g,x)$.

Lema 3.2. Bajo las definiciones y suposiciones de esta sección, se verifica que

a) $v_g = V_g$, para cada política estacionaria $g \in \mathbb{F}$.

b) Una política π^* es óptima si y sólo si la función de valor V_{π^*} satisface la EPD, es decir, $TV(\pi^*, x) = V(\pi^*, x)$, para cada $x \in X$.

Demostración. a) Por la unicidad el punto fijo v_g del operador T_g (Lema 3.1), es suficiente verificar que V_g satisface la relación $V_g = T_g V_g$. Para ello, se reescribe $V_g(x)$ como sigue:

$$V_g(x) = E_x^g\left[\sum_{t=0}^{\infty}\alpha^t c(x_t,a_t)\right] = c(x,g(x)) + \alpha E_x^g\left[\sum_{t=1}^{\infty}\alpha^{t-1}c(x_t,a_t)\right].$$

Luego, por propiedades de la esperanza condicional (Proposición D.1, Apéndice D) y la propiedad de Markov, se satisface que

$$\begin{aligned} E_x^g\left[\sum_{t=1}^{\infty}\alpha^{t-1}c(x_t,a_t)\right] &= E_x^g\left[E_x^g\left[\sum_{t=1}^{\infty}\alpha^{t-1}c(x_t,a_t)\middle| h_1\right]\right] \\ &= E_x^g\left[E_{x_1}^g\left[\sum_{t=1}^{\infty}\alpha^{t-1}c(x_t,a_t)\right]\right] \\ &= E_x^g[V_g(x_1)] \\ &= \int_X V_g(y)Q(dy|x,g(x)), \end{aligned}$$

por lo tanto, para cada $x \in X$,

$$V_g(x) = c(x,g(x)) + \alpha\int_X V_g(y)Q(dy|x,g(x)) = T_g V_g(x),$$

es decir, $V_g = T_g V_g$.

b) Suponga que π^* es una política tal que la función $u(x) \coloneqq V(\pi^*, x)$ satisface la EPD, $u = Tu$, es decir,

$$u(x) = \min_{a \in A(x)} \left\{ c(x,a) + \alpha \int_X u_y Q(dy|x,a) \right\}, \tag{3.3}$$

para cada $x \in X$. Con lo anterior, se probará que π^* es la política óptima, es decir, $u(x) \leq V(\pi, x)$ para cada política $\pi \in \Pi$ y cada estado inicial $x \in X$. Para simplificar la notación, en esta prueba se fija una política arbitraria $\pi \in \Pi$, un estado cualquiera $x \in X$, y con ello, en lo siguiente se puede definir $E \coloneqq E_x^\pi$.

Luego, para cualquier historia $h_t \in \mathbb{H}_t$, por la propiedad de Markov, se sigue que:

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

o bien,

$$\alpha^t u(x_t) - E[\alpha^{t+1} u(x_{t+1})|h_t, a_t] \leq \alpha^t c(x_t, a_t).$$

Así, tomando la esperanza en ambos lados de la desiguldad se tiene que

$$E[\alpha^t u(x_t)] - E[\alpha^{t+1} u(x_{t+1})] \leq E[\alpha^t c(x_t, a_t)],$$

y sumando sobre $t = 0, \ldots, n$, se obtiene

$$u(x) - \alpha^{n-1} E[u(x_{n-1})] \leq E\left[\sum_{t=0}^{n} \alpha^t c(x_t, a_t) \right].$$

Finalmente, tomando el límite cuando n tiende a infinito, se concluye que $u(x) \leq V(\pi, x)$, es decir, π^* es óptima.

Ahora, suponga que π^* es óptima. De esta manera, se probará que

$u(x) \coloneqq V(\pi^*, x)$ satisface (i) $u \geq Tu$, y (ii) $u \leq Tu$, es decir, que u satisface la EPD. Entonces, para probar (i) se reescribe $u(x)$ como en la prueba de la parte a), es decir, se tiene que

$$\begin{aligned} u(x) &= E_x^{\pi^*}\left[\sum_{t=0}^{\infty} \alpha^t c(x_t, a_t)\right] \\ &= \int_A \left(c(x,a) + \alpha \int_X V\left(\pi^{*(1)}, y\right) Q(dy|x,a)\right) \pi_0^*(da|x), \end{aligned}$$

donde $\pi^{*(1)} = \left\{\pi_t^{*(1)}\right\}$, se define como la *política desplazada un paso*, esto es, dada la política $\pi^* = \{\pi_t^*\}$, con $x_0 = x$ y $a_0 = a$,

$$\pi_t^{*(1)}(\cdot|h_t) \coloneqq \pi_{t+1}^*(\cdot|x_0, a_0, h_t), \quad \text{para cada } t \geq 0.$$

Entonces, bajo la hipótesis de que π^* es óptima,

$$\begin{aligned} u(x) &\geq \int_A \left(c(x,a) + \alpha \int_X u(y) Q(dy|x,a)\right) \pi_0^*(da|x) \\ &\geq \min_{a \in A(x)} \left\{c(x,a) + \alpha \int_X u(y) Q(dy|x,a)\right\} \\ &= Tu(x), \end{aligned}$$

lo cual prueba (i).

Para probar (ii), sea $f \in \mathbb{F}$ una política estacionaria arbitraria, y sea $\pi' \coloneqq (f, \pi^*)$ la política que utiliza a f al tiempo $t = 0$, y utiliza a la política óptima π^* para el tiempo $t = 1$ en adelante, es decir, $\pi_0'(x_0) \coloneqq f(x_0)$, y para cada $t \geq 1$,

$$\pi_t'(\cdot|x_0, a_0, \ldots, x_{t-1}, a_{t-1}, x_t) \coloneqq \pi_{t-1}^*(\cdot|x_1, a_1, \ldots, x_t).$$

Luego, la optimalidad de π^* implica que

$$u(x) \leq V(\pi', x) = r(x, g(x)) + \alpha \int_X Q(dy|x, g(x)),$$

para cada $x \in X$, así, dado que $f \in \mathbb{F}$ es arbitrario,

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

Esto concluye la prueba del Lema 3.2. ■

A continuación se prueba el Teorema 3.1 haciendo uso de los Lemas 3.1 y 3.2.

Demostración del Teorema 3.1

Demostración. a) Por el Lema 3.2 parte b), una política π^* es óptima, es decir, $V(\pi^*, x) = v^*(x)$ para cada $x \in X$, si y sólo si v^* satisface la EPD $v^* = Tv^*$, y la unicidad de dicha función (o punto fijo) v^* se sigue del Lema 3.1.

b) Suponga que $f \in \mathbb{F}$ es una política estacionaria que satisface la ecuación (3.2), es decir, $v^* = T_f v^*$. Entonces, por el Lema 3.1 y la parte a) del Lema 3.2 se tiene que

$$v^* = v_f = V_f,$$

además, f es óptima.
Por otro lado, si $f \in \mathbb{F}$ es óptima, entonces $v^* = v_f$ y la unicidad del punto fijo v_f implica que $v^* = T_f v^*$, es decir, f satisface la ecuación (3.2).
Esto completa la demostración del Teorema 3.1. ■

Con lo descrito hasta el momento, se ha abordado el problema con función de costos acotada, el paso natural sería probar un resultado análogo al Teorema 3.1 para el caso con función de costos no necesariamente acotada. Este resultado se presenta en la siguiente sección.

3.2. Función de Costos no Negativa

Nuevamente, se retoma la formulación del problema de control óptimo definido en la sección anterior, pero ahora considerando que la función de costos c no necesariamente es acotada superiormente, de manera que las condiciones de estructura quedan establecidas de la siguiente manera.

Condición 3.2.

(a) Para cada estado $x \in X$, el conjunto $A(x)$ es un subconjunto compacto no vacío de A.

(b) La función de costos c es no negativa y l.s.c. en $\mathbb{K}$.

(c) La ley de transición Q es fuertemente continua.

Observación 3.1. En (a), basta con que c sea acotada inferiormente en lugar de no negativa, debido a que, si $c(x,a) \geq m$, para toda $(x,a) \in \mathbb{K}$, entonces $c'(x,a) = c(x,a) - m \geq 0$, para toda $(x,a) \in \mathbb{K}$. Otra condición necesaria para el desarrollo posterior es que c sea inf-compacta, sin embargo, es fácil probar que bajo (a) y (b), la función de costos c satisface dicha condición.

Aunque para este nuevo problema se han debilitado algunas condiciones con respecto al caso con función de costos acotada, es necesario añadir una suposición adicional para garantizar que la función de valor óptimo es finita para cada $x \in X$; dicha suposición es la siguiente.

Suposición 3.1. Existe $\pi \in \Pi$ tal que, para cada $x \in X$, $V(\pi, x) < \infty$.

Claramente esta suposición se satisface si c es acotada, ya que, si $0 \leq c \leq M$, entonces para cada $x \in X$ y $\pi \in \Pi$

$$V(\pi,x) = E_x^{\pi}\left[\sum_{t=0}^{\infty} \alpha^t c(x_t, a_t)\right] \leq \sum_{t=0}^{\infty} \alpha^t M = \frac{M}{1-\alpha} < \infty.$$

Por lo anterior, se define $\Pi_0 \subseteq \Pi$ como el conjunto de políticas que satisfacen la Suposición 3.1, es decir, si $\pi \in \Pi_0$, entonces $V(\pi, x) < \infty$.

Ahora se presenta el resultado principal de esta sección.

Teorema 3.2. Si la Condición 3.2 y a Suposición 3.1 se verifican, entonces:

(a) La función de valor V^* es la solución minimal de la EOCD, es decir,

$$V^*(x) = \min_{a \in A(x)} \left\{ c(x,a) + \alpha \int_X V^*(y) Q(dy|x,a) \right\}, \quad \forall x \in X, \tag{3.4}$$

y si u es otra solución de la EOCD, entonces $u(\cdot) \geq V^*(\cdot)$.

(b) Existe un selector $f_* \in \mathbb{F}$, tal que

$$V^*(x) = c(x, f_*) + \alpha \int_X v(y) Q(dy|x, f_*), \quad \forall x \in X, \tag{3.5}$$

y la política determista estacionaria $f_*^\infty = \{f_*, f_*, \ldots\}$ es óptima; y recíprocamente, si $f_*^\infty \in \Pi_{DS}$ es óptima, entonces satisface la ecuación (3.5).

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

$$\lim_{n\to\infty} \alpha^n E_x^\pi \left[V(\pi^*, x_n)\right] = 0, \tag{3.6}$$

para cada $\pi \in \Pi_0$ y $x \in X$, entonces $V(\pi^*, \cdot) = V^*(\cdot)$; por lo tanto, π^* es la política óptima. En otras palabras, si (3.6) se satisface, entonces π^* es la politica óptima si y sólo si $V(\pi^*, \cdot)$ satisface la EOCD.

(d) Si existe una política óptima, entonces existe una política óptima determinista estacionaria.

La prueba del Teorema 3.2 requiere de varios resultados previos. El primero de ellos es sobre el intercambio entre límite y mínimo, el cual se presenta a continuación.

Lema 3.3. Para cada $n = 1, 2, \ldots$, sean u y u_n funciones l.s.c., acotadas inferiormente e ínf-compactas sobre $\mathbb{K}$. Si $u_n \nearrow u$, entonces para cada $x \in X$

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

Demostración. Se definen, para cada $x \in X$, las siguientes funciones

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

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

Luego, dado que $u_n \nearrow u$, se tiene que $l(\cdot) \leq u^*(\cdot)$.

Para demostrar la desigualdad inversa, es decir, que $l(\cdot) \geq u^*(\cdot)$, sea $x \in X$ arbitrario pero fijo, y defínase, para cada $n = 1, 2, \ldots$

$$\begin{aligned} A_n &\coloneqq \{a \in A(x) : u_n(x, a) \leq u^*(x)\}, \\ A_0 &\coloneqq \{a \in A(x) : u(x, a) = u^*(x)\}. \end{aligned}$$

Note que, para cada n, A_n es un conjunto compacto debido a la ínf-

compacidad de u_n. Además, $An \searrow A_0$, es decir,

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

En efecto, sea $a \in \bigcap_{n=1}^{\infty} A_n$, entonces para toda n,

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

así,

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

y, por la definición de $u^*(x)$, se tiene que

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

de esta manera $u(x, a) = u^*(x)$, es decir, $a \in A_0$.

Por lo tanto, dado que A_0 es igual a la intersección de una familia numerable de conjuntos compactos, se concluye que A_0 es un conjunto compacto.

Por otro lado, por el teorema de selección medible (ver Apéndice C), se tiene que, para cada $n \geq 1$, existe $a_n \in A_n$, tal que

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

Entonces, por la compacidad de A_0, existe una subsucesión $\{a_{n_k}\}$ de la sucesión $\{a_n\}$, tal que $a_{n_k} \to a_0$, para algún $a_0 \in A_0$.

Así, para toda $n_k \geq n$, se tiene que

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

Cuando $n_k \to \infty$, de la desigualdad anterior se obtiene

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

Así, si $n \to \infty$,

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

y dado que $x \in X$ fue elegido de manera aribitraria, se concluye que, para toda $x \in X$,

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

■

Definición 3.2. $M(X)^+$ denota el conjunto de funciones medibles no negativas sobre X, y, para cada $u \in M(X)^+$, Tu es la función definida sobre X como

$$Tu(x) \coloneqq \min_{a \in A(x)} \left\{ c(x, a) + \alpha \int_X u(y) Q(dy|x, a) \right\}, \tag{3.7}$$

para cada $x \in X$.

Entonces, por el Teorema 2.1 se obtiene el siguiente resultado.

Lema 3.4. Bajo la Suposición 3.2, T es un operador de $M(X)^+$ sobre sí mismo, es decir, para cada $u \in M(X)^+$, Tu también está en $M(X)^+$, más aún, existe un selector $f \in \mathbb{F}$ tal que, para cada $x \in X$

$$Tu(x) = c(x, f) + \alpha \int_X u(y) Q(dy|x, f). \tag{3.8}$$

Note que el operador $T : M(X)^+ \to M(X)^+$ es el operador de programación dinámica que se definió en la sección anterior. Más aún, el operador restringido sobre el espacio de las funciones acotadas no negativas $\left(T \left|_{B(X)^+}\right.\right)$ es una contracción bajo la norma del supremo.

Lema 3.5. Bajo la Condición 3.2 y la Suposición 3.1, se tiene

(a) Si $u \in M(X)^+$ es tal que $u \geq Tu$, entonces $u \geq V^*$.

(b) Si $u : X \to \mathbb{R}$ es una función medible tal que Tu está bien definida, además, $u \leq Tu$ y

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

para cada $\pi \in \Pi_0$ y $x \in X$, entonces $u \leq V^*$.

Demostración. (a) Sea $u \in M(X)^+$ tal que $u \geq Tu$, por el Lema 3.5 se tiene que para toda $x \in X$

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

iterando esta relación se obtiene lo siguiente

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

continuando de una manera análoga con las iteraciones, se satisface que

$$u(x) \geq E_x^f \left[\sum_{t=0}^{n-1} \alpha^t c(x_t, a_t) \right] + \alpha^n E_x^f [u(x_n)], \tag{3.9}$$

donde

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

Dado que u es no negativa, para toda $n \geq 1$ y $x \in X$, se tiene que

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

Tomando el límite cuando n tiende a infinito, se verifica que

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

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

$$u \geq V^*,$$

con lo cual se concluye la prueba de (a).

(b) Considere $\pi \in \Pi_0$ y $x \in X$ arbitrarios. Por la propiedad de Markov y bajo la suposición $Tu \geq u$,

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

por lo tanto,

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

Luego, tomando esperanzas y sumando desde $t = 0, \ldots, n-1$, para $n \in \mathbb{N}$ arbitrario, se obtiene

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

Finalmente, tomando $n \to \infty$, se tiene que

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

y dado que π y x fueron tomados de manera arbitraria, se concluye que $V^* \geq u$. ■

Para lo posterior, se consideran las siguientes definiciones y notaciones.

Definición 3.3. Sea $n \in \mathbb{N}$ arbitrario. Se define la función uniformemente acotada $c^n : \mathbb{K} \to \mathbb{R}$, como

$$c^n(x, a) \coloneqq \text{mín}\{c(x, a), n\} \tag{3.10}$$

para cada $(x, a) \in \mathbb{K}$. Por otro lado, para cada $\pi \in \Pi_0$ y $x \in X$ se define

$$V^n(\pi, x) = E_x^\pi \left[\sum_{t=0}^{\infty} \alpha^t c^n(x_t, a_t)\right]. \tag{3.11}$$

Observación 3.2. Es fácil ver que bajo la Suposición 3.2, c^n es uniformemente acotado y l.s.c., para cada $n \in \mathbb{N}$, es decir, se satisface la Suposición 3.1. Más aún, c^n es ínf-compacta sobre $\mathbb{K}$. En efecto, sea $\lambda \in \mathbb{R}$, así, defínase el conjunto $A_{\lambda,n} \coloneqq \{a \in A(x) : c^n(x,a) \leq \lambda\}$. Sea una sucesión $\{a_k\} \subset A_{\lambda,n}$, entonces, dado que $A(x)$ es compacto, existe una sub-sucesión $\{a_{k_i}\} \subset \{a_k\}$ que converge a algún $\hat{a} \in A(x)$; luego, dado que $c^n(x,a)$ es l.s.c. sobre $\mathbb{K}$, entonces para cada $x \in X$, $\liminf_{i\to\infty} c^n(x, a_{k_i}) \geq c^n(x, \hat{a})$, por lo que $c(x,\hat{a}) \leq \lambda$, es decir, $\hat{a} \in A_{\lambda,n}$, por lo tanto, $A_{\lambda,n}$ es compacto, o bien, c^n es ínf-compacto sobre $\mathbb{K}$.

El siguiente lema relaciona el problema de control con criterio definido por (3.11) y el problema original de esta sección.

Lema 3.6. Bajo las Suposiciones 3.2 y 3.1, dado $n \in \mathbb{N}$ arbitrario, si V^{n*} es la solución de la EOCD con criterio de rendimiento dado por (3.11), es decir, si

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

entonces:

(a) $V^{n*} \nearrow V^*$, siempre que $n \to \infty$.

(b) V^* es solución de la EOCD.

Demostración. Sea $n \in \mathbb{N}$ cualquiera pero fijo. Por la Observación 3.2, se satisfacen las condiciones del Teorema 3.1 y entonces $V^{n*} \in B(X)^+ \subset M(X)^+$ es tal que, $TV^{n*} = V^{n*}$.

Por otro lado, observe que por la definición de c^n, para todo $(x,a) \in \mathbb{K}$ se satisface

$$c^n(x,a) \leq c^{n+1}(x,a) \leq c(x,a),$$

entonces, para $\pi \in \Pi_0$ arbitrario y para cada $x \in X$

$$V^n(\pi,x) \leq V^{n+1}(\pi,x) \leq V(\pi,x) < \infty,$$

ahora, por la definición de V^{n*} y V^*, y dado que n es arbitrario, se verifica, para cada $n \in \mathbb{N}$ y $x \in X$, que

$$V^{n*}(x) \leq V^{(n+1)*}(x) \leq V^*(x),$$

es decir, $\{V^{n*}\}$ es una sucesión creciente y acotada en $M(X)^+$, así, existe $\widehat{V} \in M(X)^+$ tal que, $V^{n*} \nearrow \widehat{V}$ siempre que $n \to \infty$, además, $\widehat{V} \leq V^*$.

Resta probar que $T\widehat{V} = \widehat{V}$. Sea $k \in \mathbb{N}$ y considere las funciones $u_k, u : \mathbb{K} \to \mathbb{R}$ definidas por

$$u_k(x,a) \coloneqq c^k(x,a) + \alpha \int_X V^{k*}(y)Q(dy|x,a), \tag{3.12}$$

$$u(x,a) \coloneqq c(x,a) + \alpha \int_X \widehat{V}(y)Q(dy|x,a). \tag{3.13}$$

Por el teorema de convergencia monótona (Teorema A.1, Apéndice A), $u_k \nearrow u$. Además, para cada $k \in \mathbb{N}$, la función u_k es l.s.c., acotada inferiormente e ínf-compacta sobre K.

En efecto, c^k es l.s.c., acotada inferiormente e ínf-compacta como se probó en la Observación 3.2. Por otro lado, la función $v : \mathbb{K} \to \mathbb{R}$ definida por

$$v(x,a) \coloneqq \int_X V^{k*}(y)Q(dy|x,a),$$

para cada $(x,a) \in \mathbb{K}$, es no negativa y continua, ya que Q es fuertemente continua; luego, como $A(x)$ es compacto para cada $x \in X$, es fácil probar que v es ínf-compacta sobre $\mathbb{K}$ (como se hizo para c^n en la Observación 3.2). Por lo tanto, para cada $k \in \mathbb{N}$, u_k es l.s.c., acotada inferiormente e ínf-compacta.

De manera análoga se prueba que u satiface las mismas condiciones.

Así, por el Lema 3.3 se tiene que

$$\widehat{V} = \lim_{n\to\infty} V^{n*} = \lim_{n\to\infty} TV^{n*} = T\widehat{V},$$

o bien, $\widehat{V} = T\widehat{V}$, es decir, $\widehat{V}$ es solución de la EOCD.

Por el Lema 3.5, como $\widehat{V} = T\widehat{V}$, entonces $\widehat{V} \geq V^*$.

Por lo tanto,

$$\widehat{V} = V^*.$$

■

Demostración del Teorema 3.2. (a) Por el Lemma 3.6, V^* es una solución de la EOCD, y el hecho de que V^* sea la solución mínima se debe al Lema 3.5, ya que, si $u \in M(X)^+$ es tal que, $u = Tu$, entonces $u \geq V^*$.

(b) La existencia del selector $f_* \in \mathbb{F}$ se garantiza por el Lema 3.4, y es tal que, para cada $x \in X$

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

Ahora, iterando la relación anterior, como en (3.9), se muestra que

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

haciendo a n tender a infinito, se tiene que

$$V^*(x) \geq V(f_*^\infty, x),$$

para cada $x \in X$. Por otro lado, dado que V^* es la función de valor óptimo, se tiene $V^*(\cdot) \leq V(f_*^\infty, \cdot)$. Por lo tanto, $V^*(\cdot) = V(f_*^\infty, \cdot)$, y con ello, f_*^∞ es la política óptima.

Recíprocamente, sea una política determinista estacionaria $f^\infty \in \Pi_{DS}$, se verifican las siguientes relaciones

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

luego note que, por la Proposición D.1(c) del Apéndice D y la propiedad de Markov, el último elemento de la igualdad puede expresarse como

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

es decir,

$$V(f^{\infty}, x) = c(x, f) + \alpha \int_X V(f^{\infty}, y) Q(dy|x, f).$$

En particular, si $f_* \in \Pi_{DS}$ es óptima, entonces $V(f_*^{\infty}, \cdot) = V^*(\cdot)$.

(c) Si $\pi \in \Pi_0$ es una política tal que, $V(\pi^*, \cdot)$ es solución de la EOCD, por el Lema 3.5 se tiene que $V(\pi^*, \cdot) = V^*(\cdot)$.

Finalmente, (d) es una consecuencia de (a) y (b). ■

El Teorema 3.2 garantiza la existencia de la solución, mas no presenta un algoritmo explícito para la solución del problema de control óptimo vía programación dinámica. Para ello se utilizan las funciones de iteración de valores (IV) que se presentaron al final de la Sección 2.3. En [12] se prueba que v_n es la función de valor óptimo del n-ésimo costo descontado V_n, es decir,

$$v_n(x) = \inf_{\pi \in \Pi} V_n(\pi, x) = \inf_{\pi \in \Pi} E_x^{\pi} \left[\sum_{t=0}^{n-1} c(x_t, a_t) \right],$$

para cada $x \in X$. Además, se prueba que para cada $x \in X$,

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

De hecho, este método es una forma clásica de probar el Teorema 3.2, distinta a la que se ha propueto en esta tesis.

A continuación se presenta un ejemplo de un PDM que modela un problema de consumo e inversión con horizonte infinito. Para este problema se verifica la Condición 3.2 y la Suposición 3.1, por lo que es válido el uso del Teorema 3.1 y con ello garantizar teóricamente la existencia de la política óptima estacionaria. Sin embargo, como se muestra en el ejemplo numérico posterior, la obtención explícita de dicha política unicamente puede llevarse a cabo de manera numérica.

3.3. Ejemplo: Consumo e Inversión

En este ejemplo se considera un proceso $\{x_t\}$, donde, x_t representa el capital de un inversor al tiempo t, con $t = 0, 1, \ldots$. En este caso, la variable de control a_t es la cantidad que el inversor decide invertir al tiempo t, además se supone que el rendimiento de la inversión es aleatorio pero que, en promedio, la cantidad que se invierte al tiempo t genera un rendimiento igual a la cantidad invertida; bajo este contexto, es claro que $0 \leq a_t \leq x_t$, pues no se puede invertir una cantidad mayor a la que se tiene como capital. Una vez que se ha decidido cuanto invertir, el resto de capital, $x_t - a_t$, se utiliza para consumo del inversor, y se supone que el beneficio que percibe el inversor está dado por una función de utilidad exponencial $u : \mathbb{R} \to \mathbb{R}$, definida de la siguiente manera

$$u(x) = -\gamma e^{-\gamma x},$$

donde $\gamma > 0$ es conocida como el coeficiente de aversión absoluta al riesgo. En este caso, el objetivo del inversor es maximizar la utilidad, bajo el supuesto de que este proceso se puede realizar de manera indefinida.

Una vez formulado el problema, habrá que ponerlo en el contexto de los procesos de decisión de Markov. Así, la dinámica del sistema está dada por la siguiente ecuación en diferencias

$$x_{t+1} = \xi_t a_t, \qquad t = 0, 1, 2, \ldots,$$

donde, $\{\xi_t\}$ es una sucesión de variables aleatorias independientes e identicamente distribuidas tales que, $E[\xi_t] = 1$ y $Var[\xi_t] < \infty$, para cada $t = 0, 1, \ldots$.

En este caso, no se tiene una función de costos, sino una función de utilidad que es dada por

$$U(x_t, a_t) = -\gamma e^{-\gamma(x_t - a_t)}, \qquad t = 0, 1, \ldots.$$

Asi, el espacio de estados y el de acciones es el conjunto de los números reales positivos, es decir, $X = \mathbb{R}^+ = A$ y para cada $x \in X$, $A(x) = [0, x]$.

Lo que se desea es maximizar la utilidad esperada descontada, es decir,

se busca $\pi \in \Pi$ que maximice

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

para cada $x \in X$. Observe que el desarrollo teórico que se ha hecho en esta tesis ha considerado el caso cuando se desea minimizar el criterio de rendimiento, sin embargo, se puede transformar el problema de maximización por un problema de minimización realizando ligeras modificaciones, al menos para este ejemplo.

Recuerde que si U es un conjunto distinto del vacío y $u : U \to \mathbb{R}$ es una función que alcanza su máximo en U, entonces la siguiente relación es verdadera

$$\max_{x \in U}[u(x)] = -\min_{x \in U}[-u(x)],$$

más aún, si $x^* \in U$ es el punto donde u alcanza su máximo, entonces $-u$ alcanza su mínimo en x^*.

De esta manera, hallar $\pi \in \Pi$ que maximice $\bar{V}(\cdot, x)$ para cada $x \in X$ es equivalente a hallar $\pi \in \Pi$ que minimice a

$$\begin{aligned} -\bar{V}(\pi, x) &= -E_x^{\pi}\left[\sum_{t=0}^{\infty} \alpha^t U(x_t, a_t)\right] \\ &= E_x^{\pi}\left[\sum_{t=0}^{\infty} \alpha^t (-U(x_t, a_t))\right], \end{aligned}$$

para cada $x \in X$. Así, el nuevo criterio de rendimiento está dado por

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

donde $c(x, a) = -U(x, a)$, para cada $(x, a) \in \mathbb{K}$.

Por lo tanto, para resolver el problema de control óptimo es necesario verificar que se satisface la Condición 3.2 y la Suposición 3.1 a partir de la nueva función c.

En efecto, observe que para cada $x \in X$, $A(x) = [0, x]$ es un subconjunto compacto no vacío de A. Por otro lado, note que $c(x, a) = \gamma e^{-\gamma(x-a)}$ es continua y no negativa para todo $(x, a) \in \mathbb{K}$. Resta probar que la ley de transición Q es fuertemente continua. Para ello, procediendo de manera similar que en el problema LQ de la Sección 2.4, si Δ es la densidad común para cada ξ_t, se verifica para cada $B \in \mathcal{B}(X)$

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

con un cambio de varible se tiene

$$Q(B|x, a) = \int_S I_C[u]\Delta(u/a)du,$$

para cada $x \in X$ y $a \in (0, \infty)$, luego, dado que Δ es continua, se garantiza que Q es fuertemente continua.

Para verificar la Suposición 3.1, sea

$$f(x) = 0$$

para cada $x \in X$. Luego, si se considera la política determinista estacionaria $f^\infty = \{f, f, \ldots\}$, entonces

$$\begin{aligned} V(f^\infty, x) &= E_x^{f^\infty}\left[\sum_{t=0}^{\infty} \alpha^t \gamma e^{-\gamma(0_{t-1}\xi_{t-1}-0_t)}\right] \\ &= E_x^{f^\infty}\left[\sum_{t=0}^{\infty} \alpha^t \gamma\right] \\ &= \gamma \sum_{t=0}^{\infty} \alpha^t \\ &= \frac{\gamma}{1-\alpha} < \infty. \end{aligned}$$

Es decir, existe una política $\pi = f^\infty \in \Pi$, tal que $V(\pi, x) < \infty$, para cada $x \in X$.

Ahora se puede aplicar el algoritmo de programación dinámica; para ello inicialmente se deben hallar las funciones de iteración de valores:

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

para cada $x \in X$.

Aunque el ejemplo no presenta mayor complejidad en su formulación y en la verificación de hipótesis, la obtención de las funciones de iteración de valores no resulta tan sencilla. A continuación se presenta un ejemplo numérico para el cual se presentan las primeras tres iteraciones donde se aplica el algoritmo de programación dinámica.

3.3.1. Ejemplo Numérico

Para este ejemplo se hacen las siguientes consideraciones:

1. Para cada $t \geq 0$,
$$\xi_t \sim Log\text{-N}\left(-\frac{1}{2}, 1\right),$$
es decir, ξ_t sigue una distribución Log-Normal con $E[\xi_t] = 1$ y $Var[\xi_t] = e - 1$.

2. El coeficiente de aversión absoluta al riesgo es $\gamma = \frac{1}{2}$.

3. El factor de descuento α se considera igual a $\frac{1}{2}$.

Observación 3.3. La suposición 1 está basada en [15] donde se justifica el uso de esta distribución cuando se modelan procesos de inversión.

Por las suposiciones anteriores, la función de costos está dada por $c(x,a) = \frac{1}{2}e^{-\frac{1}{2}(x-a)}$, para cada $(x,a) \in \mathbb{K}$ y el criterio de rendimiento está dado por

$$V(\pi, x) = E_x^{\pi}\left[\sum_{t=0}^{2} \left(\frac{1}{2}\right)^t \frac{1}{2} e^{-\frac{1}{2}(x_t - a_t)}\right],$$

para cada $\pi \in \Pi$ y $x \in X$.

La tarea ahora es hallar las funciones IV. Así, dado que $v_0(x) = 0$, para $n = 1$ se tiene:

$$v_1(x) = \min_{a\in[0,x]} \left\{ \frac{1}{2} e^{-\frac{1}{2}(x-a)} \right\} = \frac{1}{2} e^{-\frac{1}{2}x},$$

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

$$\begin{aligned} v_2(x) &= \min_{a\in[0,x]} \left\{ \frac{1}{2} e^{-\frac{1}{2}(x-a)} + \frac{1}{2} E[v_1(a\xi)] \right\} \\ &= \min_{a\in[0,x]} \left\{ \frac{1}{2} e^{-\frac{1}{2}(x-a)} + \frac{1}{2} E\left[\frac{1}{2} e^{-\frac{1}{2}a\xi} \right] \right\}, \end{aligned}$$

haciendo una expansión en serie de Taylor de orden cuatro para $e^{-\frac{1}{2}a\xi}$, se obtiene

$$v_2(x) \approx \min_{a\in[0,x]} \left\{ \frac{1}{2} e^{-\frac{1}{2}(x-a)} + \frac{1}{4} E\left[1 - \frac{1}{2}a\xi + \frac{1}{8}a^2\xi^2 - \frac{1}{24}a^3\xi^3 + \frac{1}{384}a^4\xi^4 \right] \right\},$$

utilizando el hecho de que $E[\xi] = 1$, $E[\xi^2] = e$, $E[\xi^3] = e^3$ y $E[\xi^4] = e^6$, se verifica

$$v_2(x) \approx \min_{a\in[0,x]} \left\{ \frac{1}{2} e^{-\frac{1}{2}(x-a)} + \frac{1}{4} - \frac{1}{8}a + \frac{e}{32}a^2 - \frac{e^3}{96}a^3 + \frac{e^6}{1536}a^4 \right\}.$$

Para hallar el punto que minimiza la función entre corchetes se debe utilizar métodos numéricos implementados en el software Mathematica, dando como resultado que

$$f_2(x) = \begin{cases} 0, & si \quad 0 \le x \le 1.3863, \\ g(x), & si \quad x > 1.3863, \end{cases}$$

con $g(x) = 0.4716 - 2.01e^{-0.496-0.722x}$. De manera que

$$v_2(x) = \begin{cases} \frac{1}{2} e^{-\frac{1}{2}x} + \frac{1}{4}, & si \quad 0 \le x \le 1.3863, \\ h(x), & si \quad x > 1.3863, \end{cases}$$

con $h(x) = \frac{1}{2} e^{-\frac{1}{2}(x-g(x))} + \frac{1}{4} - \frac{1}{8}g(x) + \frac{e}{32}(g(x))^2 - \frac{e^3}{96}(g(x))^3 + \frac{e^6}{1536}(g(x))^4$.

Note que para $n = 3$, resulta complejo poder hallar

$$v_3(x) = \min_{a\in[0,x]} \left\{ \frac{1}{2} e^{-\frac{1}{2}(x-a)} + \frac{1}{2} E[v_2(a\xi)] \right\},$$

debido a la forma de $v_2(\cdot)$.

Con este ejemplo se prueba que, a pesar de la formulación tan simple con la que se presenta el problema, en ocasiones no es sencillo hallar una solución analítica simple como ocurrió en el problema LQ. Más aún, existe toda una teoría para la obtención de dichas políticas a partir de métodos numéricos. Parte de esta teoría puede consultarse en [6].

Conclusiones

El trabajo de tesis se enfocó en el estudio de procesos de decisión de Markov para los casos total y descontado. En el caso total se estudió el criterio de rendimiento considerando un horizonte finito y para el caso descontado se presentan ambos horizontes, finito e infinito. La demostración de validación de programación dinámica se basó en la construcción de un proceso martingala, se observa que es posible aprovechar otras caracterizaciones y propiedades de martingalas para el estudio de problemas de optimización, sin embargo, dichos enfoques quedan fuera del trabajo de investigación actual y se pueden retomar en trabajos futuros. Por otro lado, en el caso de horizonte infinito en el contexto de costos descontados se presenta una variante de la demostración clásica considerando problemas con costos acotados, definiendo el truncamiento de los costos y de este modo usarlos para determinar una versión de la ecuación de programación dinámica, una vez que se determinó para el caso de costos acotados. En cada capítulo se presentan ejemplos que ilustran la teoría desarrollada, uno muy conocido en la literatura, denominado LQ y otra propuesta en modelos de crecimiento económico, en específico, un problema referente a consumo e inversión. Consideramos que el problema LQ es de importancia conocerlo y presentarlo en un primer estudio de PDM debido a su gran uso en diversas áreas aplicadas y a que ha sido implementado en métodos de aproximación numérica a modelos económicos. El segundo ejemplo es una propuesta del autor, el cual ilustra que a pesar de garantizar la existencia de estrategias óptimas, no es posible determinar una solución cerrada del mismo; de este modo se muestra la importancia de implementar métodos numéricos para aproximar soluciones en el contexto de PDM.

Los problemas que se consideran como consecuencia del trabajo son los

siguientes:

1. Crear modelos usando PDM de problemas y aplicaciones con datos reales. Por ejemplo, ya que se tiene resuelto el problema LQ, partir de él para considerar el problema con horizonte aleatorio independiente del proceso y reunir datos reales de algún proceso sin control. El trabajo en este caso consistiría en la estimación de los parámetros para la modelación del proceso.

2. Desarrollar la teoría para solucionar los problemas con criterio de costo total acumulado y horizonte aleatorio dependiente del proceso. Para ello, sería conveniente implementar la teoría de tiempos de paro, similar a lo que se hace en [7] cuando se considera el criterio de costo descontado.

3. Como se observó en el ejemplo de consumo e inversión, en problemas con función de utilidad de tipo exponencial y dinámica multiplicativa, en general, no es posible determinar directamente una solución explícita. Sin embargo, en el trabajo de tesis se ilustra que es posible garantizar la existencia de estrategias óptimas y la validez de programación dinámica. Este tipo de ejemplos existen en numerosas áreas por lo que es necesario la implementación de métodos numéricos para la aproximación de la solución. Algunos de estos métodos son: aproximaciones LQ, métodos proyectivos, algoritmos genéticos, entre otros. Una referencia en el contexto de modelos económicos donde se pueden consultar dichos métodos es [6]. Este tipo de métodos abren una línea de investigación para trabajos futuros dentro del área de PDM.

Apéndice A

Miscelánea de Resultados

Definición A.1. Sea (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 $\{x_n\} \subset X$ convergente a x.

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

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

Observe que v es continua, si y sólo si, es l.s.c. y u.s.c.

Sea $L(X)$ la familia de todas las funciones sobre X que son l.s.c. y acotadas inferiormente.

Proposición A.1. Si $v, v_1, \ldots, v_n$ pertenecen a $L(X)$, entonces

(a) Las funciones αv, con $\alpha \geq 0$, $v_1 + \cdots + v_n$ y $\min_i v_i$, pertenecen a $L(X)$.

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

Demostración. La demostración puede hallarse en [2] y [5]. ■

Teorema A.1 (Teorema de Convergencia Monótona)**.** Sea $\{f_n\}$ una sucesión monótona de funciones medibles no negativas definidas sobre X, supon-

ga que convergen a una función medible f, entonces

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

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

Apéndice B

Kérneles Estocásticos

Definición B.1. Sea (X, τ) un espacio topológico. La σ-álgebra generada por τ es la σ-álgebra de Borel y será denotada por $\mathcal{B}(X)$.

Definición B.2. X es un espacio de Borel, si X es un conjunto de Borel de un espacio métrico separable y completo.

Por ejemplo, $\mathbb{R}^n$ con la topología usual es un espacio de Borel.

Definición B.3. Sean X e Y espacios de Borel. Un *kérnel estocástico* definido sobre X dado Y es una función $P(\cdot|\cdot)$ tal que:

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

(b) $P(B|\cdot)$ es una variable aleatoria (función medible) en Y para cada $B \in \mathcal{B}(X)$

La familia de todos los kérneles estocásticos será denotada por $P(X|Y)$.

Definición B.4. El kérnel estocástico $P \in P(X|A)$ es

a) **Débilmente continuo**, (o que satisface la propiedad de Feller) si la función
$$y \mapsto \int_X v(x)P(dx|y) \in C(A)$$
para cualquier $v \in C(X)$.

b) **Fuertemente continuo**, (o que satisface la propiedad fuerte de Feller) si la función
$$y \mapsto \int_X v(x)P(dx|y) \in C(A)$$

para cualquier $v \in B(X)$.

Donde $B(X)$ denota al espacio de Banach de funciones reales sobre X medibles y acotadas bajo la norma del supremo; mientras que $C(X)$ denota al espacio de Banach de funciones reales sobre X continuas y acotadas bajo la norma del supremo. Es claro que fuertemente continuo implica débilmente continuo.

Proposición B.1. Las siguientes proposiciones son equivalentes.

(a) P es fuertemente continua.

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

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

Proposición B.2. Las siguientes proposiciones son equivalentes.

(a) P es debilmente continua.

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

Teorema B.1 (Teorema de Ionescu-Tulcea)**.** Sea $X_0, X_1, \ldots,$ una sucesión de espacios de Borel y, para $n = 0, 1, \ldots,$ se define $Y_n := X_0 \times \cdots \times X_n$ e $Y := \prod_{n=0}^{\infty} X_n$. Sea ν una medida de probabildad arbitraria sobre X_0 y, para cada $n = 0, 1, \ldots,$ sea $P_n(dx_{n+1}|y_n)$ un kérnel estocástico sobre X_{n+1} dado Y_n. Entonces, existe una única medida de probabilidad P_ν sobre Y tal que, para cada rectángulo medible $B_0 \times \cdots \times B_n \in Y_n$,

$$\begin{aligned} P_\nu(B_0 \times \cdots \times B_n) &= \int_{B_0} \nu(dx_0) \int_{B_1} P_0(d_{x_1}|x_0) \int_{B_2} P_1(d_{x_2}|x_0, x_1) \\ &\quad \cdots \int_{B_n} P_{n-1}(d_{x_n}|x_0, \ldots, x_{n-1}). \end{aligned}$$

Además, para cualquier función medible u sobre Y, la función

$$x \mapsto \int u_y P_x(dy)$$

es medible sobre X_0, donde P_x representa a P_ν cuando ν es la probabilidad concentrada en $x \in X_0$.

Demostración. Ver [2] y [5]. ■

Apéndice C

Multifunciones y Selectores

Sean X y A espacios de Borel.

Una multifunción $\psi : X \to A$ es una función tal que, $\psi(x)$ es un subconjunto no vacío de A, $x \in X$. La gráfica de la multifunción ψ es el subconjunto de $X \times A$ definido como

$$graf(\psi) \coloneqq \{(x, a) : x \in X, a \in \psi(x)\}. \tag{C.1}$$

Definición C.1. Una multifunción $\psi : X \to A$ se dice que es:

(a) *Borel Medible*, si $\psi^{-1}(B) \in \mathcal{B}(X)$ para cada conjunto abierto $B \subset A$.

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

(c) *Semicontinua inferiormente* (l.s.c.), si $\psi^{-1}(C)$ es abierto en X para cada conjunto abierto $C \subset A$.

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

(e) *Compacta*, si $\psi(x)$ es compacto para cada $x \in X$.

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

$$v^* \coloneqq \operatorname*{ínf}_{a \in \psi(x)} v(x, a).$$

Además, si $v(x, \cdot)$ alcanza su mínimo en algún punto de $\psi(x)$, se escribirá mín en lugar de ínf.

Definición C.2. La función $v : graf(\psi) \to \mathbb{R}$ se dice *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 C.1. Suponga que ψ es compacta,

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

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

y además, 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 C.2. Suponga 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

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

(b) Si además, la multifunción

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

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

Apéndice D

Esperanza Condicional y Martingalas

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

1. $E[\xi|\mathcal{F}']$ es $\mathcal{F}'$-medible.
2. Para todo $A \in \mathcal{F}'$, se tiene que $\int_A E[\xi|\mathcal{F}']\, dP = \int_A \xi dP$.

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

Proposición D.1. Sean ξ, η variables aleatorias P-integrables sobre $(\Omega, \mathcal{F}, P)$ y sean $\mathcal{F}'$ y $\mathcal{F}''$ σ-álgebras contenidas en $\mathcal{F}$, entonces:

a) Si ξ es una constante k, entonces $E[\xi|\mathcal{F}'] = k$.

b) $E[\xi + \eta|\mathcal{F}'] = E[\xi|\mathcal{F}'] + E[\eta|\mathcal{F}']$.

c) $E[E(\xi|\mathcal{F}')] = E[\xi]$.

d) Si ξ es $\mathcal{F}''$-medible, entonces $E[\xi\eta|\mathcal{F}'] = \xi E[\eta|\mathcal{F}']$, en particular, $E[\xi|\mathcal{F}'] = \xi$.

e) Si $\mathcal{F}' \subset \mathcal{F}''$, entonces $E[\xi|\mathcal{F}'] = E[E[\xi|\mathcal{F}']\,|\mathcal{F}''] = E[E[\xi|\mathcal{F}'']\,|\mathcal{F}']$.

f) Si $\xi_n \geq 0$, y $\xi_n \nearrow \xi$, entonces $E\left[\xi_n|\mathcal{F}'\right] \nearrow E\left[\xi|\mathcal{F}'\right]$.

g) Si $\xi_n \geq 0$, entonces $E\left[\sum_{n=1}^{\infty} \xi_n|\mathcal{F}'\right] = \sum_{n=1}^{\infty} E\left[\xi_n|\mathcal{F}'\right]$.

Definición D.1. Una filtración es una colección de $\sigma-$álgebras $\{\mathcal{F}_n\}_{n\geq 0}$ tal que $\mathcal{F}_n \subseteq \mathcal{F}_m$, siempre que $n \leq m$. En particular, la filtración natural o canónica de un proceso $\{X_n, n \geq 0\}$ es aquella sucesión de σ-álgebras definidas por

$$\mathcal{F}_n = \sigma(X_0, \ldots, X_n), \quad n = 0, 1, \ldots.$$

Definición D.2. Un proceso estocástico $\{X_n, n \geq 0\}$ es *adaptado* a una filtración $\{\mathcal{F}_n\}_{n\geq 0}$ si la variable X_n es $\mathcal{F}_n-$medible, para cada $n = 0, 1, \ldots$.

Definición D.3. Se dice que un proceso estocástico (a tiempo discreto) $\{X_n, n \geq 0\}$ es una *martingala* respecto de una filtración $\{\mathcal{F}_n\}_{n\geq 0}$ si cumple:

1. Es integrable.

2. Es adaptado a la filtración.

3. Para cualesquiera $n \leq m$,

$$P\left(E[X_m|\mathcal{F}_n] = X_n\right) = 1$$

Ejemplo D.1. Cualquier proceso integrable y adaptado puede transformarse en una martingala.
Sea $\{X_n, n \geq 0\}$ un proceso integrable adaptado a la filtración $\{\mathcal{F}_n\}_{n\geq 0}$. Defina $Y_0 = X_0 - E[X_0]$ y

$$Y_n = \sum_{k=1}^{n} X_k - E[X_k|\mathcal{F}_{k-1}], \quad \text{para } n \geq 1.$$

Entonces el proceso $\{Y_n, n \geq 0\}$ es una martingala respecto de la filtración $\{\mathcal{F}_n\}_{n\geq 0}$.

Demostración. En efecto, X_n es integrable pues es una suma finita de variables aleatorias integrables, más aún, como cada $Y_k \in \mathcal{F}_n$ para $1 \leq k \leq n$, se tiene que $X_n \in \mathcal{F}_n$ para cada $n \geq 0$. Queda probar que X_n cumple con

la propiedad de martingala, para ello note que:

$$\begin{aligned} E[X_{n+1}|\mathcal{F}_n] &= E\left[\sum_{k=1}^{n+1}(Y_k - E[Y_k|\mathcal{F}_{k-1}])\middle|\mathcal{F}_n\right] \\ &= \sum_{k=1}^{n} E\left[Y_k - E[Y_k|\mathcal{F}_{k-1}]|\mathcal{F}_n\right] + E\left[Y_{n+1} - E[Y_{n+1}|\mathcal{F}_n]|\mathcal{F}_n\right] \\ &= \sum_{k=1}^{n}(Y_k - E[Y_k|\mathcal{F}_{k-1}]) + E[Y_{n+1}|\mathcal{F}_n] - E[Y_{n+1}|\mathcal{F}_n] \\ &= X_n + 0 = X_n, \end{aligned}$$

es decir, para cada $n \geq 0$

$$E[X_{n+1}|\mathcal{F}_n] = X_n,$$

por lo tanto, $\{X_n : n \geq 0\}$ es una martingala.

■

Definición D.4. Una variable aleatoria τ con valores en $\{0, 1, \ldots\} \cup \{\infty\}$ es un *tiempo de paro* respecto de una filtración $\{\mathcal{F}_n\}_{n\geq 0}$ si para cada $n = 0, 1, \ldots$, se cumple que $(\tau \geq n) \in \mathcal{F}_n$.

Apéndice E

Operadores Contracción

Definición E.1. Sea (V, d) un espacio métrico. Se dice que una función $T : V \to V$ es un *operador contracción* si para algún $\alpha \in \mathbb{R}$, con $0 \leq \alpha < 1$, se tiene que

$$d(Tu, Tv) \leq \alpha d(u, v),$$

para cada u, $v \in V$.

α es llamado el *módulo de* T.

Definición E.2. Dada una función $T : V \to V$, la función T^n se define de manera recursiva por

$$T^n := T(T^{n-1}),$$

para cada $n \in \mathbb{N}$, además, T^0 es la función identidad.
Por otro lado, un elemento v^* de V es llamado un *punto fijo* de T si $Tv^* = v^*$.

Proposición E.1. (Teorema del Punto Fijo de Banach). Sea (V, d) un espacio métrico completo y $T : V \to V$ un operador contracción, entonces T tiene un único punto fijo, digamos, v^*. Además, para cualesquiera $v \in V$ y $n \geq 0$,

$$d(T^n v, v^*) \leq \alpha^n d(v, v^*).$$

donde α es el modulo de T.
La funciones $v_n := Tv_{n-1} = T^n v$ son llamadas las *aproximaciones sucesivas*.

En lo siguiente, el espacio V es usualmente $B(X)$, el espacio de funciones reales, acotadas y medibles de Banach sobre el espacio de Borel X, con la norma del supremo, $\|v\| := \sup_x |v(x)|$.

Proposición E.2. Sea $T : B(X) \to B(X)$ un operador y suponga que:

a) T es monótono, es decir, si u, $v \in B(X)$ y $u \leq v$, entonces $Tu \leq Tv$.

b) Exite un número $\alpha \in [0, 1)$, tal que, $T(v + c) = Tv + \alpha c$, para toda $v \in B(X)$ y cualquier constante c.

Entonces, T es una operador contracción con módulo α.

Demostración. Sean $u, v \in V$. Dado que $\|u - v\| = \sup_x |u(x) - v(x)|$, se satisface que

(1) $u(\cdot) \leq v(\cdot) + \|u - v\|$, y

(2) $v(\cdot) \leq u(\cdot) + \|u - v\|$.

Entonces, aplicando T a (1) y bajo las condiciones a) y b), se tiene que

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

o bien,

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

para cada $x \in X$. Luego, realizando el proceso análogo sobre (2), se verifica que

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

para cada $x \in X$. Por lo anterior, se obtiene

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

para cada $x \in X$, lo cual implica el resultado deseado, es decir,

$$\|Tv - Tu\| \leq \alpha\|u - v\|,$$

para cualesquiera $u, v \in B(X)$. ■

Bibliografía

[1] Abu Alsheikh, M., Hoang, D. T., Niyato, D., Tan, H. P., & Lin, S. (2015). *Markov Decision Processes with Applications in Wireless Sensor Networks: A Survey*. Communications Surveys & Tutorials, IEEE, 17(3), 1239-1267.

[2] Ash, R. B. (1972). *Real Analysis and Probability*. Academic Press. New York.

[3] Bäuerle, N., & Rieder, U. (2011). *Markov Decision Processes with Applications to Finance*. Springer Science & Business Media.

[4] Bertsekas, D. P. (1995). *Dynamic Programming and Optimal Control* (Vol. 1, No. 2). Belmont, MA: Athena Scientific.

[5] Bertsekas, D. P., & Shreve, S. E. (1978). *Stochastic Optimal Control: The Discrete Time Case*. New York: Academic Press.

[6] Burkhard, H., & Alfred, M. (2005). *Dynamic General Equilibrium Modelling: Computational Methods and Applications*. Berlin. Springer

[7] Chatterjee, D., Cinquemani, E., Chaloulos, G., & Lygeros, J. (2008). *Stochastic Control Up to a Hitting Time: Optimality and Rolling-Horizon Implementation*. arXiv preprint arXiv:0806.3008.

[8] Cruz-Suárez, H., Ilhuicatzi-Roldán, R., & Montes-de-Oca, R. (2014). *Markov Decision Processes on Borel Spaces with Total Cost and Random Horizon*. Journal of Optimization Theory and Applications, 162(1), 329-346.

[9] Florescu, A., Bratcu, A. I., Munteanu, I., Rumeau, A., & Bacha, S. (2015). *LQG Optimal Control Applied to On-Board Energy Management System of All-Electric Vehicles*. Control Systems Technology, IEEE Transactions on, 23(4), 1427-1439.

[10] Gut, A. (2012). *Probability: a Graduate Course*. Springer Science & Business Media.

[11] Hernández-Lerma, O. (1989). *Adaptive Markov Control Processes* (Vol. 79). Springer Science & Business Media.

[12] Hernández-Lerma, O., & Lasserre, J. B. (1996). *Discrete-Time Markov Control Processes: Basic Optimality Criteria* (Vol. 30). Springer Science & Business Media.

[13] Iida, T., & Mori, M. (1996). *Markov Decision Processes With Random Horizon*. Journal of the Operations Research Society of Japan, 39(4), 592-603.

[14] López, E., Barea, R., Bergasa, L. M., & Escudero, M. S. (2003). *Navegación Topológica mediante POMDPs Incorporando Información Visual*. Departamento de Electrónica, Universidad de Alcalá.

[15] Nishimura, K., & Stachurski, J. (2012). *Stability of Stochastic Optimal Growth Models: A New Approach*. In Nonlinear Dynamics in Equilibrium Models (pp. 289-307). Springer Berlin Heidelberg.

[16] Qian, F., Huang, J., Liu, D., & Hu, S. (2015). *Adaptive Dual Control of Discrete-Time LQG Problems with Unknown-But-Bounded Parameter*. Asian Journal of Control, 17(3), 942-951.

[17] Shi, W., & Guo, J. (2014). *Application of Markov Decision Processes (MDPs) in Petroleum Industry*. GSTF Journal of Engineering Technology (JET), 2(4), 7.

[18] Stokey, N. L. (1989). *Recursive Methods in Economic Dynamics*. Harvard University Press.

[19] Van Nunen, J. A. E. E., & Stidham Jr, S. (1981). *Action-Dependent Stopping Times and Markov Decision Process with Unbounded Rewards*. Operations-Research-Spektrum, 3(3), 145-152.

[20] Zacarías, G. (2007). *Procesos de Decisión de Markov Descontados*. Tesis Licenciatura, BUAP.