

BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA
FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS
POSTGRADO EN CIENCIAS MATEMÁTICAS

**PROCESOS DE DECISIÓN DE MARKOV CON
HORIZONTE ALEATORIO**

T E S I S

QUE PARA OBTENER EL GRADO DE:

DOCTOR EN CIENCIAS MATEMÁTICAS

PRESENTA:

MARIA DEL ROCIO ILHUICATZI ROLDÁN

DIRECTOR DE TESIS:

DR. HUGO ADÁN CRUZ SUÁREZ

PUEBLA, PUE. 2013

Agradecimientos

Agradezco enormemente al Consejo Nacional de Ciencia y Tecnología (CONACYT) la beca que me fue otorgada durante la realización de estudios de doctorado, siendo este apoyo indispensable para poder destinar el tiempo y la concentración necesarios que dicho proyecto profesional exige.

Introducción

En los problemas de control se estudia un sistema dinámico cuyo comportamiento es influenciado o regulado por un cambio conveniente en algunas de las variables del sistema, las cuales son llamadas variables de control o decisión. Los controles que son aplicados en un tiempo dado son seleccionados de acuerdo a reglas conocidas como políticas. Adicionalmente se tiene una función de rendimiento u objetivo, la cual mide o evalúa en algún sentido la respuesta del sistema a la política utilizada. De esta manera, el problema de control óptimo es determinar una política que optimice la función de rendimiento.

La clase de sistemas estocásticos controlados en tiempo discreto son conocidos como Procesos de Decisión de Markov y su teoría también es identificada con los nombres de optimización secuencial estocástica, control estocástico en tiempo discreto o programación dinámica estocástica. Referente a éstos, trata el presente trabajo de tesis.

Desde sus inicios, en 1950, los Procesos de Decisión de Markov proporcionaron una gran área de investigación con una teoría rica y con diversas aplicaciones. De hecho, los Procesos de Decisión de Markov proporcionan herramientas básicas para el análisis de muchos problemas en ingeniería (véase [24]), economía (véase [22]), finanzas (véase [29]), inteligencia artificial (véase [31]), telecomunicaciones (véase [2]), entre otras.

La técnica de programación dinámica, la cual es fundamental para el análisis de los Procesos de Decisión de Markov, fue estudiada sistemáticamente por Bellman en su libro *Dinamyc Programming* (véase [5]), así como en muchos de sus artículos. Sin embargo, los expertos consideran que el primer estudio sobre los Procesos de Decisión de Markov es la investigación sobre juegos estocásticos realizada por Shapley (véase [28]).

Los Procesos de Decisión de Markov son usualmente estudiados considerando un horizonte finito y cuando éste es suficientemente grande, se modela usando un horizonte infinito. Sin embargo, existe la posibilidad de que factores externos obliguen a concluir el proceso antes de lo planeado. De esta manera, es necesario considerar al horizonte como una variable aleatoria, la cual puede ser independiente del proceso. Esta clase de aleatoriedad en el horizonte puede ayudar a modelar, por ejemplo, la bancarrota en una modelo económico, la falla de un sistema en ingeniería o una catástrofe natural en biología (véase [4] p. 267, [12], [21], [23] y [27] p. 125).

Otro enfoque para este tipo de problemas, consiste en suponer al horizonte medible con respecto a la filtración del proceso de control (horizonte dependiente del proceso). Bajo este supuesto, una clase de problemas son estudiados considerando al horizonte como un tiempo de paro (véase [8]). Igualmente, es posible estudiar problemas de paro óptimo en el contexto de los Procesos de Decisión de Markov bajo el supuesto de dependencia del horizonte (véase [3] y [27]).

Los Procesos de Decisión de Markov considerando horizonte aleatorio independiente han sido estudiados ampliamente cuando la distribución del horizonte tiene soporte finito (véase [18], [21] y [23]), y en el caso de soporte infinito se resuelve el problema sólo cuando el horizonte tiene una distribución geométrica (véase [4] p. 267 y [27] p. 126).

En este trabajo de tesis, se retoma el estudio de los Procesos de Decisión de Markov con horizonte aleatorio independiente del proceso, considerando una distribución arbitraria para el horizonte, principalmente de soporte infinito. De manera general, se supone que los espacios de estados y de acciones son de Borel y la función de costo por etapa no necesariamente es acotada. Cabe señalar que estos últimos supuestos no se han considerado en artículos previos referentes al tema (véase [18] y [23]). El objetivo principal de esta tesis es caracterizar soluciones óptimas para el problema de control.

Como es observado en [18], el problema con un horizonte aleatorio de soporte infinito, puede estudiarse de forma equivalente como un problema descontado con factor de descuento no homogéneo o variante en el tiempo con horizonte infinito. La importancia que han adquirido los problemas descontados, motiva a estudiar el problema con horizonte aleatorio siguiendo precisamente la idea del factor de descuento no homogéneo.

Entonces, la solución óptima (función de valor y política óptima) del problema de control con horizonte aleatorio es caracterizada a través de técnicas usuales de programación dinámica (véase [13] y [17]). Además, se muestra un resultado que permite acotar la función de valor óptimo del problema con horizonte aleatorio por la función de valor óptimo de un problema descontado con factor de descuento fijo.

En este trabajo de tesis, también se incluyen aplicaciones de los resultados obtenidos en distintos problemas prácticos, por ejemplo, se ha resuelto un problema de reemplazo óptimo con horizonte aleatorio. El problema de reemplazo surge principalmente en los grandes sistemas productivos, en donde se desea determinar estrategias de reemplazo de máquinas u otros dispositivos, con el fin de incurrir en costos (recompensas) mínimos (máximas). Este problema ha sido de gran interés y se ha modelado de diferentes formas (véase [9], [25] y [30]). El ejemplo de reemplazo de máquinas presentado en el trabajo de tesis es una extensión del modelo propuesto en [6] p. 7, ahora considerando un sistema formado por n -máquinas en vez de una máquina y horizonte aleatorio en vez de horizonte finito. El horizonte aleatorio en este tipo de problemas puede justificarse debido a la asociación que tienen con los grandes sistemas productivos, por ejemplo, pensando en una bancarrota o en un cambio brusco en la producción debido a los constantes cambios en las necesidades del consumidor. Entonces, para el problema de reemplazamiento óptimo con horizonte aleatorio, cuando la distribución del horizonte tiene soporte finito, se determina una estrategia óptima. En el caso de soporte infinito, se implementa la técnica de horizonte rodante para hallar una aproximación de la solución óptima.

El trabajo de tesis se encuentra organizado de la siguiente manera: en el Capítulo 1, se presenta la teoría básica de los Procesos de Decisión de Markov. En el Capítulo 2, se muestran algunos antecedentes de los Procesos de Decisión de Markov con horizonte aleatorio y se plantea formalmente el problema de control óptimo correspondiente. En este capítulo se establece una equivalencia entre los Procesos de Decisión de Markov con horizonte aleatorio y los Procesos de Decisión de Markov con horizonte finito o infinito, dependiendo si la distribución del horizonte tiene soporte finito o infinito, respectivamente; por lo que en el Capítulo 3 se estudian los Procesos de Decisión de Markov con horizonte aleatorio cuyo soporte es finito y en el Capítulo 4 se aborda el caso donde

el horizonte tiene soporte infinito. Cada uno de estos capítulos incluyen sus respectivos ejemplos. Finalmente se presentan las conclusiones.

Índice general

Agradecimientos	1
Introducción	3
Índice general	7
1. Preliminares	9
1.1. Conceptos Básicos	9
1.2. Proceso de Decisión de Markov Descontado con Horizonte Infinito	13
2. Procesos de Decisión de Markov con Horizonte Aleatorio	17
2.1. Antecedentes	17
2.2. Planteamiento del Problema	19
3. Procesos de Decisión de Markov con Horizonte Aleatorio de Soporte Finito	23
3.1. Solución por Programación Dinámica	23
3.2. Ejemplos	27
3.2.1. Modelo Lineal-Cuadrático (LQ) con Horizonte Aleatorio	27
3.2.2. Problema de Reemplazo Óptimo con Horizonte Aleatorio: un Ejemplo Numérico	30
4. Procesos de Decisión de Markov con Horizonte Aleatorio de Soporte Infinito	41
4.1. Existencia de la Política Óptima y Caracterización de la Función de Valor Óptimo	42

4.1.1. El Procedimiento de Horizonte Rodante	55
4.1.2. Una Aplicación en la Aproximación de la Solución del Problema de Reemplazo Óptimo	56
4.2. El Problema de Control Óptimo con Horizonte Aleatorio como un Problema Descontado	61
Conclusiones	71
A. Programas	75
A.1. Programa para la Obtención de la Solución Óptima del Problema de Reemplazo	75
A.2. Programa para la Obtención de la Política de Horizonte Rodante	78
Bibliografía	87

Capítulo 1

Preliminares

Los conceptos básicos sobre los Procesos de Decisión de Markov presentados en este capítulo han sido retomados de [15], sin embargo, también se pueden consultar, por ejemplo, en [7] y [27]. De [15] es retomado el Teorema 4.2.3 y presentado en este capítulo como el Teorema 1.1, el cual caracteriza la solución óptima del problema de control óptimo descontado con horizonte infinito. El resultado mencionado será requerido en el Capítulo 4.

1.1. Conceptos Básicos

Definición 1.1 *Un modelo de control de Markov es una quintupla*

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

donde

- a) X es llamado el espacio de estados (el cual, en este trabajo de tesis, se considera como un espacio de Borel, esto es, un subconjunto Boreliano de un espacio métrico separable y completo);
- b) A es llamado el espacio de acciones (también considerado como un espacio de Borel);
- c) $\{A(x) : x \in X\}$ es una familia de subconjuntos no vacíos $A(x)$ de A medibles, donde $A(x)$ denota el conjunto de controles o acciones admisibles cuando el sistema está en el estado $x \in X$, con la

propiedad de que el conjunto de parejas admisibles estado-acción

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

es un subconjunto medible de $X \times A$.

d) Q es un kernel estocástico sobre X dado \mathbb{K} llamado la ley de transición. Es decir, es una función $Q : \mathcal{B}(X) \times \mathbb{K} \rightarrow \mathbb{R}$, donde $\mathcal{B}(X)$ es la σ -álgebra de Borel de X , tal que $Q(\cdot | k)$ es una medida de probabilidad sobre X para cada $k \in \mathbb{K}$ y $Q(B | \cdot)$ es una función medible sobre \mathbb{K} para cada $B \in \mathcal{B}(X)$.

e) $c : \mathbb{K} \rightarrow \mathbb{R}$ es una función medible, llamada la función de costo por etapa.

Observación 1.1 En muchos casos, la ley de transición puede ser especificada mediante una ecuación en diferencias de la siguiente forma:

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

$t = 0, 1, 2, \dots$, donde x_0 es el estado inicial y $\{\xi_t\}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas que toman valores en un espacio Boreliano S y tienen distribución común μ , independiente del estado inicial. Entonces, la ley de transición Q es dada por

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

donde $I_B(\cdot)$ denota la función indicadora del conjunto B .

Sea $(X, A, \{A(x) : x \in X\}, Q, c)$ un modelo de control de Markov y para cada $t = 0, 1, \dots$ se define el espacio \mathbb{H}_t de historias admisibles hasta el tiempo t como

$$\mathbb{H}_0 := X,$$

$$\mathbb{H}_t := \mathbb{K}^t \times X = \mathbb{K} \times \mathbb{H}_{t-1},$$

para $t = 1, 2, \dots$

Un elemento arbitrario h_t de \mathbb{H}_t , el cual es llamado una t -historia, es un vector de la forma

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

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

Definición 1.2 Una política de control es una sucesión $\pi = \{\pi_t, t = 0, 1, \dots\}$ de kérneles estocásticos π_t sobre el conjunto de control A dado \mathbb{H}_t tal que

$$\pi_t(A(x_t) | h_t) = 1,$$

$$h_t \in \mathbb{H}_t, t = 0, 1, \dots$$

El conjunto de todas las políticas es denotado por Π .

Sea Φ el conjunto de todos los kérneles estocásticos φ tales que $\varphi(A(x) | x) = 1$ para todo $x \in X$, y sea \mathbb{F} el conjunto de todas las funciones medibles $f : X \rightarrow A$ tales que $f(x) \in A(x)$ para todo $x \in X$. Las funciones en \mathbb{F} son llamadas selectores de la multifunción $x \mapsto A(x)$.

A continuación se presenta una clasificación de las políticas.

- 1) Una política se dice que es *Markoviana Aleatorizada* si existe una sucesión $\{\varphi_t\}$ de kérneles estocásticos $\varphi_t \in \Phi$ tales que

$$\pi_t(\cdot | h_t) = \varphi_t(\cdot | x_t) \quad \text{para todo } h_t \in \mathbb{H}_t, \quad t = 0, 1, 2, \dots$$

- 2) Una política se dice que es *Estacionaria Aleatorizada* si existe $\varphi \in \Phi$ tal que

$$\pi_t(\cdot | h_t) = \varphi(\cdot | x_t) \quad \text{para todo } h_t \in \mathbb{H}_t, \quad t = 0, 1, 2, \dots$$

- 3) Una política se dice que es *Determinista* si existe una sucesión $\{g_t\}$ de funciones medibles $g_t : \mathbb{H}_t \rightarrow A$ tal que, para todo $h_t \in \mathbb{H}_t$ y $t = 0, 1, 2, \dots$, $g_t(h_t) \in A(x_t)$ y $\pi_t(\cdot | h_t)$ está concentrada en $g_t(h_t)$, es decir

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

- 4) Una política se dice que es *Markoviana Determinista* 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 todo $h_t \in \mathbb{H}_t$, $t = 0, 1, 2, \dots$. El conjunto de todas las políticas deterministas Markovianas es denotado por Π_{DM} .

- 5) Una política se dice que es *Estacionaria Determinista* 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 todo $h_t \in \mathbb{H}_t$, $t = 0, 1, 2, \dots$. El conjunto de todas las políticas deterministas estacionarias es denotado por Π_{DS} .

Sea (Ω, \mathcal{F}) el espacio medible formado por el espacio muestral canónico $\Omega = \mathbb{H}_\infty := (X \times A)^\infty$ y \mathcal{F} la σ -álgebra producto correspondiente. Los elementos de Ω son sucesiones de la forma $\omega = (x_0, a_0, x_1, a_1, \dots)$ con $x_t \in X$ y $a_t \in A$ para todo $t = 0, 1, 2, \dots$. Las proyecciones x_t y a_t de Ω a los conjuntos X y A son llamadas variables de estado y acción, respectivamente.

Sea $\pi = \{\pi_t\}$ una política arbitraria y μ una medida de probabilidad arbitraria sobre X , llamada distribución inicial. Entonces, por el Teorema de C. Ionescu-Tulcea (véase [15]), existe una única medida de probabilidad P_μ^π sobre (Ω, \mathcal{F}) cuyo soporte es \mathbb{H}_∞ , i.e., $P_\mu^\pi(\mathbb{H}_\infty) = 1$. El proceso estocástico $(\Omega, \mathcal{F}, P_\mu^\pi, \{x_t\})$ es llamado proceso de control o decisión de Markov en tiempo discreto.

El operador esperanza con respecto a P_μ^π es denotado por E_μ^π . Si μ está concentrada en un estado inicial $x \in X$, entonces P_μ^π y E_μ^π son escritos como P_x^π y E_x^π , respectivamente.

Un criterio de rendimiento o una función objetivo permite medir la efectividad de una política. En la literatura comunmente se encuentran considerados los que a continuación se definen.

- a) **Costo total esperado.** Sea $\pi \in \Pi$ y $x \in X$. El costo total esperado con horizonte finito N se define por

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

donde c_N es una función medible en X , llamada *función de costo terminal*.

- b) **Costo total esperado descontado.** Sea $\pi \in \Pi$ y $x \in X$. El costo total esperado descontado con horizonte infinito se define por

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

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

En seguida, se presentan algunos conceptos, los cuales serán considerados en las suposiciones durante el desarrollo del trabajo de tesis.

Definición 1.3 Una función $c : \mathbb{K} \rightarrow \mathbb{R}$ se dice ser *semicontinua inferiormente* en $a \in A(x)$ para cada $x \in X$ si

$$\liminf_{n \rightarrow \infty} c(x, a_n) \geq c(x, a),$$

para cualquier sucesión $\{a_n\}$ en $A(x)$ tal que $a_n \rightarrow a$. La función c se dice ser *semicontinua inferiormente sobre $A(x)$* si es *semicontinua inferiormente* en cada punto de $A(x)$. Una función c es *semicontinua superiormente* si y sólo si $-c$ es *semicontinua inferiormente*.

Definición 1.4 Una función $c : \mathbb{K} \rightarrow \mathbb{R}$ se dice ser *inf-compacta* sobre \mathbb{K} si el conjunto $\{a \in A(x) : c(x, a) \leq r\}$ es compacto para cada $x \in X$ y $r \in \mathbb{R}$.

Definición 1.5 Un kernel estocástico Q se dice ser *débilmente continuo* si la función v' definida como

$$v'(x, a) := \int_X v(y)Q(dy \mid x, a),$$

es continua y acotada sobre \mathbb{K} para cada función continua y acotada v sobre X . Q se dice ser *fuertemente continuo* si la función v' es continua y acotada sobre \mathbb{K} para cada función medible y acotada v sobre X .

1.2. Proceso de Decisión de Markov Descuento con Horizonte Infinito

Considere el modelo de control de Markov $(X, A, \{A(x) : x \in X\}, Q, c)$ y como función de rendimiento el costo total esperado descontado, esto es,

$$v(\pi, x) := E_x^\pi \left[\sum_{t=0}^{\infty} \alpha^t c(x_t, a_t) \right]. \quad (1.1)$$

Defina la función de valor óptimo como

$$V(x) := \inf_{\pi \in \Pi} v(\pi, x), \quad x \in X.$$

El problema de control óptimo descontado consiste en encontrar una política $\pi^* \in \Pi$ tal que

$$v(\pi^*, x) = V(x), \quad x \in X.$$

En este caso, π^* es identificada como una *política óptima*.

Considere la siguiente suposición:

Suposición 1.1 (a) *La función de costo por etapa c es semicontinua inferiormente, no negativa e inf-compacta sobre \mathbb{K} .*

(b) *La ley de transición Q es fuertemente continua.*

(c) *Existe una política π tal que $v(\pi, x) < \infty$, para cada $x \in X$.*

El conjunto de todas las políticas que satisfacen la Suposición 1.1(c) es denotado por Π^0 .

Para resolver el problema descontado, se considera la Suposición 1.1 y los subproblemas truncados hasta la etapa n , con $n = 1, 2, 3, \dots$, siendo el criterio de rendimiento el siguiente:

$$v_n(\pi, x) := E_x^\pi \left[\sum_{t=0}^n \alpha^t c(x_t, a_t) \right]. \quad (1.2)$$

Entonces, la función de valor óptimo para cada subproblema queda definida como

$$V_n(x) := \inf_{\pi \in \Pi} v_n(\pi, x), \quad x \in X.$$

De acuerdo con el método de *iteración de valores* (véase [15]), se tiene que

$$\begin{aligned} V_0(x) &= 0, \\ V_n(x) &= \min_{a \in A(x)} \left[c(x, a) + \alpha \int_X V_{n-1}(y) Q(dy | x, a) \right], \end{aligned} \quad (1.3)$$

$x \in X$, $n = 1, 2, 3, \dots$

Luego, el Lema 4.2.8 en [15], garantiza que $\lim_{n \rightarrow \infty} V_n(x) = V(x)$ para todo $x \in X$, obteniéndose de esa manera la función de valor óptimo para el problema descontado con horizonte infinito. Cabe señalar que el método de iteración de valores funciona sólo para modelos homogéneos en el tiempo, ya que dicha propiedad es la que permite utilizar de manera iterada las funciones de valor óptimo V_n (véase (1.3)).

El Teorema 1.1 caracteriza la función de valor óptimo del problema descontado con horizonte infinito y permite hallar la política óptima, una vez que se haya determinado la función de valor óptimo. La demostración puede consultarse en [15].

Teorema 1.1 *Suponga que la Suposición 1.1 se cumple, entonces*

- (a) *La función de valor óptimo V del problema de control óptimo descontado satisface la siguiente ecuación de optimalidad:*

$$V(x) = \min_{a \in A(x)} \left[c(x, a) + \alpha \int_X V(y) Q(dy | x, a) \right], \quad x \in X, \quad (1.4)$$

y si u es otra solución de la ecuación de optimalidad, entonces $u \geq V$.

- (b) *Existe un selector $f_* \in \mathbb{F}$ tal que $f_*(x) \in A(x)$ logra el mínimo en la ecuación (1.4) para todo $x \in X$, es decir,*

$$V(x) = c(x, f_*(x)) + \int_X V(y) Q(dy | x, f_*(x)), \quad (1.5)$$

y la política determinista estacionaria f_^∞ es óptima; recíprocamente, si $f_*^\infty \in \Pi_{DS}$ es óptima, entonces se satisface (1.5).*

- (c) *Si π^* es una política tal que $v(\pi^*, \cdot)$ es una solución a la ecuación de optimalidad (1.4) y satisface que*

$$\lim_{n \rightarrow \infty} \alpha^n E_x^{\pi^*} [v(\pi^*, x_n)] = 0, \quad \pi \in \Pi^0, \quad x \in X,$$

entonces $v(\pi^, \cdot) = V(\cdot)$; por lo tanto π^* es óptima.*

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

Capítulo 2

Procesos de Decisión de Markov con Horizonte Aleatorio

En este capítulo se mencionan algunos antecedentes de los Procesos de Decisión de Markov con horizonte aleatorio, en el contexto del presente trabajo de tesis. También, se establecen las condiciones bajo las cuales es estudiado el problema de control óptimo con horizonte aleatorio y se plantea formalmente dicho problema.

2.1. Antecedentes

En la literatura se encuentran referencias en donde se estudian problemas de control a tiempo discreto con horizonte aleatorio, modelados mediante la teoría de los Procesos de Decisión de Markov. Sin embargo, se han considerado las siguientes condiciones restrictivas: distribución de probabilidad arbitraria para el horizonte con soporte finito o distribución geométrica (distribución de probabilidad con soporte infinito).

Por ejemplo, en [27], como ya se mencionó en la introducción, se analiza el problema de control óptimo con horizonte aleatorio en el caso en que el horizonte tiene distribución geométrica y se observa que éste coincide con el problema de control óptimo descontado con horizonte infinito, mostrando una equivalencia entre los criterios de rendimiento correspondientes a cada problema. En [27], se consideran espacios de

estados y acciones discretos.

En [23], se estudia un problema de consumo-inversión con incertidumbre en el horizonte usando un modelo de control de Markov, en donde, el principal objetivo es analizar el consumo inicial óptimo con respecto a diferentes distribuciones del horizonte con soporte finito. Para ello se supone un orden en las distribuciones de probabilidad de acuerdo con un factor de riesgo identificado en dichas distribuciones y se demuestra que a mayor riesgo en la distribución del horizonte, menor es el consumo inicial óptimo. También en [23], de manera introductoria, se menciona el caso en que la distribución del horizonte tiene distribución geométrica.

Otro artículo que considera aleatoriedad en el horizonte con soporte finito es [21], en donde se resuelve un problema de tipo lineal-cuadrático (LQ) matricial con horizonte aleatorio. Los problemas LQ se aplican principalmente en ingeniería, en donde el costo total puede ser interpretado como una abstracción de la energía. El problema es referido en este caso como problema regulador LQ (por ejemplo véase [24]).

Una referencia que estudia el problema suponiendo una distribución arbitraria para el horizonte es [18], en donde se considera el proceso de decisión de Markov no homogéneo con horizonte aleatorio, espacio de estados numerable y función de costo por etapa acotada. En dicho artículo se hacen observaciones sobre la verdadera problemática referente a la obtención de la solución óptima en el caso en que la distribución del horizonte tenga soporte infinito. En este caso, el problema con horizonte aleatorio se identifica con un problema de horizonte infinito no homogéneo, para el cual en general no es posible encontrar una solución exacta. Entonces, como una alternativa encontrada en la literatura, se proporciona un algoritmo basado en la técnica de horizonte rodante para aproximar la política óptima (o ϵ -óptima) cuando el soporte de la distribución del horizonte es infinito, dando como ejemplo una aplicación a modelos de control de inventarios.

Considerando estos antecedentes del problema de control óptimo con horizonte aleatorio, recuérdese que uno de los objetivos principales de este trabajo de tesis es abordar casos más generales que los ya estudiados, haciéndolo principalmente en dos sentidos:

1. Suponiendo espacios de estados y acciones de Borel y costos por etapa no acotados.

2. Considerando una distribución arbitraria para el horizonte (de soporte finito o infinito).

2.2. Planteamiento del Problema

Sea τ una variable aleatoria discreta asociada a un espacio de probabilidad $(\Omega', \mathcal{F}', P)$. Suponga que la distribución de τ es conocida, dada por $P(\tau = t) := \rho_t$, $t = 0, 1, 2, \dots, T$, donde T es un número entero positivo o $T = \infty$. Considere un modelo de control de Markov $(X, A, \{A(x) : x \in X\}, Q, c)$ y defina como criterio de rendimiento

$$j^\tau(\pi, x) := E \left[\sum_{t=0}^{\tau} c(x_t, a_t) \right],$$

$\pi \in \Pi$, $x \in X$, donde E denota el valor esperado con respecto a la distribución conjunta del proceso $\{(x_t, a_t)\}$ y τ .

j^τ es identificado como un costo total esperado con horizonte aleatorio.

Luego, considere el correspondiente problema de control óptimo. Para ello, defina la función de valor óptimo como

$$J^\tau(x) := \inf_{\pi \in \Pi} j^\tau(\pi, x), \quad (2.1)$$

$x \in X$.

De esa manera, el problema de control óptimo con horizonte aleatorio consiste en encontrar una política $\pi^* \in \Pi$ tal que $j^\tau(\pi^*, x) = J^\tau(x)$, $x \in X$.

Considere la siguiente suposición.

Suposición 2.1 *Para cada $x \in X$ y $\pi \in \Pi$ el proceso inducido $\{(x_t, a_t) \mid t = 0, 1, 2, \dots\}$ es independiente de τ .*

Entonces, bajo la Suposición 2.1 se tiene que

$$\begin{aligned}
j^\tau(\pi, x) &= E \left[\sum_{t=0}^{\tau} c(x_t, a_t) \right] \\
&= E \left[E \left[\sum_{t=0}^{\tau} c(x_t, a_t) \middle| \tau \right] \right] \\
&= \sum_{n=0}^T E_x^\pi \left[\sum_{t=0}^n c(x_t, a_t) \right] \rho_n \\
&= \sum_{t=0}^T \sum_{n=t}^T E_x^\pi [c(x_t, a_t)] \rho_n \\
&= E_x^\pi \left[\sum_{t=0}^T P_t c(x_t, a_t) \right], \tag{2.2}
\end{aligned}$$

$\pi \in \Pi$, $x \in X$, donde $P_t := \sum_{n=t}^T \rho_n = P(\tau \geq t)$, $t = 0, 1, 2, \dots, T$. Con esto se muestra una equivalencia entre un problema de control óptimo con horizonte aleatorio τ y un problema de control óptimo con horizonte $T + 1$, con costo por etapa no homogéneo dado por $P_t c$ y costo terminal igual a cero.

Debido a las diferentes técnicas utilizadas en cada caso, se tiene la necesidad de abordar el problema de manera separada, destinando el Capítulo 3 para el caso $T < \infty$ y el Capítulo 4 para el caso $T = \infty$.

A continuación presentamos dos casos interesantes encontrados en la literatura, los cuales motivan el presente trabajo de tesis.

- I. Obsérvese que cuando el horizonte tiene distribución geométrica, con parámetro p , $0 < p < 1$, tenemos que $P(\tau = n) = pq^n$, $n = 0, 1, 2, \dots$, donde $q = 1 - p$. Por lo tanto,

$$\begin{aligned}
P_t &= \sum_{n=t}^{\infty} pq^n \\
&= pq^t \sum_{n=0}^{\infty} q^n, \\
&= q^t
\end{aligned}$$

$t = 0, 1, 2, \dots$. Entonces, sustituyendo P_t en (2.2), se obtiene el criterio del problema descontado con horizonte infinito con factor de descuento igual a q (véase (1.1) en el Capítulo 1), tal y como se ha mencionado en [18], [23] y [27]. En este caso puede aplicarse el Teorema 1.1. De esta forma se sustenta la idea de que un problema con horizonte aleatorio, aun con otra distribución, pueda estudiarse como un problema descontado, siendo ésta la manera en que se ha abordado el problema.

II. Cuando el horizonte tiene distribución binomial negativa con parámetros p y $r = 2$, donde $0 < p < 1$, esto es,

$$P(\tau = n) = \binom{2+n-1}{n} p^2 q^n,$$

$n = 0, 1, 2, \dots$, con $q = 1 - p$, resulta que

$$\begin{aligned} P_t &= \sum_{n=t}^{\infty} (n+1) p^2 q^n \\ &= p^2 q^t \sum_{n=0}^{\infty} (n+t+1) q^n \\ &= p^2 q^t \left(\frac{q}{(1-q)^2} + \frac{t+1}{1-q} \right) \\ &= q^{t+1} + (1-q)(t+1)q^t, \end{aligned}$$

sustituyendo P_t en (2.2), se obtiene que

$$j^\tau(\pi, x) = qv(\pi, x) + (1-q) \frac{\partial(qv(\pi, x))}{\partial q},$$

donde v es el criterio de rendimiento para el problema descontado con factor de descuento q dado en (1.1). El resultado sólo permite expresar el criterio de rendimiento del problema con horizonte aleatorio en términos del criterio de rendimiento del problema descontado con horizonte infinito y, obviamente, no se sigue una relación similar entre las respectivas funciones de valor óptimo. En la Sección 2 del Capítulo 4, se logra establecer una relación entre la función de valor óptimo de un problema con horizonte aleatorio

de soporte infinito con la función de valor óptimo de un problema descontado con horizonte infinito, bajo una condición en la distribución del horizonte.

Capítulo 3

Procesos de Decisión de Markov con Horizonte Aleatorio de Soporte Finito

En este capítulo se estudia el caso en que la distribución del horizonte tiene soporte finito, es decir, cuando $T < +\infty$. En este caso, el problema de control óptimo puede resolverse mediante programación dinámica. En la primera sección de éste capítulo se proporciona el teorema correspondiente a programación dinámica adaptado al problema con horizonte aleatorio en el contexto dado. En la segunda sección se muestran dos ejemplos en donde se ha aplicado el resultado.

3.1. Solución por Programación Dinámica

Recuerde que el criterio original para el problema de control óptimo con horizonte aleatorio es

$$j^\tau(\pi, x) := E \left[\sum_{t=0}^{\tau} c(x_t, a_t) \right],$$

$\pi \in \Pi$ y $x \in X$, donde τ es una variable aleatoria con distribución conocida tal que $P(\tau = t) = \rho_t$, $t = 0, 1, 2, \dots, T$, y que la función de valor óptimo es denotada por $J^\tau(x)$, $x \in X$.

Suponga que $T < +\infty$. Entonces, como ya se observó en (2.2), en este caso el criterio de rendimiento se reduce al siguiente:

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

$\pi \in \Pi$, $x \in X$, donde $P_t = P(\tau \geq t)$.

Considere la siguiente suposición.

Suposición 3.1 (a) *La función de costo por etapa c es semicontinua inferiormente, no negativa e inf-compacta en \mathbb{K}*

(b) *La ley de transición Q es fuertemente continua o débilmente continua.*

El siguiente teorema proporciona la ecuación de programación dinámica que permite resolver el problema de control óptimo con horizonte aleatorio cuando la distribución del horizonte tiene soporte finito ($T < \infty$).

Teorema 3.1 *Sean J_0, J_1, \dots, J_{T+1} funciones sobre X definidas por*

$$J_{T+1}(x) := 0$$

y para $t = T, T-1, \dots, 0$,

$$J_t(x) := \min_{a \in A(x)} \left[P_t c(x, a) + \int_X J_{t+1}(y) Q(dy | x, a) \right], \quad x \in X. \quad (3.1)$$

Bajo la Suposición 3.1, estas funciones son medibles y para cada $t = 0, 1, \dots, T$, existe $f_t \in \mathbb{F}$ tal que $f_t(x) \in A(x)$ alcanza el mínimo en (3.1) para todo $x \in X$, esto es

$$J_t(x) = P_t c(x, f_t(x)) + \int_X J_{t+1}(y) Q(dy | x, f_t(x)),$$

$x \in X$ y $t = 0, 1, \dots, T$. Entonces la política Markoviana determinista $\pi^* = \{f_0, \dots, f_T\}$ es óptima y la función de valor óptimo está dada por

$$J^\tau(x) = j^\tau(\pi^*, x) = J_0(x),$$

$x \in X$.

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

$$C_t(\pi, x) := E^\pi \left[\sum_{n=t}^T P_n c(x_n, a_n) \middle| x_t = x \right], \quad t = 0, 1, \dots, T,$$

$$C_{T+1}(\pi, x) := 0.$$

Observe que

$$j^\pi(\pi, x) = C_0(\pi, x). \quad (3.2)$$

Ahora, se demostrará por inducción que para todo $x \in X$ y $t = 0, 1, \dots, T + 1$

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

y con $\pi = \pi^*$

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

Realizando inducción hacia atrás, se tiene que para $t = T + 1$

$$C_{T+1}(\pi, x) = 0 = J_{T+1}(x).$$

Luego, suponga que para algún $t = T, T - 1, \dots, 0$

$$C_{t+1}(\pi, x) \geq J_{t+1}(x), \quad x \in X.$$

Ahora,

$$\begin{aligned} C_t(\pi, x) &= E^\pi \left[P_t c(x_t, a_t) + \sum_{n=t+1}^T P_n c(x_n, a_n) \middle| x_t = x \right] \\ &= \int_A \left[P_t c(x, a) + \int_X C_{t+1}(\pi, y) Q(dy \mid x, a) \right] \pi_t(da \mid x), \end{aligned}$$

y por la hipótesis de inducción

$$\begin{aligned} C_t(\pi, x) &\geq \int_A \left[P_t c(x, a) + \int_X J_{t+1}(y) Q(dy \mid x, a) \right] \pi_t(da \mid x) \\ &\geq \min_{a \in A(x)} \left[P_t c(x, a) + \int_X J_{t+1}(y) Q(dy \mid x, a) \right] \end{aligned}$$

Por lo tanto se concluye que $C_t(\pi, x) \geq J_t(x)$.

Por otro lado, si $C_{t+1}(\pi, x) = J_{t+1}(x)$, $x \in X$ con $\pi = \pi^*$, $\pi_t(\cdot \mid h_t)$

es la medida de Dirac concentrada en $f_t(x_t)$, entonces se mantiene la igualdad en los cálculos anteriores, obteniendo $C_t(\pi^*, x) = J_t(x)$. Luego, si $C_t(\pi, x) \geq J_t(x)$, en particular para $t = 0$ y de (3.2)

$$j^\tau(\pi, x) \geq J_0(x)$$

y para $\pi = \pi^*$

$$j^\tau(\pi^*, x) = J_0(x),$$

quedando probado de esa manera el Teorema 3.1. ■

El siguiente cambio de variable permitirá analizar el problema de control óptimo con horizonte aleatorio como un problema descontado con factor de descuento variante en el tiempo, en el caso $T = \infty$ presentado en el Capítulo 4. Sea $U_{T+1}(x) := J_{T+1}(x)$ y

$$U_t(x) := \frac{J_t(x)}{P_t},$$

$t = 0, 1, 2, \dots, T$, con $P_t = P(\tau \geq t)$. Sustituyendo en (3.1), se obtiene la siguiente expresión:

$$P_t U_t(x) = \min_{a \in A(x)} \left[P_t c(x, a) + \int_X P_{t+1} U_{t+1}(y) Q(dy | x, a) \right],$$

de donde, considerando que $P_t > 0$, se obtiene que

$$U_t(x) = \min_{a \in A(x)} \left[c(x, a) + \frac{P_{t+1}}{P_t} \int_X U_{t+1}(y) Q(dy | x, a) \right].$$

Como $P_0 = 1$, se tiene que

$$U_0(x) = J_0(x).$$

Entonces la ecuación de programación dinámica (3.1) es equivalente a

$$U_t(x) = \min_{a \in A(x)} \left[c(x, a) + \alpha_t \int_X U_{t+1}(y) Q(dy | x, a) \right], \quad (3.3)$$

donde

$$\alpha_t := \frac{P_{t+1}}{P_t}, \quad t = 0, 1, 2, \dots, T. \quad (3.4)$$

Además, obsérvese que puede obtenerse la siguiente interpretación probabilística:

$$\begin{aligned}\alpha_t &= \frac{P(\tau \geq t+1, \tau \geq t)}{P(\tau \geq t)} \\ &= P(\tau \geq t+1 \mid \tau \geq t).\end{aligned}$$

3.2. Ejemplos

En esta sección se presentan dos modelos en donde se ha resuelto el problema de control óptimo con horizonte aleatorio, suponiendo que la distribución del horizonte tiene soporte finito.

3.2.1. Modelo Lineal-Cuadrático (LQ) con Horizonte Aleatorio

Considere $X = A = A(x) = \mathbb{R}$. En este tipo de modelos, la función de costo es dada por la siguiente forma cuadrática:

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

$(x, a) \in \mathbb{K}$, donde $q \geq 0$ y $r > 0$. La ley de transición es inducida por una ecuación en diferencias (véase Observación 1.1), generalmente expresada de la siguiente manera lineal:

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

$t = 0, 1, 2, \dots, \tau$, con x_0 conocido. En este caso, $\gamma, \beta \in \mathbb{R}$ y $\{\xi_t\}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas que toman valores en $S = \mathbb{R}$ con función de densidad Δ continua y acotada, tal que $E[\xi_0] = 0$ y $E[\xi_0^2] = \sigma^2 < +\infty$, donde ξ_0 es un elemento genérico de la sucesión $\{\xi_t\}$. Además, se supone que la distribución del horizonte τ tiene soporte finito, esto es $P(\tau = k) = \rho_k$, $k = 0, 1, 2, \dots, T$, $T < \infty$.

Lema 3.1 *El Modelo LQ con horizonte aleatorio satisface la Suposición 3.1.*

Demostración. Obviamente, c es no negativa y continua. Luego, sea $A_\lambda(c) := \{a \in A(x) : c(x, a) \leq \lambda\}$, $\lambda \in \mathbb{R}$. Entonces,

$$A_\lambda(c) = \begin{cases} \emptyset & \text{si } \lambda < qx^2 \\ \{0\} & \text{si } \lambda = qx^2 \\ \left[-\sqrt{\frac{\lambda - qx^2}{r}}, \sqrt{\frac{\lambda - qx^2}{r}}\right] & \text{si } \lambda > qx^2 \end{cases}$$

Como $A_\lambda(c)$ es compacto para cada $x \in \mathbb{R}$, se tiene que c es una función inf-compacta en \mathbb{K} . Resta probar que la ley de transición es fuertemente continua. Sea $v : X \rightarrow \mathbb{R}$ una función medible y acotada. Observe que

$$\begin{aligned} v'(x, a) &= \int_X v(y)Q(dy | x, a) \\ &= \int_{-\infty}^{\infty} v(\gamma x + \beta a + s)\Delta(s)ds. \end{aligned}$$

Haciendo el cambio de variable $u = \gamma x + \beta a + s$, se obtiene que

$$v'(x, a) = \int_{-\infty}^{\infty} v(u)\Delta(u - \gamma x - \beta a)du.$$

Sea $\{(x_k, a_k)\}$ una sucesión tal que $\lim_{k \rightarrow \infty} (x_k, a_k) = (x', a')$. Por el Teorema de la Convergencia Dominada y por ser Δ continua, $\lim_{k \rightarrow \infty} v'(x_k, a_k) = v'(x', a')$, por lo tanto v' es continua sobre \mathbb{K} y acotada. Así, la ley de transición para el Modelo LQ es fuertemente continua. ■

Lema 3.2 *La política óptima π^* y la función de valor óptimo J^τ para el modelo LQ con horizonte aleatorio son dadas por $\pi^* = (f_0, f_1, \dots, f_T)$, donde:*

$$f_t(x) = \frac{-\alpha_t C_{t+1} \gamma \beta}{r + \alpha_t C_{t+1} \beta^2} x, \quad t = T, T-1, \dots, 0,$$

y $J^\tau(x) = C_0 x^2 + D_0$, $x \in X$, donde las constantes C_t y D_t satisfacen las siguientes relaciones de recurrencia:

$$\begin{aligned} C_{T+1} &= 0, \\ C_t &= \frac{qr + \alpha_t C_{t+1} (q\beta^2 + r\gamma^2)}{r + \alpha_t C_{t+1} \beta^2}, \quad t = T, T-1, \dots, 0, \end{aligned}$$

y

$$\begin{aligned} D_{T+1} &= 0, \\ D_t &= \alpha_t(C_{t+1}\sigma^2 + D_{t+1}), \quad t = T, T-1, \dots, 0, \end{aligned}$$

con

$$\alpha_t = \frac{P(\tau \geq t+1)}{P(\tau \geq t)}.$$

Demostración. En este caso, usando (3.3), la ecuación de programación dinámica resulta ser la siguiente:

$$U_t(x) = \min_{a \in \mathbb{R}} [qx^2 + ra^2 + \alpha_t E[U_{t+1}(\gamma x + \beta a + \xi)]] . \quad (3.5)$$

Para $t = T$ en (3.5), con $U_{T+1}(x) = 0$, se obtiene que $f_T(x) = 0$ y $U_T(x) = qx^2$, $x \in X$.

Para $t = T-1$, sustituyendo U_T en (3.5), se obtiene que

$$\begin{aligned} U_{T-1}(x) &= \min_{a \in \mathbb{R}} [qx^2 + ra^2 + \alpha_{T-1} E[q(\gamma x + \beta a + \xi)^2]] \\ &= \min_{a \in \mathbb{R}} [qx^2 + ra^2 + \alpha_{T-1} q(\gamma^2 x^2 + \beta^2 a^2 + 2\gamma\beta ax + \sigma^2)] \\ &= \min_{a \in \mathbb{R}} [(r + \alpha_{T-1} q\beta^2)a^2 + 2\alpha_{T-1} q\gamma\beta xa + (\alpha_{T-1} q\gamma^2 + q)x^2 + \alpha_{T-1} q\sigma^2], \end{aligned}$$

derivando la expresión entre corchetes con respecto a la variable a e igualando a cero para la minimización, se obtiene que

$$f_{T-1}(x) = \frac{-\alpha_{T-1} C_T \gamma \beta}{r + \alpha_{T-1} C_T \beta^2} x,$$

donde $C_T = q$. Por lo tanto $U_{T-1}(x) = C_{T-1}x^2 + D_{T-1}$, donde

$$C_{T-1} = \frac{qr + \alpha_{T-1} C_T (q\beta^2 + r\gamma^2)}{r + \alpha_{T-1} C_T \beta^2}$$

y $D_{T-1} = C_T \alpha_{T-1} \sigma^2$.

Continuando con este procedimiento, se sigue que

$$f_{T-2}(x) = \frac{-\alpha_{T-2} C_{T-1} \gamma \beta}{r + \alpha_{T-2} C_{T-1} \beta^2} x$$

y $U_{T-1}(x) = C_{T-2}x^2 + D_{T-2}$, donde

$$C_{T-2} = \frac{qr + \alpha_{T-2}C_{T-1}(q\beta^2 + r\gamma^2)}{r + \alpha_{T-2}C_{T-1}\beta^2}$$

y $D_{T-2} = \alpha_{T-2}(C_{T-1}\sigma^2 + D_{T-1})$.

Finalmente, con $t = 0$, se obtiene que

$$f_0(x) = \frac{-\alpha_0 C_1 \gamma \beta}{r + \alpha_0 C_1 \beta^2} x,$$

y $U_0(x) = C_0x^2 + D_0$, donde

$$C_0 = \frac{qr + \alpha_0 C_1 (q\beta^2 + r\gamma^2)}{r + \alpha_0 C_1 \beta^2}$$

y $D_0 = \alpha_0(C_1\sigma^2 + D_1)$. Como la Suposición 3.1 se cumple, queda demostrado el Teorema 3.1. ■

3.2.2. Problema de Reemplazo Óptimo con Horizonte Aleatorio: un Ejemplo Numérico

En este ejemplo, se considera un sistema formado por n máquinas y por simplicidad se puede suponer que las máquinas son similares y funcionan independientemente una de otra. Para cada máquina se identifica un proceso estocástico de deterioro independiente, cuyos posibles niveles de deterioro son denotados por $1, 2, 3, \dots, D$, donde se supone que D es un número entero positivo. El nivel uno indica que la máquina está en perfectas condiciones y se considera que el nivel i es mejor que el nivel $i + 1$, $i = 1, 2, 3, \dots, D - 1$.

Suponga que la matriz de probabilidades de transición $P = (p_{i,j})_{D \times D}$ es conocida (idéntica para las n -máquinas por ser similares), donde $p_{i,j}$ representa la probabilidad de que una máquina pase del nivel i al nivel j . Debido a que una máquina no puede pasar a un mejor nivel de deterioro, se tiene que $p_{i,j} = 0$ si $j < i$.

Sea $g : \{1, 2, 3, \dots, D\} \rightarrow \mathbb{R}$ una función conocida, donde $g(i)$, $i = 1, 2, 3, \dots, D$ representa el costo de operación de una máquina cuando se encuentra en el i -ésimo nivel de deterioro. Para esta función g se supone que

$$g(1) \leq g(2) \leq \dots \leq g(D),$$

es decir, se supone que g es no decreciente.

También, se considera que al inicio de cada periodo de tiempo puede tomarse una de las siguientes dos acciones:

- a) operar la k -ésima máquina, $k = 1, 2, \dots, \mathbf{n}$, con el nivel de deterioro en que se encuentra durante ese periodo de tiempo, o
- b) reemplazarla por una nueva pagando un costo fijo $R > 0$.

Además, considere que el horizonte de operación del sistema es aleatorio, y puede ser asociado con una variable aleatoria τ independiente del proceso seguido por el sistema, cuya distribución de probabilidad $P(\tau = t) = \rho_t$, $t = 0, 1, 2, \dots, T$, es conocida, donde T es un número entero positivo.

De esa forma, el problema de reemplazo óptimo consiste en determinar políticas de reemplazo que minimicen el costo total de operación del sistema. Dicho problema se ha modelado como un problema de control óptimo con horizonte aleatorio de soporte finito. A continuación se define el modelo de control de Markov correspondiente al problema de reemplazo óptimo.

Para definir el espacio de estados X , obsérvese que al inicio de un periodo de tiempo arbitrario, el estado del sistema puede registrarse mediante el arreglo $(d_1, d_2, \dots, d_{\mathbf{n}})$ donde d_k , $k = 1, 2, \dots, \mathbf{n}$, representa el nivel de deterioro con el cual la k -ésima máquina está operando. Así,

$$X = \{(d_1, d_2, \dots, d_{\mathbf{n}}) | d_k \in \{1, 2, \dots, D\}, k = 1, 2, \dots, \mathbf{n}\}, \quad (3.6)$$

donde la cardinalidad de X es $D^{\mathbf{n}}$.

Para definir el espacio de acciones A , puede considerarse como una acción el siguiente arreglo: $(a_1, a_2, \dots, a_{\mathbf{n}})$ con $a_k = 0$ o $a_k = 1$, donde $a_k = 0$ significa dejar que la k -ésima máquina opere con el nivel de deterioro d_k y $a_k = 1$ significa reemplazarla. Por lo tanto

$$A = A(x) = \{(a_1, a_2, \dots, a_{\mathbf{n}}) | a_k \in \{0, 1\}, k = 1, 2, \dots, \mathbf{n}\}, \quad (3.7)$$

donde su cardinalidad resulta ser $2^{\mathbf{n}}$.

Luego, para una máquina arbitraria k , sea $P^0 = (p_{i,j}^0)_{D \times D}$ la matriz de transición del proceso de deterioro cuando la máquina no es reemplazada. Obsérvese que $P^0 = P$. Sea $P^1 = (p_{i,j}^1)_{D \times D}$ la matriz de transición cuando

la máquina es reemplazada. En este caso, $p_{i,j}^1 = 1$, si $j = 1$ y $p_{i,j}^1 = 0$ en otro caso (con seguridad, cuando la máquina k es reemplazada, pasa al nivel uno o de perfectas condiciones). Defina $Q^a = (q_{i,j}^a)_{D^n \times D^n}$ como la matriz de transición del estado i al estado j del sistema, cuando la acción $a \in A$ es tomada, $i, j \in X$. Por la independencia de los procesos de deterioro de las máquinas, se tiene que

$$q_{(i_1, i_2, \dots, i_n), (j_1, j_2, \dots, j_n)}^{(a_1, a_2, \dots, a_n)} = p_{i_1, j_1}^{a_1} \cdot p_{i_2, j_2}^{a_2} \cdot \dots \cdot p_{i_n, j_n}^{a_n}. \quad (3.8)$$

De ese modo queda determinada la ley de transición para el proceso de control de Markov.

Finalmente, se debe definir la función de costo por etapa. Para ello obsérvese que para cualquier periodo de tiempo t

$$c(x_t, a_t) = \sum_{k=1}^n \gamma(x_{k,t}, a_{k,t}),$$

donde

$$\gamma(x_{k,t}, a_{k,t}) = \begin{cases} g(x_{k,t}) & \text{si } a_{k,t} = 0 \\ g(1) + R & \text{si } a_{k,t} = 1. \end{cases}$$

En este caso $x_{k,t}$ y $a_{k,t}$ representan el estado y la acción en el tiempo t de la máquina k , respectivamente.

Por lo tanto, la función objetivo para el problema de reemplazo óptimo está dada por la siguiente expresión:

$$j^\tau(\pi, x) := E \left[\sum_{t=0}^{\tau} c(x_t, a_t) \right],$$

$\pi \in \Pi$ y $x \in X$, donde τ es el horizonte de planeación del problema.

Entonces, usando el Teorema 3.1 de este capítulo, la política óptima de reemplazo se obtiene a partir de la siguiente ecuación de programación dinámica:

$$\begin{aligned} J_{T+1}(x) &= 0, \\ J_t(x) &= \min_{a \in A(x)} \left[P_t c(x, a) + \sum_{y \in X} J_{t+1}(y) q_{x,y}^a \right], \end{aligned} \quad (3.9)$$

$t = 0, 1, \dots, T$, donde $J_0(x)$ es el valor óptimo esperado correspondiente al estado inicial x .

Observación 3.1 *La suposición de medibilidad y existencia de selectores en el Teorema 3.1 se cumple, debido a que X y A son finitos (véase [27], p. 90).*

Para iterar la ecuación (3.9) se ha implementado un programa en Matlab para facilitar los cálculos numéricos. Dicho programa es presentado en el Apéndice de este trabajo de tesis como el programa A.1.

A continuación se presenta el algoritmo que se siguió para la elaboración del programa.

1. Introducir los siguientes datos:
 - n , el número de máquinas.
 - D , el número de niveles de deterioro.
 - P , la matriz de probabilidades de transición del proceso de deterioro de una máquina.
 - $g(i)$, el costo por operar una máquina con el nivel de deterioro i , $i = 1, 2, 3, \dots, D$.
 - R , el costo por unidad reemplazada.
 - T , el valor máximo que puede tomar el horizonte.
 - ρ_t , $t = 1, 2, 3, \dots, T$ distribución de probabilidad del horizonte.
2. Construir el espacio de estados X , usando (3.6).
3. Construir el espacio de acciones A , usando (3.7).
4. Calcular las matrices de transición Q^a para cada $a \in A$, por medio de (3.8).
5. Obtener la política óptima y la función de valor óptimo, como sigue
 - 5.1. Hacer $t = T + 1$ y $J_t(x) = 0$ para cada $x \in X$.
 - 5.2. Si $t = 0$, terminar, entonces $J_t(x)$ es la función de valor óptimo y $\{f_0, f_1, \dots, f_{T+1}\}$ es la política óptima. En otro caso, ir al paso 5.3.

5.3. Reemplazar t por $t - 1$ y calcular $J_t(x)$ para $x \in X$ por medio de la ecuación

$$J_t(x) = \min_{a \in A(x)} \left[P_t c(x, a) + \sum_{y \in X} J_{t+1}(y) q_{x,y}^a \right],$$

haciendo

$$f_t(x) = a,$$

para algún

$$a \in \arg \min_{a \in A(x)} \left[P_t c(x, a) + \sum_{y \in X} J_{t+1}(y) q_{x,y}^a \right].$$

Regresar al paso 5.2.

Ejemplos numéricos

Considere el problema de reemplazo óptimo con horizonte aleatorio y los siguientes valores numéricos: $\mathbf{n} = 3$, $D = 3$,

$$P = \begin{pmatrix} 0.4 & 0.3 & 0.3 \\ 0 & 0.3 & 0.7 \\ 0 & 0 & 1 \end{pmatrix},$$

$$g = (5, 7, 29), R = 4, T = 6 \text{ y } \rho = (0.1, 0.1, 0.3, 0.2, 0.15, 0.15).$$

En la Tabla 3.1, se presenta el costo total esperado óptimo para cada estado inicial.

En la Tabla 3.2, se proporciona la acción óptima por etapa dependiendo del estado en que se encuentre el sistema.

Adicionalmente se ha establecido lo siguiente: Para un número de máquinas \mathbf{n} , considere el estado de perfectas condiciones del sistema como el estado inicial, esto es $x_0 = (1, 1, \dots, 1)_{\mathbf{n} \times 1}$. Sea $J_{\mathbf{n}}^r(x_0)$ el valor óptimo esperado para el problema de reemplazo óptimo correspondiente. En la Tabla 3.3, se muestra el valor esperado óptimo para distintos valores de \mathbf{n} , cuya gráfica es mostrada en la Figura 3.1.

Note que existe una aparente relación lineal entre el número de máquinas y el valor esperado óptimo, con estado inicial x_0 . Además, puede

Estado inicial x	Valor óptimo $J_\tau(x)$	Estado inicial x	Valor óptimo $J_\tau(x)$
1 1 1	67.8226	1 1 2	70.4790
1 1 3	70.4790	1 2 1	70.4790
1 2 2	73.1353	1 2 3	73.1353
1 3 1	70.4790	1 3 2	73.1353
1 3 3	73.1353	2 1 1	70.4790
2 1 2	73.1353	2 1 3	73.1353
2 2 1	73.1353	2 2 2	75.7917
2 2 3	75.7917	2 3 1	73.1353
2 3 2	75.7917	2 3 3	75.7917
3 1 1	70.4790	3 1 2	73.1353
3 1 3	73.1353	3 2 1	73.1353
3 2 2	75.7917	3 2 3	75.7917
3 3 1	73.1353	3 3 2	75.7917
3 3 3	75.7917		

Tabla 3.1: Valor óptimo para cada estado inicial

observarse que en este caso particular, es posible escribir la siguiente relación:

$$J_n^\tau(x_0) = \mathbf{n}J_1^\tau(x_0).$$

En un caso más general, considere la Tabla 3.4, la cual muestra los valores óptimos esperados para todos los posibles estados iniciales cuando se tiene una máquina. Entonces, computacionalmente puede verificarse que con un estado inicial, $x_0 = (d_1, d_2, \dots, d_n)$, para distintos valores de \mathbf{n} , $\mathbf{n} > 1$, se tiene que

$$J_n^\tau(x_0) = J_1^\tau(d_1) + J_1^\tau(d_2) + \dots + J_1^\tau(d_n),$$

lo cual permite dividir al problema con \mathbf{n} -máquinas en \mathbf{n} problemas de una sólo máquina (resultado que no se probó teóricamente, por estar fuera de los objetivos de la tesis), reduciendo el problema de la dimensionalidad. Cabe señalar, que el programa presenta dificultades relacionados con las dimensiones de los arreglos necesarios para el guardado de la información con $D = 3$ niveles de deterioro y un número de máquinas $\mathbf{n} > 5$.

Etapa Estado	1	2	3	4	5	6
1 1 1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
1 1 2	0 0 0	0 0 0	0 0 0	0 0 0	0 0 1	0 0 1
1 1 3	0 0 1	0 0 1	0 0 1	0 0 1	0 0 1	0 0 1
1 2 1	0 0 0	0 0 0	0 0 0	0 0 0	0 1 0	0 1 0
1 2 2	0 0 0	0 0 0	0 0 0	0 0 0	0 1 1	0 1 1
1 2 3	0 0 1	0 0 1	0 0 1	0 0 1	0 1 1	0 1 1
1 3 1	0 1 0	0 1 0	0 1 0	0 1 0	0 1 0	0 1 0
1 3 2	0 1 0	0 1 0	0 1 0	0 1 0	0 1 1	0 1 1
1 3 3	0 1 1	0 1 1	0 1 1	0 1 1	0 1 1	0 1 1
2 1 1	0 0 0	0 0 0	0 0 0	0 0 0	1 0 0	1 0 0
2 1 2	0 0 0	0 0 0	0 0 0	0 0 0	1 0 1	1 0 1
2 1 3	0 0 1	0 0 1	0 0 1	0 0 1	1 0 1	1 0 1
2 2 1	0 0 0	0 0 0	0 0 0	0 0 0	1 1 0	1 1 0
2 2 2	0 0 0	0 0 0	0 0 0	0 0 0	1 1 1	1 1 1
2 2 3	0 0 1	0 0 1	0 0 1	0 0 1	1 1 1	1 1 1
2 3 1	0 1 0	0 1 0	0 1 0	0 1 0	1 1 0	1 1 0
2 3 2	0 1 0	0 1 0	0 1 0	0 1 0	1 1 1	1 1 1
2 3 3	0 1 1	0 1 1	0 1 1	0 1 1	1 1 1	1 1 1
3 1 1	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0
3 1 2	1 0 0	1 0 0	1 0 0	1 0 0	1 0 1	1 0 1
3 1 3	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1
3 2 1	1 0 0	1 0 0	1 0 0	1 0 0	1 1 0	1 1 0
3 2 2	1 0 0	1 0 0	1 0 0	1 0 0	1 1 1	1 1 1
3 2 3	1 0 1	1 0 1	1 0 1	1 0 1	1 1 1	1 1 1
3 3 1	1 1 0	1 1 0	1 1 0	1 1 0	1 1 0	1 1 0
3 3 2	1 1 0	1 1 0	1 1 0	1 1 0	1 1 1	1 1 1
3 3 3	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1

Tabla 3.2: Políticas óptimas

\mathbf{n}	$J_{\mathbf{n}}^r(x_0)$
1	22.6075
2	45.2151
3	67.8226
4	90.4301
5	113.0377

Tabla 3.3: Valor óptimo con estado inicial $x_0 = (1, 1, 1)$ para diferentes valores de \mathbf{n}

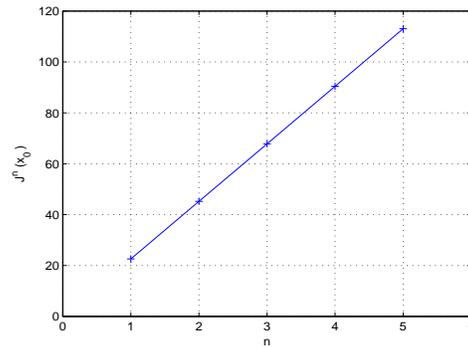


Figura 3.1: Valor óptimo con estado inicial $x_0 = (1, 1, 1)$ como una función de \mathbf{n}

Estado inicial x_0	$J_1^r(x_0)$
1	22.6075
2	25.2639
3	25.2639

Tabla 3.4: Función de valor óptimo para una máquina

Presentamos a continuación otro ejemplo considerando una máquina, con una distribución uniforme discreta sobre el conjunto $\{0, 1, 2, \dots, 49\}$ para el horizonte, 5 niveles de deterioro y la siguiente matriz de proba-

bilidades de transición:

$$P = \begin{pmatrix} 0.1 & 0.3 & 0.3 & 0.2 & 0.1 \\ 0 & 0.2 & 0.3 & 0.3 & 0.2 \\ 0 & 0 & 0.2 & 0.3 & 0.5 \\ 0 & 0 & 0 & 0.6 & 0.4 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Los resultados obtenidos son para cada una de las dos siguientes funciones de costo: $g_1 = (5, 6, 10, 25, 50)$ y $g_2 = (5, 6, 7, 35, 50)$.

Para la función de costo g_1 , la política óptima se resume a lo siguiente:

- No reemplazar la máquina en cualquier etapa estando en el nivel 1 o 2.
- Reemplazar la máquina en cualquier etapa estando en el nivel 3, 4 o 5.

Por lo tanto, podría decirse que el nivel óptimo de reemplazo es el 3.

Para la función de costo g_2 , se obtiene como política óptima lo siguiente:

- No reemplazar en cualquier etapa estando en el nivel 1 o 2.
- No reemplazar en la etapa 0 a la 45 y reemplazar en las restantes estando en el nivel 3.
- Reemplazar en cualquier etapa estando en el nivel 4 o 5.

Con esto, queremos hacer notar que no siempre puede determinarse un nivel óptimo de reemplazo. En general, tal y como se menciona en el Teorema 3.1, la política óptima es Markoviana determinista y sólo en casos particulares parece ser estacionaria.

Finalmente, señalamos dos aspectos que pueden modificarse en el modelo de reemplazo óptimo para generar otros posibles problemas.

Primeramente, podrían suponerse que los procesos de deterioro para las máquinas son únicamente independientes sin pedir idénticamente distribuidos, lo cual permitiría considerar sistemas formados por máquinas no necesariamente similares. Una segunda modificación es, considerar

costos no lineales de reemplazo de las máquinas, pensando por ejemplo, que podrían considerarse descuentos en el reemplazo de máquinas al mayoreo. Ambas modificaciones pueden realizarse haciendo ligeros cambios en el programa.

Capítulo 4

Procesos de Decisión de Markov con Horizonte Aleatorio de Soporte Infinito

En este capítulo se estudia el problema de control óptimo con horizonte aleatorio en el caso en que la distribución del horizonte tiene soporte infinito, esto es, cuando $T = +\infty$. Aquí se muestra una caracterización de la función de valor óptimo y se garantiza la existencia de la política óptima. También se presenta una aplicación de los resultados teóricos en una aproximación de la solución del problema de reemplazo óptimo de máquinas, presentado en el capítulo anterior. Además se proporcionan condiciones sobre la distribución del horizonte para poder acotar la función de valor óptimo del problema con costo total esperado y horizonte aleatorio mediante la función de valor óptimo de un problema con costo total esperado descontado y horizonte infinito.

4.1. Existencia de la Política Óptima y Caracterización de la Función de Valor Óptimo

En el caso en que $T = +\infty$, se tiene que

$$j^\tau(\pi, x) = E_x^\pi \left[\sum_{t=0}^{\infty} P_t c(x_t, a_t) \right], \quad (4.1)$$

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

Obsérvese que $P_t = \prod_{k=0}^t \alpha_{k-1}$, $t = 0, 1, 2, \dots$, donde se ha considerado que $\alpha_{-1} := P_0 = 1$ y α_k , $k = 0, 1, 2, \dots$, es definido en (3.4).

Entonces,

$$j^\tau(\pi, x) = E_x^\pi \left[\sum_{t=0}^{\infty} \prod_{k=0}^t \alpha_{k-1} c(x_t, a_t) \right], \quad (4.2)$$

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

Sea $v_n^\tau(\pi, x)$ el valor esperado del costo total de la etapa n en adelante, $n = 0, 1, 2, \dots$, aplicado a (4.2) con la condición inicial $x_n = x$, donde x es un elemento genérico de X . Esto es,

$$\begin{aligned} v_n^\tau(\pi, x) &= E_x^\pi \left[\sum_{t=n}^{\infty} \prod_{k=0}^t \alpha_{k-1} c(x_t, a_t) \right] \\ &= \prod_{k=0}^{n-1} \alpha_{k-1} E_x^\pi \left[\sum_{t=n}^{\infty} \prod_{k=n}^t \alpha_{k-1} c(x_t, a_t) \right], \end{aligned}$$

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

Obviamente se tiene que $v_0^\tau(\pi, x) = j^\tau(\pi, x)$, $x \in X$. Por lo tanto, el análisis de estas últimas funciones, proporcionará resultados para el criterio original, y para llevar a cabo el proceso de optimización, para cada $n = 0, 1, 2, \dots$, es suficiente considerar

$$v_n^\tau(\pi, x) := E_x^\pi \left[\sum_{t=n}^{\infty} \prod_{k=n}^t \alpha_{k-1} c(x_t, a_t) \right], \quad (4.3)$$

$\pi \in \Pi$, $x \in X$ y

$$V_n^\tau(x) := \inf_{\pi \in \Pi} v_n^\tau(\pi, x). \quad (4.4)$$

También, hacemos notar que la expresión (4.3) puede interpretarse como el valor esperado de un costo total descontado, en donde el factor de descuento es α_t , $t = 0, 1, 2, \dots$ (factor de descuento no homogéneo), y si se considera el factor homogéneo, es decir, que $\alpha_t = \alpha$, para toda $t = 0, 1, 2, \dots$, se obtiene el criterio descontado. En los resultados presentados, puede hacerse esta simplificación, quedando éstos de acuerdo con la teoría establecida para el problema descontado, bajo los supuestos considerados.

Para $N > n \geq 0$, defina

$$v_{n,N}^\tau(\pi, x) := E_x^\pi \left[\sum_{t=n}^N \prod_{k=n}^t \alpha_{k-1} c(x_k, a_k) \right], \quad (4.5)$$

con $\pi \in \Pi$, $x \in X$, y

$$V_{n,N}^\tau(x) := \inf_{\pi \in \Pi} v_{n,N}^\tau(\pi, x), \quad (4.6)$$

$x \in X$.

Considere la siguiente suposición:

Suposición 4.1 (a) *La función de costo por etapa c es semicontinua inferiormente, no negativa e inf-compacta en \mathbb{K} y la ley de transición Q es fuertemente continua o débilmente continua (igual que la Suposición 3.1).*

(b) *Existe una política $\pi \in \Pi$ tal que $j^\tau(\pi, x) < \infty$, para cada $x \in X$.*

El conjunto de todas las políticas $\pi \in \Pi$ que satisfacen la Suposición 4.1(b) se denota por Π^0 .

Observación 4.1

i) En la Suposición 4.1, se supone que la función de costo es no negativa. Esta suposición puede cambiarse, sin pérdida de generalidad, pidiendo c acotada por abajo. A saber, si $c \geq m$ para alguna constante m , entonces el problema con costo por etapa $c' := c - m$, el cual es no negativo, es equivalente a un problema con costo por etapa c .

ii) Observe que si $c \geq m$ y si la Suposición 4.1(b) se cumple, como

$$j^\tau(\pi, x) \geq m \sum_{t=0}^{\infty} P_t = m \sum_{t=0}^{\infty} (1+t)\rho_t = m(1 + E[\tau]),$$

entonces $E[\tau] < \infty$. Recíprocamente, en el caso en que c es acotada, si $E[\tau] < \infty$, entonces la Suposición 4.1(b) se cumple.

Los Lemas 4.1 y 4.2 mostrados a continuación son similares a los Lemas 4.2.4 y 4.2.6, en [15], respectivamente. Las pruebas correspondientes se presentan únicamente por completitud del trabajo de tesis.

Lema 4.1 *Para cada $N > n \geq 0$, sea w_n y $w_{n,N}$ funciones sobre \mathbb{K} , las cuales son no negativas, semicontinuas inferiormente e inf-compactas sobre \mathbb{K} . Si $w_{n,N} \uparrow w_n$ cuando $N \rightarrow \infty$, entonces*

$$\lim_{N \rightarrow \infty} \min_{a \in A(x)} w_{n,N}(x, a) = \min_{a \in A(x)} w_n(x, a),$$

$x \in X$.

Demostración. Como $w_{n,N} \uparrow w_n$ cuando $N \rightarrow \infty$, se tiene que

$$\lim_{N \rightarrow \infty} \min_{a \in A(x)} w_{n,N}(x, a) \leq \min_{a \in A(x)} w_n(x, a),$$

$x \in X$. Para probar la desigualdad contraria defina

$$\begin{aligned} l_n(x) &:= \lim_{N \rightarrow \infty} \min_{a \in A(x)} w_{n,N}(x, a) && \text{y} \\ w_n^*(x) &:= \min_{a \in A(x)} w_n(x, a), \end{aligned}$$

$x \in X$. Considere $x \in X$ arbitrario fijo y defina los siguientes conjuntos.

$$A_{n,N} := \{a \in A(x) : w_{n,N}(x, a) \leq l_n(x)\}, \quad N = n + 1, n + 2, n + 3, \dots,$$

$$A_n := \{a \in A(x) : w_n(x, a) = w_n^*(x)\}.$$

Por la hipótesis de inf-compacidad y como $w_{n,N} \uparrow w_n$, cada uno de estos conjuntos resulta ser no vacío y compacto, además se tiene que $A_{n,N} \downarrow A_n$. Para cada $N > n$, la Proposición D.6 en el Apéndice de [15], garantiza la existencia de $a_{n,N} \in A_{n,N}$ tal que

$$w_{n,N}(x, a_{n,N}) = \min_{a \in A(x)} w_{n,N}(x, a). \quad (4.7)$$

Por la compacidad y como $A_{n,N} \downarrow A_n$, existe $a_n \in A_n$ y una subsucesión $\{a_{n,N_i}\}$ de $\{a_{n,N}\}$ tal que $a_{n,N_i} \rightarrow a_n$. Luego, como $\{w_{n,N}\}$ es monótona creciente, para cualquier $N > n$, se tiene que

$$w_{n,N_i}(x, a_{n,N_i}) \geq w_{n,N}(x, a_{n,N_i}),$$

para todo $N_i \geq N$. Por (4.7), se obtiene que

$$\min_{a \in A(x)} w_{n,N_i}(x, a) \geq w_{n,N}(x, a_{n,N_i}).$$

Haciendo $i \rightarrow \infty$ y por la semicontinuidad inferior de $w_{n,N}$, resulta que

$$l_n(x) \geq w_{n,N}(x, a_n),$$

lo cual, por tener que $w_{n,N} \uparrow w_n$, implica que

$$l_n(x) \geq w_n(x, a_n) = w_n^*(x).$$

Como se consideró x arbitrario, la desigualdad es válida para todo $x \in X$, concluyendo así la demostración. ■

Definición 4.1 Sea $M(X)^+$ el cono de las funciones medibles no negativas sobre X . Para cada $u \in M(X)^+$, $T_n u$, $n = 0, 1, 2, \dots$, es el operador sobre X definido como

$$T_n u(x) = \min_{a \in A(x)} \left[c(x, a) + \alpha_n \int_X u(y) Q(dy \mid x, a) \right],$$

$x \in X$.

Lema 4.2 *Considere que la Suposición 4.1 (a) se cumple. Para cada $u \in M(X)^+$, $T_n u \in M(X)^+$. Además, existe $f_n \in \mathbb{F}$ tal que*

$$T_n u(x) = c(x, f_n(x)) + \alpha_n \int_X u(y) Q(dy | x, f_n(x)),$$

$x \in X$.

Demostración. Considere la Suposición 4.1 (a) con la ley de transición Q débilmente continua y sea $u \in M(X)^+$. Por ser Q débilmente continua $\int_X u(y) Q(dy | x, a)$ resulta ser medible no negativa. Como la suma y el mínimo de funciones medibles no negativas es medible no negativa se concluye que $T_n u \in M(X)^+$. Ahora se probará que la función

$$a \mapsto \int_X u(y) Q(dy | x, a) \tag{4.8}$$

es semicontinua inferiormente sobre $A(x)$ para todo $x \in X$. Para ello, sea u_k una sucesión de funciones medibles continuas acotadas sobre X tal que $u_k \uparrow u$, y sea a^l una sucesión en $A(x)$ que converge a $a \in A(x)$. Entonces, por ser Q débilmente continua

$$\begin{aligned} \liminf_{l \rightarrow \infty} \int_X u(y) Q(dy | x, a^l) &\geq \liminf_{l \rightarrow \infty} \int_X u_k(y) Q(dy | x, a^l) \\ &= \int_X u_k(y) Q(dy | x, a), \end{aligned}$$

para toda k . Haciendo $k \rightarrow \infty$, por el Teorema de la Convergencia Monótona, se obtiene que

$$\liminf_{l \rightarrow \infty} \int_X u(y) Q(dy | x, a^l) \geq \int_X u(y) Q(dy | x, a),$$

esto es (4.8) es semicontinua inferiormente. Como la suma de dos funciones semicontinuas inferiormente resulta ser también semicontinua inferiormente, se concluye que

$$u'(x, a) := c(x, a) + \alpha_n \int_X u(y) Q(dy | x, a)$$

es semicontinua inferiormente. Entonces, por la Proposición A.1(c) en el Apéndice de [15] el conjunto $D := \{a \in A(x) : u'(x, a) \leq r\}$ es

compacto para cada $x \in X$ y $r \in \mathbb{R}$, teniendo que u' es inf-compacta en \mathbb{K} . Finalmente, usando la Proposición D.6 dada en el Apéndice de [15] se obtiene el resultado. La prueba cuando Q es fuertemente continua es de manera similar, el caso se presenta en [15]. ■

Lema 4.3 *Considere que la Suposición 4.1(a) se cumple y sea $\{u_n\}$ una sucesión en $M(X)^+$.*

(a) *Si $u_n \geq T_n u_{n+1}$, $n = 0, 1, 2, \dots$, entonces $u_n \geq V_n^\tau$, $n = 0, 1, 2, \dots$*

(b) *Si $u_n \leq T_n u_{n+1}$ y*

$$\lim_{N \rightarrow \infty} E_x^\pi \left[\prod_{j=n}^{n+N} \alpha_{j-1} u_{n+N}(x_{n+N}) \right] = 0, \quad (4.9)$$

$n = 0, 1, 2, \dots$, $\pi \in \Pi^0$, entonces $\alpha_{n-1} u_n \leq V_n^\tau$, $n = 0, 1, 2, \dots$

Demostración.

(a) Sea $\{u_n\}$ una sucesión en $M(X)^+$ tal que $u_n \geq T_n u_{n+1}$, entonces, por el Lema 4.2

$$u_n(x) \geq c(x, f_n(x)) + \alpha_n \int_X u_{n+1}(y) Q(dy | x, f_n(x)),$$

$x \in X$. Iterando esta desigualdad, se obtiene que

$$\begin{aligned} u_n(x) &\geq E_x^\pi \left[c(x_n, f_n(x_n)) + \sum_{t=n+1}^{N-1} \prod_{j=n+1}^t \alpha_{j-1} c(x_t, f_t(x_t)) \right] \\ &\quad + \prod_{j=n+1}^N \alpha_{j-1} E_x^\pi [u(x_N)], \end{aligned} \quad (4.10)$$

$x \in X$, con

$$E_x^\pi [u(x_N)] = \int_X u(y) Q^N(dy | x_n, f_n(x_n)),$$

donde $Q^N(\cdot | x_n, f_n(x_n))$ denota el kernel de transición de N -pasos del proceso de Markov $\{x_t\}$ cuando se usa la política $\pi = \{f_k\}$,

iniciando en la etapa n . Como u es no negativa, $\alpha_k \leq 1$ y $x_n = x$, se obtiene de (4.10) que

$$u_n(x) \geq E_x^\pi \left[\alpha_{n-1} c(x_n, f_n(x_n)) + \sum_{t=n+1}^{N-1} \prod_{j=n}^t \alpha_{j-1} c(x_t, f_t(x_t)) \right].$$

Por lo tanto, haciendo $N \rightarrow \infty$, resulta que

$$u_n(x) \geq v_n^\tau(\pi, x) \geq V_n^\tau(x),$$

$x \in X$.

(b) Suponga que $u_n \leq T_n u_{n+1}$, $n = 0, 1, 2, \dots$. Entonces

$$u_n(x) \leq c(x, a_n) + \alpha_n \int_X u_{n+1}(y) Q(dy | x, a_n), \quad (4.11)$$

$\pi = \{a_k\} \in \Pi$, $x \in X$. Ahora, usando la propiedad Markoviana, se obtiene que

$$\begin{aligned} & E_x^\pi \left[\prod_{j=n}^{n+t+1} \alpha_{j-1} u_{n+t+1}(x_{n+t+1}) \mid h_{n+t}, a_{n+t} \right] \\ &= \prod_{j=n}^{n+t+1} \alpha_{j-1} \int_X u_{n+t+1}(y) Q(dy | x_{n+t}, a_{n+t}) \\ &= \prod_{j=n}^{n+t} \alpha_{j-1} \left(c(x_{n+t}, a_{n+t}) + \alpha_{n+t} \int_X u_{n+t+1}(y) Q(dy | x_{n+t}, a_{n+t}) \right) \\ &\quad - \prod_{j=n}^{n+t} \alpha_{j-1} c(x_{n+t}, a_{n+t}). \end{aligned}$$

Utilizando la desigualdad (4.11), se sigue que

$$\begin{aligned} & \prod_{j=n}^{n+t} \alpha_{j-1} \left(c(x_{n+t}, a_{n+t}) + \alpha_{n+t} \int_X u_{n+t+1}(y) Q(dy | x_{n+t}, a_{n+t}) \right) \\ & \geq \prod_{j=n}^{n+t} \alpha_{j-1} u_{n+t}(x_{n+t}). \end{aligned}$$

Equivalentemente,

$$\begin{aligned} & \prod_{j=n}^{n+t} \alpha_{j-1} c(x_{n+t}, a_{n+t}) \\ \geq & - \left[\prod_{j=n}^{n+t+1} \alpha_{j-1} E_x^\pi \left[u_{n+t+1}(x_{n+t+1}) \mid h_{n+t}, a_{n+t} \right] - \prod_{j=n}^{n+t} \alpha_{j-1} u_{n+t}(x_{n+t}) \right]. \end{aligned}$$

De esa manera, tomando esperanzas y sumando sobre $t = 0, 1, 2, \dots, N-1$, se obtiene que

$$\begin{aligned} \sum_{t=0}^{N-1} E_x^\pi \left[\prod_{j=n}^{n+t} \alpha_{j-1} c(x_{n+t}, a_{n+t}) \right] &= \sum_{t=n}^{N+n-1} E_x^\pi \left[\prod_{j=n}^t \alpha_{j-1} c(x_t, a_t) \right] \\ &\geq \alpha_{n-1} u_n(x) - \prod_{j=n}^{n+N} \alpha_{j-1} E_x^\pi [u_{n+N}(x_{n+N})]. \end{aligned}$$

Haciendo $N \rightarrow \infty$ y usando (4.9), se sigue el resultado.

■

Lema 4.4 *Considere que la Suposición 4.1 se cumple. Entonces, para cada $n \geq 0$ y $x \in X$,*

$$V_{n,N}^\tau(x) \uparrow V_n^\tau(x) \quad \text{cuando} \quad N \rightarrow \infty$$

y V_n^τ es semicontinua inferiormente.

Demostración. Usando la siguiente ecuación de programación dinámica

$$U_t(x) = \min_{a \in A(x)} \left[c(x, a) + \alpha_t \int_X U_{t+1}(y) Q(dy \mid x, a) \right], \quad (4.12)$$

para $t = N-1, N-2, \dots, n$, con $U_N(x) = 0$, $x \in X$, (véase (3.3)), por medio del Teorema 3.1 se obtiene que

$$V_{n,N}^\tau(x) = U_n(x),$$

$$V_{s,N}^\tau(x) = U_s(x)$$

para $n \leq s < N$. Además, se puede probar por inducción hacia atrás que U_s , $n \leq s < N$, es semicontinua inferiormente y por lo tanto $V_{n,N}^\tau(\cdot)$ también es semicontinua inferiormente.

Luego, obsérvese que para $t = n$, (4.12) se escribe como

$$V_{n,N}^\tau(x) = \min_{a \in A(x)} \left[c(x, a) + \alpha_n \int_X V_{n+1,N}^\tau(y) Q(dy | x, a) \right], \quad (4.13)$$

Además, por la no negatividad de c , para cada $n = 0, 1, 2, \dots$, la sucesión $\{V_{n,N}^\tau : N = n, n+1, \dots\}$ es no decreciente. Esto implica que existe una función $u_n \in M(X)^+$ tal que para cada $x \in X$,

$$V_{n,N}^\tau(x) \uparrow u_n(x), \quad \text{cuando } N \rightarrow \infty.$$

Ahora, se probará que $V_n^\tau(x) = u_n(x)$. Claramente se tiene que

$$V_{n,N}^\tau(x) \leq v_{n,N}^\tau(\pi, x) \leq v_n^\tau(\pi, x),$$

$x \in X$ y $\pi \in \Pi$. Por lo tanto

$$V_{n,N}^\tau(x) \leq V_n^\tau(x),$$

$N > n$, entonces

$$u_n \leq V_n^\tau.$$

Además, como u_n es el supremo de una sucesión de funciones semicontinuas inferiormente, también resulta ser semicontinua inferiormente. Usando el Lema 4.1 y haciendo $N \rightarrow \infty$ en (4.13), se obtiene que

$$u_n(x) = \min_{a \in A(x)} \left[c(x, a) + \alpha_n \int_X u_{n+1}(y) Q(dy | x, a) \right], \quad (4.14)$$

$n = 0, 1, 2, \dots$ y $x \in X$. Finalmente, por el Lema 4.3 (a) se tiene que

$$u_n \geq V_n^\tau,$$

obteniendo que

$$u_n = V_n^\tau,$$

concluyendo de esta manera la prueba del Lema 4.4. ■

Teorema 4.1 *Considere que la Suposición 4.1 se cumple, entonces*

(a) la función de valor óptimo V_n^τ , $n = 0, 1, 2, \dots$, satisface la ecuación de optimalidad

$$V_n^\tau(x) = \min_{a \in A(x)} \left[c(x, a) + \alpha_n \int_X V_{n+1}^\tau(y) Q(dy | x, a) \right], \quad (4.15)$$

$x \in X$, y si $\{u_n\}$ es cualquier otra sucesión que satisface la ecuación (4.15), entonces $u_n \geq V_n^\tau$.

(b) Existe una política $\pi^* = \{f_n \in \mathbb{F} \mid n \geq 0\}$ tal que, para cada $n = 0, 1, 2, \dots$, el control $f_n(x) \in A(x)$ alcanza el mínimo en (4.15), i.e.

$$V_n^\tau(x) = c(x, f_n(x)) + \alpha_n \int_X V_{n+1}^\tau(y) Q(dy | x, f_n(x)), \quad (4.16)$$

$x \in X$, y la política π^* es óptima.

Demostración.

(a) La prueba del Lema 4.4 garantiza que la sucesión $\{V_n^\tau\}$ satisface la ecuación de optimalidad en (4.15), y por el Lema 4.3 (a), si $\{u_n\}$ satisface

$$u_n(x) = \min_{a \in A(x)} \left[c(x, a) + \alpha_n \int_X u_{n+1}(y) Q(dy | x, a) \right],$$

se concluye que $u_n \geq V_n^\tau$.

(b) la existencia de $f_n \in \mathbb{F}$ que satisface (4.16) es asegurada por el Lema 4.2. Ahora, iterando (4.16) con $x_n = x \in X$, se obtiene que

$$\begin{aligned} V_n^\tau(x) &= E_x^\pi \left[c(x_n, f_n(x_n)) + \sum_{t=n+1}^{N-1} \prod_{j=n+1}^t \alpha_{j-1} c(x_t, f_t(x_t)) \right] \\ &\quad + \prod_{j=n+1}^N \alpha_{j-1} E_x^\pi [u(x_N)] \\ &\geq E_x^\pi \left[\sum_{t=n}^{N-1} \prod_{j=n}^t \alpha_{j-1} c(x_t, f_t(x_t)) \right], \end{aligned}$$

$n \geq 0$ y $N > n$. Entonces, haciendo $N \rightarrow \infty$ se obtiene que

$$V_n^\tau(x) \geq v_n^\tau(\pi^*, x),$$

$x \in X$ y $\pi^* = \{f_k\}$. Además, en particular para π^* ,

$$V_n^\tau(x) \leq v_n^\tau(\pi^*, x),$$

$x \in X$. Por lo tanto, π^* es óptima.

De esa manera se ha probado el Teorema 4.1. ■

A continuación se presentan dos modelos para los cuales se cumplen las conclusiones del Lema 4.4 y el Teorema 4.1.

- 1) **Modelo LQ con horizonte aleatorio.** Considere el modelo LQ presentado en el Capítulo 3, ahora suponiendo que la distribución del horizonte τ tiene soporte infinito con $E[\tau] < \infty$.

Lema 4.5 *El Modelo LQ con horizonte aleatorio satisface la Suposición 4.1.*

Demostración. La Suposición 4.1(a) se ha probado en el Capítulo 3. Ahora, para probar la Suposición 4.1(b), considere la política estacionaria $h(x) = -\frac{\gamma}{\beta}x$ y $x_0 = x \in X$. En este caso, al aplicar la dinámica del sistema, esto es $x_{t+1} = \gamma x_t + \beta a_t + \xi_t$, tenemos que

$$\begin{aligned} x_1 &= \gamma x_0 + \beta a_0 + \xi_0 = \gamma x - \beta \frac{\gamma}{\beta} x + \xi_0 = \xi_0 \\ x_2 &= \gamma x_1 + \beta a_1 + \xi_1 = \gamma \xi_0 - \beta \frac{\gamma}{\beta} \xi_0 + \xi_1 = \xi_1 \\ x_3 &= \gamma x_2 + \beta a_2 + \xi_2 = \gamma \xi_1 - \beta \frac{\gamma}{\beta} \xi_1 + \xi_2 = \xi_2 \\ &\vdots \\ x_t &= \gamma x_{t-1} + \beta a_{t-1} + \xi_{t-1} = \gamma \xi_{t-2} - \beta \frac{\gamma}{\beta} \xi_{t-2} + \xi_{t-1} = \xi_{t-1}. \end{aligned}$$

Entonces

$$\begin{aligned}
j^\tau(h, x) &= E_x^h \left[\sum_{t=0}^{\infty} P_t c(x_t, a_t) \right] \\
&= E_x^h \left[\sum_{t=0}^{\infty} P_t \left(q + r \frac{\gamma^2}{\beta^2} \right) \xi_{t-1}^2 \right] \\
&= \left(q + r \frac{\gamma^2}{\beta^2} \right) \sigma^2 \sum_{t=0}^{\infty} P_t \\
&= \left(q + r \frac{\gamma^2}{\beta^2} \right) \sigma^2 (E[\tau] + 1) < \infty.
\end{aligned}$$

■

- 2) **Modelo de Consumo-Inversión con función de utilidad de tipo Logaritmo.** Considere un inversionista, quien desea dividir su riqueza actual x_t entre una inversión a_t y un consumo $x_t - a_t$ para cada periodo $t = 0, 1, 2, \dots, \tau$, donde τ es una variable aleatoria con un soporte infinito. Se supone $A = [0, 1]$ y que el conjunto de inversiones admisibles está dado por

$$A(x) = \begin{cases} [0, x) & \text{si } x \in (0, 1] \\ 0 & \text{si } x = 0. \end{cases}$$

En este caso el espacio de estados es $X = [0, 1]$. La función de utilidad es definida por

$$u(x - a) = \begin{cases} \ln(x - a) & \text{si } x \in (0, 1] \\ 0 & \text{si } x = 0. \end{cases}$$

La relación entre la inversión y el capital acumulado es dada por

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

$t = 0, 1, 2, \dots, \tau$ con $x_0 = x \in X$, donde $\{\xi_t\}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas que toman valores en $(0, 1)$, con función de densidad Δ continua, tal que $E[|\ln \xi_0|] = K < \infty$, donde ξ_0 es un elemento genérico de la

sucesión $\{\xi_t\}$.

En el caso de maximización, se obtienen resultados equivalentes a los ya mostrados con cambios adecuados en las suposiciones. Por ejemplo, la Suposición 4.1 debe cambiarse por la siguiente:

Suposición 4.2 (a) *La recompensa por etapa g es semicontinua superiormente, no positiva y sup-compacta sobre \mathbb{K} .*

(b) *Q es fuertemente o débilmente continuo.*

(c) *Existe una política $\pi \in \Pi$ tal que $j^\tau(\pi, x) > -\infty$ para cada $x \in X$.*

Lema 4.6 *El Modelo de Consumo-Inversión con función de utilidad de tipo Logaritmo satisface la Suposición 4.2.*

Demostración. Sea $A_\lambda(u) = \{a \in A(x) : u(x - a) \geq \lambda\}$, $\lambda \in \mathbb{R}$. Observe que, para $x = 0$

$$A_\lambda(u) = \begin{cases} \{0\} & \text{si } \lambda \leq 0 \\ \emptyset & \text{si } \lambda > 0 \end{cases}$$

y, para $x \in (0, 1]$

$$A_\lambda(u) = \begin{cases} [0, x - \exp(\lambda)] & \text{si } \lambda \leq \ln x \\ \emptyset & \text{si } \lambda > \ln x. \end{cases}$$

Como $A_\lambda(u)$ es cerrado y compacto para cada $x \in X$ se tiene que u es sup-compacta y usando la Proposición A.1 en el Apéndice de [15], también es semicontinua superiormente. Así, la Suposición 4.2(a) se cumple.

Ahora, sea $v : X \rightarrow \mathbb{R}$ una función continua y acotada y defina

$$\begin{aligned} v'(x, a) &= \int_X v(y)Q(dy | x, a) \\ &= \int_0^1 v(as)\Delta(s)ds, \end{aligned}$$

$x \in X$ y $a \in A(x)$. Realizando el cambio de variable $u = as$, se obtiene que

$$v'(x, a) = \int_{-\infty}^{\infty} v(u) I_{[0,a]}(u) \Delta\left(\frac{u}{a}\right) \frac{du}{a},$$

si $a \neq 0$ y $v'(x, 0) = v(0)$. Como Δ es continua, v' es continua para $a \neq 0$. Sea $\{a_n\}$ una sucesión arbitraria tal que $a_n \rightarrow 0$, así $\lim_{n \rightarrow \infty} v'(x, a_n) = \lim_{n \rightarrow \infty} \int_0^1 v(a_n s) \Delta(s) ds = v(0)$, entonces v' es continua para $a = 0$. De esa manera, se verifica que v' es continua y acotada sobre \mathbb{K} para cada función continua y acotada v sobre X , i.e. Q es débilmente continua. Por otro lado, usando $h(x) = 0$, $x \in X$, se obtiene que $j^\tau(h, x) = P_0 \ln x > -\infty$, $x \in X$. ■

Considerando al problema de horizonte aleatorio como un problema no homogéneo en el costo, puede utilizarse como una alternativa de análisis la transformación de un modelo no homogéneo a un modelo homogéneo (véase [14] p. 13). En [13] se hizo el estudio del problema no homogéneo (sin considerar no homogeneidad en el factor de descuento) mediante ambas técnicas: la utilizada en este trabajo de tesis y la anteriormente mencionada. Entonces, se hace la observación de que efectivamente se puede obtener nuevamente el mismo teorema que caracteriza a la solución óptima pero de una forma aislada, esto es, sin resultados previos. La técnica elegida para este trabajo de tesis proporciona incluso metodologías para determinar la solución óptima, tal y como lo muestra el Lema 4.4.

4.1.1. El Procedimiento de Horizonte Rodante

El Lema 4.4 puede ser aplicado para obtener $J^\tau = V_0^\tau$, como el límite de la sucesión $\{V_{0,N}^\tau\}$. Pero, en la práctica hallar este límite es bastante complicado debido a la no homogeneidad del problema. El Lema 4.4 y el Teorema 4.1 deberán usarse conjuntamente para determinar la solución óptima, no obstante, quedan como dos importantes resultados de existencia de la solución óptima. Sin embargo, obsérvese que el Lema 4.4 puede ser aplicado para aproximar J^τ en casos prácticos. Para aproximar π^* , por ejemplo, puede utilizarse el procedimiento de horizonte rodante, el cual es presentado a continuación con la finalidad de utilizarlo en un

ejemplo posterior.

El Procedimiento de Horizonte Rodante

El procedimiento de horizonte rodante es el método más común aplicado en la práctica para aproximar soluciones a problemas de control óptimos no homogéneos con horizonte infinito (véase [32]). Primero se fija un horizonte N , luego se procede a resolver el problema correspondiente de N periodos y se implementa la primera decisión óptima encontrada, se rueda hacia adelante un periodo y se repite el procedimiento a partir del estado actual y así sucesivamente. Enseguida, se presenta el algoritmo de horizonte rodante.

Algoritmo

1. Hacer $m = 0$ y $n = N$.
2. Hallar la política $\pi^* = (\pi_m^*, \pi_{m+1}^*, \dots, \pi_{n-1}^*)$, la cual es óptima para los periodos de m a n , y sea $\hat{\pi}_m = \pi_m^*$.
3. Hacer $m = m + 1$ y $n = n + 1$.
4. Regresar al Paso 2.

La política $\hat{\pi} = (\hat{\pi}_0, \hat{\pi}_1, \hat{\pi}_2, \dots)$ es una *política de horizonte rodante*.

La política de horizonte rodante en algunos casos puede ser no óptima, un ejemplo se presenta en [1]. También en [1] se provee un error teórico para el procedimiento de horizonte rodante aplicado a procesos de decisión de Markov no homogéneos con horizonte infinito.

4.1.2. Una Aplicación en la Aproximación de la Solución del Problema de Reemplazo Óptimo

Considere el modelo de reemplazo óptimo de \mathbf{n} -máquinas con horizonte aleatorio presentado en el Capítulo 3, pero en este caso, suponga que la distribución del horizonte tiene soporte infinito.

Lema 4.7 *El modelo de reemplazo óptimo de \mathbf{n} -máquinas con horizonte aleatorio satisface la Suposición 4.1.*

Demostación. Como X y A son finitos, entonces la Suposición 4.1(a) se cumple. Además, existe M tal que $c(x, a) \leq M$ para cada $x \in X$ y $a \in A(x)$. Por lo tanto $j^\tau(\pi, x) \leq M \sum_{t=0}^{\infty} P_t = M(1 + E[\tau])$. Como se supone que $E[\tau] < \infty$, entonces $j^\tau(\pi, x) < \infty$ para cada $\pi \in \Pi$ y $x \in X$, es decir, la Suposición 4.1(b) se cumple. ■

Implementación del algoritmo de horizonte rodante en el problema de reemplazo óptimo de máquinas

Se presenta únicamente la parte de la obtención de la política de horizonte rodante, el programa completo se muestra en el Apéndice como el Programa A.2.

1. Hacer $n = 0$.
2. Hacer $t = N + 1$ y $U_t(x) = 0$ para cada $x \in X$.
3. Si $t = 0$, hacer $\hat{f}_n = f_0$, $n = n + 1$ y regresar al Paso 2. En otro caso, ir al Paso 4.
4. Reemplazar t por $t - 1$ y calcular $U_t(x)$ para cada $x \in X$ por medio de la ecuación

$$U_t(x) = \min_{a \in A(x)} \left[c(x, a) + \alpha_{t+n} \sum_{y \in X} U_{t+1}(y) q_{x,y}^a \right], \quad ,$$

(véase (3.3)), haciendo

$$f_t(x) = a,$$

para algún

$$a \in \arg \min_{a \in A(x)} \left[c(x, a) + \alpha_{t+n} \sum_{y \in X} U_{t+1}(y) q_{x,y}^a \right].$$

Regresar al Paso 3.

$\hat{f} = (\hat{f}_0, \hat{f}_1, \hat{f}_2, \dots)$ es la política de horizonte rodante obtenida correspondiente a un N fijo, donde N es un número entero.

Ejemplo numérico

Considere el problema de reemplazo óptimo con horizonte aleatorio y los siguientes valores numéricos: $\mathbf{n} = 3$, $D = 3$,

$$P = \begin{pmatrix} 0.4 & 0.3 & 0.3 \\ 0 & 0.3 & 0.7 \\ 0 & 0 & 1 \end{pmatrix},$$

$g = (5, 7, 29)$, $R = 4$, $\rho_k = P(\tau = k) = -\frac{(1-p)^{k+1}}{(k+1)\ln p}$, $k = 0, 1, 2, \dots$, donde $0 < p < 1$ con $p = 0.8$ (Distribución de tipo Logaritmo) y $N = 20$.

En la Tabla 4.1, se presenta la acción de horizonte rodante dependiendo de la etapa y el estado del sistema.

En este caso numérico, se observa que la política obtenida es estacionaria (independiente del tiempo) y que es posible determinar un nivel de reemplazo óptimo para cada máquina. En este caso, debe reemplazarse una máquina si se encuentra en el nivel 3.

La función de valor óptimo se obtiene evaluando la política de horizonte rodante en el criterio de rendimiento.

Luego, el Teorema 4.4 garantiza que

$$V_{0,N}^\tau(x) \uparrow J^\tau(x) \quad \text{cuando} \quad N \rightarrow \infty.$$

Entonces, para N^* suficientemente grande $V_{0,N^*}^\tau(x)$ aproxima a $J^\tau(x)$, $x \in X$. Dicho valor puede ser usado como referencia para decidir si una política arbitraria puede ser utilizada para aproximar la solución óptima, considerando una tolerancia de error. En particular, para una política de horizonte rodante, se podría obtener $j^\tau(\hat{\pi}, x)$ y comparar con $V_{0,N^*}^\tau(x)$, $x \in X$, pidiendo que la máxima diferencia no sobrepase el error permitido.

En la Tabla 4.2, la columna número uno muestra la evaluación sólo de las primeras cinco acciones debido a que se hace presente en el programa la maldición de la dimensionalidad y no permite seguir evaluando (véase [26]), y en la segunda columna se presenta una aproximación de la función de valor óptimo usando el Lema 4.4 con $N^* = 30$. N^* se determinó a

Etapa Estado	1	2	3	4	5	6	7	...
1 1 1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	
1 1 2	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	
1 1 3	0 0 1	0 0 1	0 0 1	0 0 1	0 0 1	0 0 1	0 0 1	
1 2 1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	
1 2 2	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	
1 2 3	0 0 1	0 0 1	0 0 1	0 0 1	0 0 1	0 0 1	0 0 1	
1 3 1	0 1 0	0 1 0	0 1 0	0 1 0	0 1 0	0 1 0	0 1 0	
1 3 2	0 1 0	0 1 0	0 1 0	0 1 0	0 1 0	0 1 0	0 1 0	
1 3 3	0 1 1	0 1 1	0 1 1	0 1 1	0 1 1	0 1 1	0 1 1	
2 1 1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	
2 1 2	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	
2 1 3	0 0 1	0 0 1	0 0 1	0 0 1	0 0 1	0 0 1	0 0 1	
2 2 1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	
2 2 2	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	
2 2 3	0 0 1	0 0 1	0 0 1	0 0 1	0 0 1	0 0 1	0 0 1	
2 3 1	0 1 0	0 1 0	0 1 0	0 1 0	0 1 0	0 1 0	0 1 0	
2 3 2	0 1 0	0 1 0	0 1 0	0 1 0	0 1 0	0 1 0	0 1 0	
2 3 3	0 1 1	0 1 1	0 1 1	0 1 1	0 1 1	0 1 1	0 1 1	
3 1 1	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	
3 1 2	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	
3 1 3	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	
3 2 1	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	
3 2 2	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	
3 2 3	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	
3 3 1	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	
3 3 2	1 1 0	1 1 0	1 1 0	1 1 0	1 1 0	1 1 0	1 0 0	
3 3 3	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	

Tabla 4.1: Política de horizonte rodante

partir de una diferencia no mayor a un $\epsilon = 1 \times 10^{-19}$ entre los valores óptimos sucesivos $V_{0,N^*}^\tau(x)$ y $V_{0,N^*+1}^\tau(x)$.

Es pertinente señalar que el programa fue elaborado en Maple, por la bondad que presenta en la manipulación de valores de gran precisión,

Estado inicial x	$j^\tau(\hat{\pi}, x)$	$V_{0,N^*}^\tau(x)$
1 1 1	17.44943	17.45081
1 1 2	19.60531	19.60669
1 1 3	21.26398	21.26537
1 2 1	19.60531	19.60669
1 2 2	21.76119	21.76257
1 2 3	23.41986	23.42125
1 3 1	21.26398	21.26537
1 3 2	23.41986	23.42125
1 3 3	25.07854	25.07992
2 1 1	19.60531	19.60669
2 1 2	21.76119	21.76257
2 1 3	23.41986	23.42125
2 2 1	21.76119	21.76257
2 2 2	23.91707	23.91845
2 2 3	25.57574	25.57713
2 3 1	23.41986	23.42125
2 3 2	25.57574	25.57713
2 3 3	27.23442	27.23580
3 1 1	21.26398	21.26537
3 1 2	23.41986	23.42125
3 1 3	25.07854	25.07992
3 2 1	23.41986	23.42125
3 2 2	25.57574	25.57713
3 2 3	27.23442	27.23580
3 3 1	25.07854	25.07992
3 3 2	27.23442	27.23580
3 3 3	28.89310	28.89448

Tabla 4.2: Aproximación de la función de valor óptimo

ya que para el cálculo de los α_t , $t = 0, 1, 2, \dots$ fue necesario trabajar hasta con 400 dígitos. Aun así, el programa obtiene la política de horizonte rodante en tiempo aceptable (aproximadamente 35 segundos para el ejemplo mostrado). El principal problema que presenta, como ya se mencionó, es en la evaluación de la política de horizonte rodante, en don-

de, se requiere de arreglos de tamaño $D^{(n(k-1))}$ para la k -ésima etapa para el guardado de la información.

4.2. El Problema de Control Óptimo con Horizonte Aleatorio como un Problema Descontado

En esta sección, se proporciona una condición que permite acotar al problema con horizonte aleatorio τ por un problema descontado apropiado.

Recuerde que α_t , $t = 0, 1, 2 \dots$, se define como

$$\alpha_t = \frac{P_{t+1}}{P_t}, \quad (4.17)$$

donde $P_t = P(\tau \geq t)$, y considere la siguiente suposición.

Suposición 4.3 $\{\alpha_t\}_{t=0}^{\infty} \subset (0, 1)$ es una sucesión tal que $\lim_{t \rightarrow \infty} \alpha_t = \bar{\alpha}$ y $\alpha_t \leq \bar{\alpha}$.

Observación 4.2 En el caso de maximización, la suposición correspondiente a la Suposición 4.3 es la siguiente: $\{\alpha_t\}_{t=0}^{\infty} \subset (0, 1)$ es una sucesión tal que $\lim_{t \rightarrow \infty} \alpha_t = \bar{\alpha}$ y $\alpha_t \geq \bar{\alpha}$.

A continuación se presentan dos ejemplos de distribuciones de probabilidad para τ que satisfacen la Suposición 4.3.

1. Considere τ con distribución de tipo logaritmo, i.e. $\rho_k = -\frac{(1-p)^{k+1}}{(k+1)\ln p}$,

$k = 0, 1, 2, \dots$, donde $0 < p < 1$. Observe que

$$\begin{aligned}
 \alpha_t &= \frac{\sum_{k=t+1}^{\infty} \rho_k}{\sum_{k=t}^{\infty} \rho_k} \\
 &= \frac{\sum_{k=t+1}^{\infty} \rho_k + \rho_t - \rho_t}{\sum_{k=t}^{\infty} \rho_k} \\
 &= 1 - \frac{\rho_t}{\sum_{k=t}^{\infty} \rho_k} \\
 &= 1 - \frac{(1-p)^{t+1}}{(t+1) \sum_{j=0}^{\infty} \frac{(1-p)^{t+j+1}}{t+j+1}} \\
 &= 1 - \frac{1}{\sum_{j=0}^{\infty} \frac{(1-p)^j}{1+\frac{j}{t+1}}}.
 \end{aligned}$$

Por lo tanto

$$\bar{\alpha} = \lim_{t \rightarrow \infty} \alpha_t = 1 - \frac{1}{\sum_{j=0}^{\infty} (1-p)^j} = 1 - p.$$

Ahora, se verificará que $\alpha_{t-1} \leq \alpha_t$ para cada $t = 1, 2, 3, \dots$. Esto es, se probará que

$$1 - \frac{1}{\sum_{j=0}^{\infty} \frac{(1-p)^j}{1+\frac{j}{t}}} \leq 1 - \frac{1}{\sum_{j=0}^{\infty} \frac{(1-p)^j}{1+\frac{j}{t+1}}}$$

o, equivalentemente que

$$\sum_{j=0}^{\infty} \frac{(1-p)^j}{1+\frac{j}{t}} \leq \sum_{j=0}^{\infty} \frac{(1-p)^j}{1+\frac{j}{t+1}}. \quad (4.18)$$

Como para cada $j = 0, 1, 2, \dots$ se satisface que

$$\frac{1}{1+\frac{j}{t}} \leq \frac{1}{1+\frac{j}{t+1}}$$

entonces se cumple (4.18), quedando completamente probada la Suposición 4.3.

2. Sea τ con distribución binomial negativa, cuyos parámetros son q y r , donde $0 \leq q \leq 1$ y $r \in \mathbb{N}$. En este caso,

$$\begin{aligned}\alpha_t &= 1 - \frac{\rho_t}{\sum_{k=t}^{\infty} \rho_k} \\ &= 1 - \frac{\frac{(r+t-1)!}{t!}}{\sum_{j=0}^{\infty} \frac{(r+t+j-1)!q^j}{(t+j)!}} \\ &= 1 - \frac{\left(1 + \frac{1}{t}\right)\left(1 + \frac{2}{t}\right) \cdots \left(1 + \frac{r-1}{t}\right)}{\sum_{j=0}^{\infty} \left(1 + \frac{j+1}{t}\right)\left(1 + \frac{j+2}{t}\right) \cdots \left(1 + \frac{j+r-1}{t}\right)q^j},\end{aligned}$$

de donde

$$\bar{\alpha} = \lim_{t \rightarrow \infty} \alpha_t = 1 - \frac{1}{\sum_{j=0}^{\infty} q^j} = q.$$

Ahora, se probará que $\alpha_t \geq \alpha_{t+1}$. Esto es que

$$1 - \frac{\frac{(r+t-1)!}{t!}}{\sum_{j=0}^{\infty} \frac{(r+t+j-1)!q^j}{(t+j)!}} \geq 1 - \frac{\frac{(r+t)!}{(t+1)!}}{\sum_{j=0}^{\infty} \frac{(r+t+j)!q^j}{(t+j+1)!}}$$

o equivalentemente que

$$(t+1) \sum_{j=0}^{\infty} \frac{(r+t+j)!q^j}{(t+j+1)!} \leq (r+t) \sum_{j=0}^{\infty} \frac{(r+t+j-1)!q^j}{(t+j)!}. \quad (4.19)$$

Observe que para cada j

$$(t+1) \frac{(r+t+j)!}{(t+j+1)!} \leq (r+t) \frac{(r+t+j-1)!}{(t+j)!},$$

entonces se cumple (4.19). Por lo tanto $\alpha_t \downarrow \bar{\alpha}$, entonces se satisface la suposición dada en la Observación 4.2.

Ahora, considere el modelo de control de Markov $(X, A, \{A(x) : x \in X\}, Q, c)$ y el criterio de rendimiento como

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

$\pi \in \Pi$, $x \in X$ y $\bar{\alpha} = \lim_{t \rightarrow \infty} \alpha_t$ (véase la Suposición 4.3). Sea

$$V(x) := \inf_{\pi \in \Pi} v(\pi, x), \quad (4.20)$$

$x \in X$.

El siguiente lema, permite acotar la función de valor óptimo de un problema con horizonte aleatorio mediante la función de valor óptimo de un problema descontado.

Lema 4.8 *Considere que las Suposiciones 1.1(c), 4.1(a) y 4.3 se cumplen. Entonces $J^\tau(x) \leq V(x)$, $x \in X$.*

Demostración. Por el Teorema 4.2.3 en [15], existe $f \in \mathbb{F}$ tal que

$$V(x) = c(x, f(x)) + \bar{\alpha} \int_X V(y) Q(dy | x, f(x)), \quad (4.21)$$

$x \in X$. Iterando (4.21), se obtiene que

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

$n \geq 1$, $x \in X$, donde

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

$Q^n(\cdot | x, f)$ denota el kernel de transición del proceso de Markov $\{x_t\}$ de n -pasos cuando $f \in \mathbb{F}$ es usada. Por la Suposición 4.3, resulta que

$$\sum_{t=0}^{n-1} \bar{\alpha}^t c(x_t, f(x_t)) \geq \sum_{t=0}^{n-1} \prod_{k=0}^t \alpha_{k-1} c(x_t, f(x_t)) = \sum_{t=0}^{n-1} P_t c(x_t, f(x_t)),$$

por lo tanto

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

$n \geq 1, x \in X$. Como V es no negativa,

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

$n \geq 1, x \in X$. Haciendo $n \rightarrow \infty$, resulta que

$$V(x) \geq j^\tau(f, x) \geq J^\tau(x),$$

$x \in X$. ■

La siguiente parte se desarrolla con el objetivo de hallar la diferencia entre las funciones que se han relacionado en el Lema 4.8. El resultado se muestra en el Teorema 4.2, al final de esta sección.

Sea $f^* \in \mathbb{F}$ la política óptima estacionaria del problema descontado y considere la siguiente suposición:

Suposición 4.4 *Existen números positivos m y k , con $1 \leq k < \frac{1}{\alpha}$, y una función $w \in M(X)^+$ tal que para todo $(x, a) \in \mathbb{K}$,*

(a) $c(x, a) \leq mw(x)$, y

(b) $\int_X w(y)Q(dy | x, a) \leq kw(x)$.

Observación 4.3

i) La Suposición 4.4 implica la Suposición 4.1(b) (véase la Proposición 4.3.1 en [15]).

ii) Observe que

$$E_x^\pi [w(x_t)] \leq k^t w(x), \quad (4.22)$$

$t = 0, 1, 2, \dots$ y $x \in X$.

Para $t = 0$, (4.22) trivialmente se cumple, y para $t \geq 1$, usando la Suposición 4.4(b), se obtiene que

$$E_x^\pi [w(x_t) | h_{t-1}, a_{t-1}] = \int_X w(y)Q(dy | x_{t-1}, a_{t-1}) \leq kw(x_{t-1}).$$

Tomando esperanza de nueva cuenta, resulta que

$$E_x^\pi [w(x_t)] \leq k E_x^\pi [w(x_{t-1})]. \quad (4.23)$$

Iterando (4.23), se obtiene (4.22).

Lema 4.9 *Bajo la Suposición 4.1(a), 4.3 y 4.4,*

$$0 \leq V(x) - j^\tau(f^*, x) \leq mw(x) \sum_{t=0}^{\infty} (\bar{\alpha}^t - P_t) k^t.$$

Demostración. Primeramente, observe que

$$\sum_{t=0}^{\infty} (\bar{\alpha}^t - P_t) k^t \leq \sum_{t=0}^{\infty} (\bar{\alpha}k)^t < \infty,$$

como $0 < \bar{\alpha}k < 1$ por la Suposición 4.4. Entonces, bajo la Suposición 4.4(a) y la desigualdad (4.22), para una política estacionaria $f \in \mathbb{F}$, se obtiene que

$$\begin{aligned} v(f, x) - j^\tau(f, x) &= \sum_{t=0}^{\infty} (\bar{\alpha}^t - P_t) E_x^f [c(x_t, f(x_t))] \\ &\leq m \sum_{t=0}^{\infty} (\bar{\alpha}^t - P_t) E_x^f [w(x_t)] \\ &\leq mw(x) \sum_{t=0}^{\infty} (\bar{\alpha}^t - P_t) k^t. \end{aligned}$$

Tomando $f = f^*$, donde f^* es la política óptima estacionaria del problema descontado, se concluye la demostración. ■

Sean $D_n : \mathbb{K} \rightarrow \mathbb{R}$ las funciones de discrepancia definidas como

$$D_n(x, a) := c(x, a) + \alpha_n \int_X V_{n+1}^\tau(y) Q(dy | x, a) - V_n^\tau(x),$$

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

Teorema 4.2 *Bajo la Suposición 4.1(a), 4.3, 4.4 y si $P(\tau < +\infty) = 1$,*

$$V(x) - J^\tau(x) \leq mw(x) \sum_{t=0}^{\infty} (\bar{\alpha}^t - P_t) k^t + \sum_{t=0}^{\infty} \prod_{k=0}^t \alpha_{k-1} E_x^\pi [D_t(x_t, f^*)], \quad (4.24)$$

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

Demostración. Para $x \in X$, por el Lema 4.9,

$$\begin{aligned} V(x) - J^\tau(x) &= V(x) - j^\tau(f^*, x) + j^\tau(f^*, x) - J^\tau(x) \\ &\leq mw(x) \sum_{t=0}^{\infty} (\bar{\alpha}^t - P_t) k^t + j^\tau(f^*, x) - J^\tau(x). \end{aligned}$$

Ahora, para $\pi \in \Pi$ y $x \in X$,

$$\begin{aligned} v_n^\tau(\pi, x) &= E_x^\pi \left[\sum_{t=n}^{\infty} \prod_{k=n}^t \alpha_{k-1} c(x_t, a_t) \right] \\ &= E_x^\pi \left[\sum_{t=n}^{\infty} \prod_{k=n}^t \alpha_{k-1} D_t(x_t, a_t) \right] \\ &\quad - E_x^\pi \left[\sum_{t=n}^{\infty} \prod_{k=n}^{t+1} \alpha_{k-1} \left(\int_X V_{t+1}^\tau(y) Q(dy \mid x_t, a_t) - V_t^\tau(x_t) \right) \right] \\ &= E_x^\pi \left[\sum_{t=n}^{\infty} \prod_{k=n}^t \alpha_{k-1} D_t(x_t, a_t) \right] \\ &\quad - E_x^\pi \left[\sum_{t=n}^{\infty} \left(\prod_{k=n}^{t+1} \alpha_{k-1} E_x^\pi [V_{t+1}^\tau(x_{t+1}) \mid h_t, a_t] - \prod_{k=n}^t \alpha_{k-1} V_t^\tau(x_t) \right) \right] \\ &= E_x^\pi \left[\sum_{t=n}^{\infty} \prod_{k=n}^t \alpha_{k-1} D_t(x_t, a_t) \right] \\ &\quad - \sum_{t=n}^{\infty} \left[\prod_{k=n}^{t+1} \alpha_{k-1} E_x^\pi [V_{t+1}^\tau(x_{t+1})] - \prod_{k=n}^t \alpha_{k-1} E_x^\pi [V_t^\tau(x_t)] \right]. \quad (4.25) \end{aligned}$$

Observe que, para algún entero positivo M ,

$$\sum_{t=n}^{\infty} \left[\prod_{k=n}^{t+1} \alpha_{k-1} E_x^\pi [V_{t+1}^\tau(x_{t+1})] - \prod_{k=n}^t \alpha_{k-1} E_x^\pi [V_t^\tau(x_t)] \right]$$

$$\begin{aligned}
&= \lim_{M \rightarrow \infty} \sum_{t=n}^M \left[\prod_{k=n}^{t+1} \alpha_{k-1} E_x^\pi [V_{t+1}^\tau(x_{t+1})] - \prod_{k=n}^t \alpha_{k-1} E_x^\pi [V_t^\tau(x_t)] \right] \\
&= \lim_{M \rightarrow \infty} \left[\prod_{k=n}^{M+1} \alpha_{k-1} E_x^\pi [V_{M+1}^\tau(x_{M+1})] - \alpha_{n-1} V_n^\tau(x_n) \right] \\
&= \lim_{M \rightarrow \infty} \left[\frac{P_{M+1}}{P_{n-1}} E_x^\pi [V_{M+1}^\tau(x_{M+1})] - \alpha_{n-1} V_n^\tau(x_n) \right],
\end{aligned}$$

y por la Suposición $P(\tau < +\infty) = 1$,

$$\lim_{M \rightarrow \infty} \frac{P_{M+1}}{P_{n-1}} = 0.$$

Entonces se obtiene en (4.25) que

$$v_n^\tau(\pi, x) = E_x^\pi \left[\sum_{t=n}^{\infty} \prod_{k=n}^t \alpha_{k-1} D_t(x_t, a_t) \right] + \alpha_{n-1} V_n^\tau(x_n).$$

Como $\alpha_{n-1} \leq 1$ y $x = x_n$, se obtiene que

$$v_n^\tau(\pi, x) - V_n^\tau(x) \leq \sum_{t=n}^{\infty} \prod_{k=n}^t \alpha_{k-1} E_x^\pi [D_t(x_t, a_t)].$$

Finalmente, como $v_0^\tau(\pi, x) = j^\tau(\pi, x)$, y tomando $\pi = f^*$, se obtiene (4.2). ■

Obsérvese que a partir de los resultados obtenidos en esta última sección, específicamente del Lema 4.8, y junto con el Lema 4.4, para N^* suficientemente grande se tiene que

$$V_{0, N^*}^\tau(x) \leq J^\tau(x) \leq V(x), \quad (4.26)$$

$x \in X$.

En la práctica, es difícil hallar la solución óptima para el problema con horizonte aleatorio en el caso de soporte infinito. Incluso se ha propuesto una manera de aproximar la política óptima mediante el procedimiento de horizonte rodante, auxiliado del Lema 4.4, uno de los resultados teóricos principales del trabajo de tesis. En principio, la idea sería probar

cualquier política y utilizar (4.26) para decidir si es viable. Por ejemplo, podría probarse la política óptima estacionaria del problema descontado f^* en el problema de horizonte aleatorio, ya que el Lema 4.9 asegura que $j^\tau(f^*, x) \leq V(x)$, $x \in X$.

Conclusiones

En la tesis presentada se estudió un problema de control óptimo con el costo total esperado como criterio de rendimiento, en donde se consideró al horizonte de planeación como una variable aleatoria independiente del proceso de control de Markov. Para ello, se supusieron espacios de estados y acciones de Borel y costo por etapa no acotado, abarcando de esta manera casos más generales que los ya estudiados (véase [18]).

Otro aspecto importante fue considerar una distribución arbitraria para el horizonte, principalmente cuando ésta tuviera soporte infinito, siendo este el caso menos abordado en la literatura por presentar cierto grado de complejidad.

Sin embargo, el caso en que la distribución del horizonte tuviera soporte finito, fue presentado de manera más general en el Capítulo 3, por los supuestos considerados en los espacios de estados y acciones, y la función de costo, siendo esto necesario para el análisis del caso de soporte infinito. Entonces, se garantizó la validez de la técnica de programación dinámica para la obtención de la solución óptima. En este mismo capítulo, se presentaron dos ejemplos que pudieran ilustrar el resultado, siendo el problema de reemplazo de máquinas un ejemplo trabajado desde el modelado del problema hasta la presentación de resultados numéricos obtenidos mediante un programa elaborado en Matlab. Dicho trabajo es presentado en el artículo *Optimal Replacement in a System of n -Machines with Random Horizon* (véase [20]).

En el Capítulo 4, se muestran los principales resultados del trabajo de tesis, partiendo del planteamiento del problema con horizonte aleatorio de soporte infinito como un problema descontado con factor de descuento no homogéneo con horizonte infinito. El uso del factor de descuento homogéneo en el costo total esperado descontado con horizonte infinito puede justificarse matemáticamente para garantizar la convergencia de la

serie infinita presente en el criterio de rendimiento y en algunos problemas puede tener alguna interpretación económica. Sin embargo, en el Capítulo 4 se ha mostrado que puede considerarse alternativamente un costo total esperado con horizonte aleatorio, teniendo un efecto matemático similar al ya mencionado, siendo ésta una consecuencia importante del trabajo de tesis.

A continuación se enlistan los principales resultados del trabajo de tesis.

1. La convergencia de las funciones de valor óptimo de problemas finitos de n etapas, $n = 0, 1, 2, 3, \dots$ a la función de valor óptimo del problema de control con horizonte aleatorio (Lema 4.4).
2. La caracterización de la función de valor óptimo y la existencia de la política óptima (Teorema 4.1).
3. La implementación del procedimiento de horizonte rodante para la aproximación de la política óptima.
4. El acotamiento de la función de valor óptimo del problema con horizonte aleatorio por la función de valor óptimo del problema descontado (Lema 4.8).

La implementación del procedimiento de horizonte rodante fue probado en el problema de reemplazo óptimo de máquinas y el Lema 4.4 fue utilizado para obtener una referencia del error numérico de aproximación. El trabajo realizado es presentado en el artículo *Rolling Horizon Procedures for the Solution of an Optimal Replacement Problem of n -Machines with Random Horizon* (véase [19]).

Un tercer artículo, titulado *Markov Decision Processes on Borel Spaces with Total Cost and Random Horizon*, fue estructurado con los puntos 1, 2 y 4 antes mencionados (véase [11]), siendo el de mayor importancia por contener los principales resultados teóricos establecidos en la tesis.

Matemáticamente hablando resulta interesante medir la sensibilidad en otras componentes del modelo de control de Markov. Por ejemplo, partiendo de un sistema estocástico, bajo ciertas condiciones en la ley de transición es posible comparar las soluciones óptimas del problema original con un sistema determinista adecuado (véase [10]). En nuestro caso, en el punto 4 se hizo el análisis de dicha sensibilidad.

Finalmente, se enlistan algunas posibles extensiones del trabajo de tesis.

- Estudiar el problema bajo diferente bloque de suposiciones en el modelo de control de Markov. Por ejemplo, podrían considerarse las suposiciones del Capítulo 8 de [16], adaptadas al problema con horizonte aleatorio.
- Proponer una forma de aproximación de la solución del problema con horizonte aleatorio mediante el uso de la política óptima del problema descontado asociado a dicho problema y medir su efectividad de acuerdo con los acotamientos que se obtuvieron para su correspondiente función de valor óptimo.
- Demostrar que en el problema de reemplazo óptimo, el considerar máquinas independientes y costos de operación y reemplazo de máquinas también independientes reduce a resolver el problema separadamente para cada máquina.
- Resolver el problema de reemplazo óptimo considerando máquinas no independientes con costos no lineales de reemplazo y horizonte aleatorio.

Apéndice A

Programas

A.1. Programa para la Obtención de la Solución Óptima del Problema de Reemplazo

```
clear;
N=6; % valor máximo que puede tomar el horizonte+1
dh=[.1,.1,.3,.2,.15,.15]; % distribución del horizonte
pt(1)=1;
for i=2:N % Cálculo de las p_t
    pt(i)=pt(i-1)-dh(i-1);
end
D=3; % Niveles de deterioro de la máquina
n=3; % número de máquinas
c=[5,7,29]; % costo por operar cualquier máquina en cada nivel de deterioro
R=4; % costo por reemplazar una máquina
P0=[.4,.3,.3;
    0,.3,.7;
    0,0,1]; % matriz de transición cuando se decide no reemplazar
P1=ones(D,1);
P1=[P1,zeros(D,D-1)]; % matriz de transición cuando se decide reemplazar una máquina
```

```

for i=n:-1:1, % obtención los estados del sistema
    cont=1;
    while cont<=D^n,
        for j=1:D,
            for k=1:D^(n-i)
                E(cont,i)=j;
                cont=cont+1;
            end
        end
    end
end
E,% espacio de estados (el renglón k será el estado número k)
for i=n:-1:1, % obtención de las posibles acciones
    cont=1;
    while cont<=2^n,
        for j=0:1,
            for k=1:2^(n-i)
                A(cont,i)=j;
                cont=cont+1;
            end
        end
    end
end
A,%espacio de acciones (el renglón k será la acción número k)
for i=1:2^n,% q(j,k,i) es la probabilidad de pasar del estado
número j al número k del sistema dada la acción número i
    for j=1:D^n
        for k=1:D^n
            q(j,k,i)=1;
            for s=1:n
                if A(i,s)==0;
                    q(j,k,i)=P0(E(j,s),E(k,s))*q(j,k,i);
                else
                    q(j,k,i)=P1(E(j,s),E(k,s))*q(j,k,i);
                end
            end
        end
    end
end
end

```

```

    end
end
for j=1:D^n
    V(N+1,j)=0;% V(i,j) es la función de valor en la etapa
i-1 del estado j-ésimo (E(:,j))
end
% cálculo de la función de valor óptima y política óptima
for i=N:-1:1 % para cada etapa
    for j=1:D^n % para cada estado
        for s=1:2^n % para cada acción
            vp(s)=0;

            % valor esperado de v(y) dado el estado número j
y la acción número s
            for k=1:D^n
                vp(s)=vp(s)+V(i+1,k)*q(j,k,s);
            end
            C(s)=0; %c(j,s)
            for k=1:n
                if A(s,k)==0
                    C(s)=c(E(j,k))+C(s);
                else
                    C(s)=c(1)+R+C(s);
                end
            end
            v(s)=pt(i)*C(s)+vp(s); % argumento del min en
la ecuación de programación dinámica
        end
        V(i,j)=v(1);
        P(i,j)=1;
        for s=2:2^n % búsqueda del mínimo o acción óptima
            if v(s)<V(i,j)
                V(i,j)=v(s);
                P(i,j)=s; % número de acción que es óptima
para el estado j en la etapa i-1
            end
        end
    end
end

```

```

        end
    end
    V, % función de valor óptimo
    P, % política óptima

```

A.2. Programa para la Obtención de la Política de Horizonte Rodante

```

> restart:
> Digits:=400: # número de dígitos utilizados para el cálculo
de los alfas
> HR:=20: # longitud del horizonte rodante (N)
> H:=500: # cantidad de alfas calculados
> alfa:=Array(1..H): # arreglo para los alfas
> Pt:=Array(1 .. H):
> p:=.8: # parámetro de la distribución del horizonte (tipo
logaritmo)
> fda:=0:
> pt:=1:
> Pt[1]:=1:
> for t from 1 to H do # cálculo de los alfas
    fda:=fda-(1-p)^t/(t*ln(p)):
    pt1:=1-fda:
    alfa[t]:=pt1/pt:
    Pt[t]:=pt1:
    pt:=pt1:
end do:
> fda: # si fda=1, ha fallado el cálculo de los alfas y hay
que aumentar el número de dígitos para mayor precisión
> alfa;
> d:=3: # número de niveles de deterioro
> n:=3: # número de máquinas
> c:=[5, 7, 29]: # c[i] es el costo por operar una máquina en
el nivel i
> R:=4: # costo por reemplazar una máquina
> P0:=Matrix([[0.4, 0.3, 0.3],

```

```

                [0, 0.3, 0.7],
                [0, 0, 1]]):# Matriz de transición cuando se de-
cide no reemplazar una máquina
> P1:=Matrix([[1, 0, 0],
              [1, 0, 0],
              [1, 0, 0]]):# Matriz de transición cuando se de-
cide reemplazar una máquina
> E:=Array(1 .. d^n, 1 .. n):
> for i from n by -1 to 1 do # construcción del espacio de es-
tados
    cont:=1:
    while cont <= d^n do
        for j from 1 to d do
            for k to d^(n-i) do
                E[cont,i]:=j:
                cont:=cont+1:
            end do:
        end do:
    end do:
end do:
> E:
> A := Array(1..2^n,1..n):
> for i from n by -1 to 1 do #construcción del espacio de ac-
ciones
    cont:=1:
    while cont <= 2^n do
        for j from 0 to 1 do
            for k to 2^(n-i) do
                A[cont,i]:=j:
                cont:=cont+1:
            end do:
        end do:
    end do:
end do:
> A:
> q := Array(1..d^n, 1..d^n, 1..2^n): # q[j,k,i] es la proba-
bilidad de pasar del estado número j al número k del sistema,

```

```

dada la acción i
> for i from 1 to 2^n do #construcción de la matriz de tran-
sición de estados del sistema dada cada acción
  for j from 1 to d^n do
    for k from 1 to d^n do
      q[j,k,i]:=1:
      for s from 1 to n do
        if A[i,s]=0 then
          q[j,k,i]:=P0[E[j,s], E[k,s]]*q[j, k,i]:
        else
          q[j,k,i]:=P1[E[j,s],E[k,s]]*q[j,k,i]:
        end if:
      end do:
    end do:
  end do:
end do:
> Verror:=Array(1..d^n): # aquí se guardan las diferencias de
la función de valor óptimo para dos tamaños truncados conse-
cutivos del horizonte
> epsilon:=10^(-19): # cota de error permitida para indicar la
aproximación de la función de valor
> V1:=Array(1..d^n):
> V2:=Array(1..d^n):
> Vaprox:=Array(1..d^n):# aproximación de la función de valor
óptimo
> Error:=2*epsilon:
> ind:=0:
> cont:=0:
> for H1 from 1 to H do # obtención de la función de valor óp-
timo de los subproblemas finitos de H1 etapas
  if Error > epsilon then
    V:=Array(1..H1+1,1..d^n):
    for i from 1 to d^n do
      V[H1+1,i]:=0:
    end do:
    for i from H1 by -1 to 1 do
      for j from 1 to d^n do

```

```

v:=Array(1..2^n):
vp:=Array(1..2^n):
C:=Array(1..2^n):
for s from 1 to 2^n do
  vp[s]:=0:
  for k from 1 to d^n do
    vp[s]:=vp[s]+V[i+1,k]*q[j,k,s]:
  end do:
  C[s]:=0:
  for k from 1 to n do
    if A[s,k] = 0 then
      C[s]:=c[E[j,k]]+C[s]:
    else
      C[s]:=c[1]+R+C[s]
    end if
  end do:
  v[s]:=C[s]+alfa[i]*vp[s]:
end do:
V[i,j]:=v[1]:
for s from 2 to 2^n do
  if v[s] < V[i,j] then
    V[i,j]:=v[s]:
  end if
end do:
end do:
cont:=cont+1:
if cont=1 then
  for k from 1 to d^n do
    V2[k]:=V[1,k]:
  end do:
  Error:=2*epsil:
else
  for k from 1 to d^n do
    V1[k]:=V2[k]:
    V2[k]:=V[1,k]:
  end do:

```

```

        for k to d^n do
            Verror[k]:=abs(V2[k]-V1[k]):
        end do:
        Error:=Verror[1]:
        for k from 2 to d^n do
            if Error<Verror[k] then
                Error:=Verror[k]:
            end if
        end do:
    end if
else
    ind:=1:
    Haprox:=H1:
    H1:=H+1:
    Vaprox:=V2
end if
end do
> ind: # indica si se pudo obtener una aproximación de la función de valor óptimo con el error dado
> Haprox:
> V:=Array(1..HR+1,1..d^n):
> for i from 1 to d^n do
    V[HR+1,i]:=0:
end do:
> P:=Array(1..HR,1..d^n):
> numpr:=15: # número de acciones de la política de horizonte rodante calculadas (a lo más H-HR-1 debido al número de alfas disponibles)
> Polrod:=Array(1..numpr,1..d^n):
> for l from 0 to numpr-1 do
    for i from HR by -1 to 1 do
        for j from 1 to d^n do
            v:=Array(1..2^n):
            vp:=Array(1..2^n):
            C:=Array(1..2^n):
            for s from 1 to 2^n do
                vp[s]:=0:
            end do:
        end do:
    end do:
end do:

```

```

        for k from 1 to d^n do
            vp[s]:=vp[s]+V[i+1,k]*q[j,k,s]:
        end do:
        C[s]:=0:
        for k from 1 to n do
            if A[s,k]=0 then
                C[s]:=c[E[j,k]]+C[s]:
            else
                C[s]:=c[1]+R+C[s]:
            end if
        end do:
        v[s]:=C[s]+alfa[i+1]*vp[s]:
    end do:
    V[i, j]:=v[1]:
    P[i, j]:=1:
    for s from 2 to 2^n do:
        if v[s]<V[i,j] then
            V[i,j]:=v[s]:
            P[i, j]:=s:
        end if
    end do:
end do:
end do:
for k from 1 to d^n do:
    Polrod[l+1,k]:=P[1,k]:
end do:
end do:
> Polrod: #Política de horizonte rodante
> EvalPR:=Array(1..d^n): # evaluación de la política de hori-
zonte rodante en el criterio de rendimiento
for i from 1 to d^n do
    EvalPR[i]:= 0:
    for j from 1 to n do
        if A[Polrod[1,i],j]=0 then
            EvalPR[i]:=EvalPR[i]+c[E[i, j]]
        else
            EvalPR[i]:=EvalPR[i]+c[1]+R:
        end if
    end do
end do

```

```

        end if
    end do:
end do:
na:=5: # etapas evaluadas, se requiere de un arreglo de tamaño  $d^{(n*(s-1))}$  para la etapa s
for s from 2 to na do
    if s=2 then
        Pr:=Array(1..dn,1..d(n*(s-1))):
        for t from 1 to dn do
            for u from 1 to d(n*(s-1)) do
                Pr[t,u]:=q[t,u,Polrod[s-1,t]]:
            end do:
        end do:
    else
        Pra:=Array(1..dn,1..d(n*(s-2))):
        for t from 1 to dn do
            for u from 1 to d(n*(s-2)) do
                Pra[t,u]:=Pr[t,u]:
            end do:
        end do:
        Pr:=Array(1..dn,1..d(n*(s-1))):
        for t from 1 to dn do
            cont:=0:
            cont2:=0:
            for u from 1 to d(n*(s-2)) do
                cont2:=cont2+1:
                for v to dn do
                    cont:=cont+1:
                    Pr[t,cont]:=Pra[t,u]*q[cont2,v,Polrod[s-1,cont2]]:
                end do:
            end do:
            if cont2=dn then
                cont2:=0:
            end if
        end do:
    end do:
end if:

```

```

cxa:=Array(1 .. d^n):
for k to d^n do
  for j from 1 to n do
    if A[Polrod[s,k],j]=0 then
      cxa[k]:=cxa[k]+c[E[k,j]]:
    else
      cxa[k]:=cxa[k]+c[1]+R:
    end if
  end do:
end do:
Pr1:=Array(1..d^n,1..d^n):
for i from 1 to d^n do
  for j from 1 to d^n do
    for k from j by d^n to d^(n*(s-1)) do
      Pr1[i,j]:=Pr1[i,j]+Pr[i,k]:
    end do:
  end do:
end do:
for i from 1 to d^n do
  Ecxa:=0:
  for j from 1 to d^n do
    Ecxa:=Ecxa+Pr1[i,j]*cxa[j]:
  end do:
  EvalPR[i]:=EvalPR[i]+Pt[s-1]*Ecxa:
end do:
end do:
> EvalPR; # resultado de la evaluación de la política de hori-
zonte rodante en el criterio de rendimiento
> Vaprox; # aproximación de la función de valor óptimo por me-
dio de la resolución de un problema truncado

```


Bibliografía

- [1] Alden J. M., Smith R. L. “Rolling Horizon Procedures in Nonhomogeneous Markov Decision Process”. *Operations Research*. Vol. 40, pp. 183-194. 1992.
- [2] Altman E., “Applications of Markov Decisions Processes in Communication Networks”. In *Handbook of Markov Decision Processes: Methods and Applications* by Eugene Feinberg and Shwartz Kluwer, pp. 489-536. 2001.
- [3] Bather J., “Decision Theory: An Introduction to Dynamic Programming and Sequential Decision”. John Wiley & Sons. 2000.
- [4] Bäuerle N., Rieder U., “Markov Decision Processes with Applications to Finance”. Springer-Verlag. 2010.
- [5] Bellman R. E., “Dinamyc Programming”, Princeton University Press. 1957.
- [6] Bertsekas D. P., “Dynamic Programming and Optimal Control”. Vol. 1. Athena Scientific. 1995.
- [7] Bertsekas D. P., Shreve S. E., “Stochastic Optimal Control: The Discrete-Time Case”. Athena Scientific. 1996.
- [8] Chatterjee D., Cinquemani E., Chaloulos G., Lygeros J., “Stochastic Control up to a Hitting Time: Optimality and Rolling-Horizon Implementation”. arxiv 0806.3008. 2009.
- [9] Childress S., Durango-Cohen P., “On Parallel Machine Replacement Problems with General Replacement Functions and Stochastic Deterioration”. *Naval Research Logistics*. Vol. 52, pp. 409-419. 2005.

- [10] Cruz-Suárez H., Ilhuicatzí-Roldán R. “Stochastic Optimal Control for Small Noise Intensities: The Discrete-Time Case”. WSEAS Transactions on Mathematics. Vol. 9, pp. 120-129. 2010.
- [11] Cruz-Suárez H., Ilhuicatzí-Roldán R., Montes-de-Oca R., “Markov Decision Processes on Borel Spaces with Total Cost and Random Horizon”. Journal of Optimization Theory and Applications. 2012.
- [12] Ermoliev Y., Ermolieva T., Fischer G., Makowski M., “Extreme Events, Discounting and Stochastic Optimization”. Ann Oper Res. Vol. 177, pp. 9-19. 2009.
- [13] Guo X., Hernandez-del-Valle A., Hernández-Lerma O., “Nonstationary Discrete-Time Deterministic and Stochastic Control Systems with Infinite Horizon”. International Journal of Control. Vol. 83, pp. 1751-1757. 2010.
- [14] Hernández-Lerma O., “Adaptive Markov Control Processes”. Springer-Verlag. 1989.
- [15] Hernández-Lerma O., Lasserre J. B., “Discrete-Time Markov Control Processes: Basic Optimality Criteria”. Springer-Verlag. 1996.
- [16] Hernández-Lerma O., Lasserre J. B., “Further Topics on Discrete-Time Markov Control Processes”. Springer-Verlag. 1999.
- [17] Hinderer K., “Foundation of Non-stationary Dynamic Programming with Discrete Time Parameter”. Springer-Verlag. 1970.
- [18] Iida T., Mori M., “Markov Decision Processes with Random Horizon”. Journal of the Operations Research. Vol. 39, pp. 592-603. 1996.
- [19] Ilhuicatzí-Roldán R., Cruz-Suárez H., “Rolling Horizon Procedures for the solution of an Optimal Replacement Problem of n -Machines with Random Horizon”. Investigación Operacional (por aparecer).
- [20] Ilhuicatzí-Roldán R., Cruz-Suárez H., “Optimal Replacement in a System of n -Machines with Random Horizon”. Proyecciones Journal of Mathematics. Vol. 31, pp. 219-233. 2012.

- [21] Kozłowski E., “The Linear-Quadratic Stochastic Optimal Control Problem with Random Horizon at the Finite Number of Infinitesimal Events”. *Annales UMCS Informatica*. Vol. 1, pp. 103–115. 2010.
- [22] Le Van C., Dana R. A., “Dynamic Programming in Economics”. Kluwer Academic Publishers. 2003.
- [23] Levhari D., Mirman L. J., “Savings and Consumption with Uncertain Horizon”. *Journal of Political Economy*. Vol. 85, pp. 265–281. 1977.
- [24] Mori K., “Approximate Linear Quadratic Regulator Problem and its Applications to Optimal Control in Discrete-Time LTI Systems”. *Intelligente Control and Innovative, Lecture Notes in Electrical Engineering*. Vol. 110, pp.53-64.
- [25] Nair S. K., Hopp W. J., “A Model for Equipment Replacement due to Technological Obsolescence”. *European Journal of Operational Research*. Vol. 63, pp. 207-221. 1992.
- [26] Powell W. B., “Approximate Dynamic Programming: Solving the Curses of Dimensionality”, John Wiley & Sons. 2007.
- [27] Puterman M. L., “Markov Decision Process: Discrete Stochastic Dynamic Programming”. John Wiley & Sons. 1994.
- [28] Shapley L. S., “Stochastic games”. *Proceedings of the National Academy of Sciences*, pp. 1095-1100. 1953.
- [29] Schal M., “Markov Decision Processes in Finance and Dynamic Options”. In *Handbook of Markov Decision Processes: Methods and Applications* by Eugene Feinberg and Shwartz Kluwer, pp. 460-488. 2001.
- [30] Sheti S. P., Sorger G., Zhou X. Y., “Stability of Real-Time Lot-Scheduling and Machine Replacement Policies with Quality Levels”. *IEEE Transactions on Automatic Control*. Vol. 45, pp. 1-17. 2000.
- [31] Van Roy B., “Neuro-Dynamic Programming: Overview and Recent-trends”. In *Handbook of Markov Decision Processes: Methods and*

Applications by Eugene Feinberg and Shwartz Kluwer”, pp. 431-460,
20

- [32] Wang B., Liu D., “Rolling Horizon Procedure for Large-Scale Equipment Replacement Problems with Multiple Alternatives”. Control and Decision Conference, Chinese. 2011.