"El saber de mis hijos hará mi grandeza"

UNIVERSIDAD DE SONORA

División de Ciencias Exactas y Naturales

Programa de Posgrado en Matemáticas

Aproximación Numérica para Iteración de Valores en Procesos de Control Markoviano con Espacios Generales y Costos Acotados

T E S I S

Que para obtener el grado académico de:

Doctor en Ciencias
(Matemáticas)

Presenta:

M.C. Joaquín Humberto López Borbón

Director de Tesis: Dr. Óscar Vega Amaya

Hermosillo, Sonora, México, 21 de Marzo del 2015.

Universidad de Sonora

Repositorio Institucional UNISON

SINODALES

Dr. Carlos Pacheco Gonzalez
CINVESTAV-IPN, Ciudad de México, DF.

Dr. Óscar Vega Amaya
UNISON, Hermosillo, Sonora, México.

Dr. Fernando Luque Vázquez
UNISON, Hermosillo, Sonora, México.

Dr. Adolfo Minjarez Sosa
UNISON, Hermosillo, Sonora, México.

Dr. Hector Jasso Fuentes
CINVESTAV-IPN, Ciudad de México, DF.

Agradecimientos

Primeramente agradezco al nuevo objetivo de mi vida, mi hijo Joaquinito, a mis hijas Katy y Mayrani por su cariño y ternura y a Irene con todo mi amor. A todos ellos, en minima reciprocidad al tiempo que les ocupe al realizar este trabajo.

Agradezco también a mis hijos Carlitos, Carolina e Ivan, en prueba de que "*cuando se quiere, se puede*". Y en general extender mis agradecimientos a toda mi familia.

De manera muy especial, quiero agradecerle a mi director de tesis Dr. Óscar Vega Amaya, por el tiempo dedicado a dirigir este trabajo de investigación y sobre todo por sus atinados controles en este proceso de investigación. Quiero dar las gracias a los integrantes del comité revisor de este trabajo: Dr. Carlos Pacheco Gonzalez, Dr. Fernando Luque Vázquez, Dr. Adolfo Minjarez Sosa y Dr. Hector Jasso Fuentes, sin duda, sus comentarios, sugerencias y correcciones me han permitido culminar este trabajo. Y en general agradezco a mis profesores del Posgrado del Departamento de Matemáticas UNISON.

Asimismo, agradezco los apoyos institucionales y becas recibidas del Departamento de Matemáticas UNISON y CONACyT.

Finalmente agradezco a "*juan pueblo*", que con sus impuestos me han pagado mi formación académica. Sin embargo, a ellos les llega el mínimo de los avances científico a máximo costo. ¡Creo que nunca podré pagarte "*juan pueblo*"!

Contenido

Introducción

"Estudia el pasado si quieres pronosticar el futuro".

Confucio (551 AC-478 AC).

"El destino es el que baraja las cartas, pero nosotros somos los que jugamos".

Shakespeare (1564-1616).

"Nunca se puede predecir un acontecimiento físico con una precisión absoluta".

Max Planck (1858-1947).

"Mientras más conscientes estemos de las dificultades, más debemos luchar por optimizar el esfuerzo".

Fidel Castro Ruz.

Una amplia gama de problemas en economía, investigación de operaciones, explotación de recursos renovables y no renovables, etc., son modelados como problemas de optimización de sistemas estocásticos que evolucionan en el tiempo ([4], [6], [10], [11], [12], [19], [23], [25], [30]). Estos problemas consisten en la toma de decisiones secuenciales con el fin de conducir al sistema hacia un comportamiento óptimo. La elección de la mejor decisión entre las distintas alternativas (denominadas acciones o controles) requiere pensar en los efectos que cada una de ellas puede generar. Los efectos inmediatos a menudo pueden verse, pero los efectos a largo plazo no siempre son trasparentes, de tal manera que algunos controles pueden tener respuestas inmediatas muy pobres pero a la larga pueden ser los mejores. Así que el problema consiste en escoger aquellos controles que proporcionen un adecuado balance entre los costos (ó beneficios) inmediatos y los futuros. La dificultad del decisor radica en la incertidumbre sobre el futuro ya que los resultados de los controles son aleatorios.

Sistemas dinámicos que involucran la toma de decisiones secuenciales en ambiente de incertidumbre, cuya evolución hacia estados futuros sólo depende del estado y control actual, pero no de la historia pasada, son modelados mediante modelos de control markoviano (MCM). Un MCM consta de un espacio de estados, un espacio de controles, una ley de transición y una función de costo. Al inicio de cada etapa (tiempo trascurrido entre toma de deciones consecutivas), el controlador observa el estado en que se encuentra el sistema y escoge un control factible, provocando que el sistema se mueva a otro estado de acuerdo a la ley de transición, lo cual genera un costo. Cuando la observación de estados y la aplicación de controles se hace en tiempos discretos se dice que el MCM es a tiempo discreto. Es en el contexto de estos modelos donde desarrollaremos nuestro trabajo.

En los MCM el problema principal es encontrar la regla de decisión (política) que específica el mejor control a aplicar para cualquier estado del sistema. Cada

política define un proceso controlado y un costo total acumulado durante el proceso dado por una función objetivo ó índice de funcionamiento. Entonces el problema de control óptimo (PCO) consiste en encontrar la política que lleve al sistema a un comportamiento óptimo de acuerdo a un índice de funcionamiento (o función objetivo) especificado. Los índices de funcionamiento más usuales son el índice en costo descontando y el índice en costo promedio.

Actualmente, la teoría de los procesos de control markoviano está muy desarrollada y proporciona un marco sumamente flexible para el análisis de muchos problemas de optimización secuencial en diversas áreas ([4], [6], [10], [11], [12], [19], [23], [25], [30]). Sin embargo, es bien conocido que para sistemas con espacio de estados muy grandes o continuos, la teoría de los procesos controlados sufre el fenómeno conocido como la "*maldición de la dimensión*" [21], la cual hace imposible la obtención numérica de políticas óptimas en un tiempo computacional razonable. Por tal motivo muchos investigadores han propuesto métodos de aproximación con la esperanza de romper o mitigar la maldición de la dimensión obteniendo políticas sub-óptimas o que al menos sean "*buenas*" soluciones ([1], [5], [8], [9], [15], [17], [20], [21], [22], [24], [26]). Sin embargo, la mayor parte de estos trabajos están concentrados en espacios discretos, principalmente en el caso finito, o bién en el criterio de costo descontado.

El principal enfoque para resolver un PCO consiste en encontrar una solución de la ecuación de optimalidad, que en el caso descontado toma la forma

$$V = TV \tag{1}$$

donde el operador T denominado operador de programación dinámica es aplicado en un espacio de funciones.

Un método estandar para obtener una solución de la ecuación de optimalidad es iteración de valores (IV), el cual consiste en aplicar iterativamente el operador T en un espacio de funciones especificado ([4], [10], [11], [12], [19]). El algoritmo IV genera una sucesión de funciones aproximantes de acuerdo a la siguiente ecuación recursiva

$$V_{n+1} = TV_n, \quad n \in \mathbb{N}_0, \tag{2}$$

bajo ciertas condiciones el operador T es un operador de contracción en el espacio de funciones considerado, de manera que el teorema de punto fijo de Banach asegura la existencia de una función V^* en dicho espacio tal que $V^* = TV^*$. Sin embargo, el algoritmo IV tiene dos obstáculos computacionales: (I) el algoritmo tiene que parar en un número finito de iteraciones; (II) tiene que obtener TV_n para cada elemento del espacio de estados. Para espacios de estados continuo o finito muy grande, la implementación computacional TV_n es imposible, de modo que sólo queda obtener aproximaciones de TV_n y obtener políticas sub-óptimas con dos errores, uno por parar y el otro por aproximar.

En la literatura se han considerado una variedad de esquemas de aproximación, entre ellos, podemos mencionar: algoritmo iteración de valores aproximados (IVA) ([8], [17], [20]), programación neuro-dinámica [5] o aprendizaje reforzado [26], entre otros. La idea es combinar esquemas adecuados de aproximación de funciones con el algoritmo estandar IV.

En este trabajo estudiaremos el enfoque IVA, donde el esquema de aproximación es representado por medio de un operador L, donde Lv es la aproximación de la función v, el cual es aplicado entre dos ejecuciones consecutivas de T en el algoritmo IV. Existen dos algoritmos IVA los cuales son generados dependiendo de cual operador T ó L actua primero. Dichos algoritmos IVA son dados por los operadores compuestos $\widetilde{T} = LT$ y $\widehat{T} = TL$, los cuales definen las respectivas ecuaciones de los algoritmos

$$\widetilde{V}_{n+1} = \widetilde{T}\widetilde{V}_n, \quad n \in \mathbb{N}_0, \tag{3}$$

y

$$\widehat{V}_{n+1} = \widehat{T}\widehat{V}_n, \quad n \in \mathbb{N}_0. \tag{4}$$

Las propiedades del operador L determinan fuertemente la bondad de los algoritmos IVA. Por ejemplo, la convergencia de las sucesiones (3) y (4) no se puede garantizar a menos que L sea no-expansivo con respecto a una norma adecuada ([8], [9], [21]).

Cuando un algoritmo IVA, por ejemplo el (3), es detenido en la n-ésima iteración, se obtiene la política estacionaria f_n que optimiza $\widetilde{T}\widetilde{V}_n$ (dicha política f_n se le denomina $\widetilde{V}_n{-}greedy$) y se aproxima la función valor óptimo V^* por medio de la función de valor V_{f_n} correspondiente a la política f_n. Entonces, de forma natural surgen los siguientes problemas:

(D1) El primero es la convergencia de las funciones $(\widetilde{V}_n)_{n\in\mathbb{N}}$.

(D2) Asegurada la convergencia anterior, digamos a $\widetilde{V}^*$, el segundo problema es encontrar la cota de la desviación $\widetilde{V}^* - V^*$ expresada en términos de la precision de la aproximación proporcionada por el operador L.

(D3) El tercero, quizás el más importante desde el punto de vista práctico, consiste en encontrar cotas para el error de aproximación $e(f_n) := V_{f_n} - V^*$.

Obviamente se plantean los mismos problemas para el el algoritmo IVA (4) remplazando $\widetilde{T}$, $\widetilde{V}_n$ y $\widetilde{V}^*$ por $\widehat{T}$, $\widehat{V}_n$ y $\widehat{V}^*$, respectivamente.

Para el PCO en costo descontado, los artículos [1], [17] y [24] proveen cotas de desempeño interesantes para el algoritmo IVA. Sus enfoques y resultados difieren de los presentados en este trabajo en varias formas. Para comenzar, Almudevar [1] desarrolla un enfoque abstracto y primero estudia aproximaciones para problemas de punto fijo en general, y luego aplica sus resultados a varios tipos de problemas de decisión markoviana. Sin embargo, Almudevar sólo obtiene resultados asintóticos y sus cotas no están ligadas a la precisión con que el operador L aproxima a los elementos del modelo de control original. Por otro lado, Munos [17] proporciona cotas L_p expresadas en términos de ciertas cantidades que el denomina "coeficientes de concentración de la probabilidad de transición", pero éstas son difíciles de calcular o acotar excepto en algunos casos simples; además, como en el artículo de Almudevar [1], las cotas no son expresadas en términos de la precisión del esquema de aproximación. Stachurski [24] elimina esta última desventaja bajo la suposición adicional de que el operador de programación dinámica

transforma funciones monótonas en funciones monótonas, lo cual limita la utilidad de su resultado. Adicionalmente Almudevar [1] y Munos [17] requieren para el análisis del caso de espacio de estados continuos que la ley de transición del sistema admita una función de densidad, la cual no es necesaria en el presente trabajo.

El PCO en costo promedio es mas complicado desde el punto de vista matemático; de hecho algunos autores lo consideran no como un sólo problema, sino como una colección de problemas [2]. La ecuación de optimalidad en este caso es de la forma

$$\rho^* + h^* = Th^*, \tag{5}$$

donde ρ^* es una constante, h^* pertenece a cierto espacio de funciones y T es el operador de programación dinámica correspondiente. Si el par (ρ^*, h^*) satisface (5) y ρ^* es acotada, entonces se prueba utilizando argumentos estándares que ρ^* es el costo promedio óptimo y que la política estacionaria f^* obtenida al optimizar en el lado derecho de la ecuación de optimalidad es una política estacionaria óptima.

Un método estándar para obtener una solución de la ecuación de optimalidad (5) es el de aproximaciones sucesivas o iteración de valores (IV) en costo promedio, que consiste en aplicar iterativamente el operador T en un espacio de funciones especificado ([4], [10], [11], [12], [19]). El algoritmo IV en costo promedio genera una sucesión de funciones aproximantes de acuerdo a la siguiente ecuación recursiva

$$J_{n+1} = TJ_n, \quad n \in \mathbb{N}_0, \tag{6}$$

comenzando con $J_0 \equiv 0$, y se obtiene una solución (ρ^*, h^*) como límite de las sucesiones

$$\begin{aligned} \rho_{n+1}(z) &= J_{n+1}(z) - J_n(z), \quad n \in \mathbb{N}_0, \\ h_{n+1}(\cdot) &= J_{n+1}(\cdot) - J_{n+1}(z), \quad n \in \mathbb{N}_0, \end{aligned}$$

donde z es un estado arbitrario fijo. Usando el operador de programación dinámica estas sucesiones están relacionadas como sigue

$$\rho_{n+1}(z) + h_{n+1}(\cdot) = Th_n(\cdot), \quad n \in \mathbb{N}_0,$$

donde $h_0 \equiv 0$. Note que, $\rho_{n+1}(z) = Th_n(z)$ para todo $n \in \mathbb{N}_0$, porque $h_n(z) = 0$. Entonces se tiene otra forma equivalente del algoritmo IV en costo promedio

$$h_{n+1}(\cdot) = T_z h_n(\cdot), \quad n \in \mathbb{N}_0, \tag{7}$$

donde el operador es definido como

$$T_z u(\cdot) := Tu(\cdot) - Tu(z), \tag{8}$$

para funciones u que pertenecen a cierto espacio de funciones apropiado. Note que el par (ρ^*, h^*) con $h^*(z) = 0$ satisface la ecuación de optimalidad (5) si y sólo si h^* es punto fijo de T_z. Para consultar resultados más generales y recientes respecto al criterio costo promedio, consulte por ejemplo [11] y [12].

Sin embargo, al igual que en el PCO en costo descontado, este procedimiento también tiene los dos obstáculos computacionales mencionados para el caso

descontado: (I) tiene que parar después de un número finito de iteraciones; (II) tiene que obtener TJ_n para cada elemento del espacio de estados. La implementación computacional de TJ_n en (6) ó $T_z h_n$ en (7) en espacios continuos o discretos muy grandes es imposible, haciendo también imposible obtener una representación numérica exacta de la sucesión $\{(\rho_n, h_n)\}_{n\in\mathbb{N}}$.

Para evitar esta dificultad, en forma similar al caso descontado, se consideran los siguientes algoritmos IVA

$$\widetilde{h}_{n+1} = \widetilde{T}_z \widetilde{h}_n := L T \widetilde{h}_n - L T \widetilde{h}_n(z), \quad n \in \mathbb{N}_0 \tag{9}$$

y

$$\widehat{h}_{n+1} = \widehat{T}_z \widehat{h}_n := T L \widehat{h}_n - T L \widehat{h}_n(z), \quad n \in \mathbb{N}_0 \tag{10}$$

donde T es el operador de programación dinámica correspondiente y L es un operador de aproximación.

Como en el caso descontado, si el algoritmo IVA (9) es detenido en la n-ésima iteración, se obtiene la política estacionaria f_n, $\widetilde{h}_n - greedy$ que optimiza $\widetilde{T}\widetilde{h}_n$, y se aproxima el costo promedio optimo ρ^* por medio del costo promedio $J(f_n)$ incurrido por la política f_n. Entonces, debido a los dos obstáculos computacionales del algoritmo IV, se generan los siguientes problemas:

(P1) El primero es la convergencia de la sucesión $(\widetilde{\rho}_n, \widetilde{h}_n)_{n\in\mathbb{N}}$.

(P2) Asegurada la convergencia anterior, digamos a $(\widetilde{\rho}^*, \widetilde{h}^*)$, el segundo problema es encontrar las cotas de las desviaciones $\widetilde{\rho}^* - \rho^*$ y $\widetilde{h}^* - h^*$ expresadas en términos de la precisión o calidad de aproximación que proporciona el operador L.

(P3) El tercero quizás, el más importante desde el punto de vista práctico, consiste en encontrar cotas para el error de aproximación $e(f) := J(f_n) - \rho^*$.

Se plantean los mismos problemas para el el algoritmo IVA dado en (10) remplazando $\widetilde{T}$, $(\widetilde{\rho}_n, \widetilde{h}_n)$ y $(\widetilde{\rho}^*, \widetilde{h}^*)$, por $\widehat{T}$, $(\widehat{\rho}_n, \widehat{h}_n)$ y $(\widehat{\rho}^*, \widehat{h}^*)$, respectivamente.

El objetivo de esta tesis es proporcionar métodos computables para obtener aproximaciones a la política óptima en modelos de control markoviano con respecto a los índices en costo descontado y en costo promedio para procesos con espacio de Borel y costos acotados. Especificamente, obtiener cotas de desempeño del algoritmo IVA para dichos índices en espacio continuos y costos acotados.

Para tal propósito en este trabajo nos restringimos a la clase de operadores de aproximación con las siguientes propiedades: (i) las funciones contantes son puntos fijos del operador; (ii) es un operador lineal y positivo; (iii) tiene ciertas propiedades de continuidad (definición 3.1). Algunos esquemas de aproximación ampliamente usados en la literatura son representados por operadores con tales propiedades, algunos de estos son: aproximadores constantes por pedazos, interpoladores lineales y multilineales, aproximadores basados en kernel ([9], [24]). El resultado clave de este trabajo es que un operador con estas propiedades define una probabilidad de transición en el espacio de estados del sistema controlado, lo

cual tiene las siguientes consecuencias importantes: (a) se puede ver la función aproximante como el "*valor promedio*" de la función que se desea aproximar con respecto a la transición de probabilidad inducida por el operador; (b) el paso de aproximación en el algoritmo IVA puede pensarse como una perturbación del modelo original. Por estas propiedades a los operadores que satisfacen (i)-(iii) les llamaremos ***promediadores.***

El presente trabajo sigue el enfoque descrito en párrafos anteriores y busca solucionar los problemas (D1)-(D3) y (P1)-(P3) con costos acotados y en espacios de Borel utilizando promediadores para realizar la aproximación. Se supone que el MCM satisface ciertas condiciones para que los operadores de programación dinámica T y T_z sean mapeos de contracción. Se demuestra que los promediadores son no-expansivos con respecto a las normas consideradas (proposición 3.3). Así los operadores $\widehat{T}$ y $\widetilde{T}$ para el caso descontado y los operadores $\widehat{T}_z$ y $\widetilde{T}_z$ para el caso promedio son mapeos de contracción que garantizan la convergencia en los problemas (D1) y (P1) respectivamente.

Respecto a los problemas (D2)-(D3) y (P2)-(P3), debido a que los promediadores definen una probabilidad de transición (proposición 3.1), permiten ver la acción de aproximación como una perturbación del MCM original. En términos de la precisión del promediador, se obtienen cotas de aproximación para la funcion de costo en una etapa y la ley de transición.

En la literatura los hechos anteriores han pasado desapercibidos, exepto por el artículo de Gordon [9], en el cual se refiere a la propiedad que permite ver los promediadores como un proceso de Markov como una *"propiedad intrigante",* e ignora completamente la otra propiedad. Además, Gordon no deduce cotas de error para el algoritmo IVA a partir de las propiedades del promediador; sólo deduce que sea no-expansivo.

La tesis está estructurada en siete capítulos de la siguiente manera:

El capítulo 1 contiene la formulación de un PCO*,* se presenta una breve descripción de sus principales elementos: modelo de control, políticas de control admisibles e índices de funcionamiento. Se introduce la notación y terminología que será utilizada.

El capítulo 2 es un resumen de resultados que serán nuestro punto de partidada: condiciones de compacidad y continuidad, condiciones de ergodicidad, ecuaciones de optimalidad y el algoritmo iteración de valores para los criterios descontado y promedio.

En el capítulo 3 se introducen la clase de operadores de aproximación denominados *promediadores*, se demuestra que dichos promediadores definen una probabilidad de transición, resultado clave para el logro de los objetivos de esta tesis. Se obtienen dos MCM perturbados y se definen las constantes de aproximación de estos MCM's perturbados.

En el capítulo 4 se estudia el índice en costo descontado y se resulven los problemas (D1)-(D3). Se presenta el agoritmo IVA utilizando los promediadores. Luego se obtienen las cotas de desempeño del algoritmo IVA, que constituye la primera parte central de esta tesis.

En capítulo 5 se estudia el índice en costo promedio y se resulven los problemas (P1)-(P3). Se introduce el algoritmo IVA y las suposiciones bajo las cuales este algoritmo funciona. Se obtienen las cotas de desempeño del algoritmo IVA, que es la segunda parte central esta tesis, con la excepción de las cotas para $\widetilde{h}^* - h^*$ y $\widehat{h}^* - h^*$.

En el capítulo 6 se ilustran tanto el enfoque como los resultados del presente trabajo en un ejemplo de sistemas de inventarios con demanda exponencial. Se implementa computacionalmente el sistema, tanto para costo descontado como para costo promedio y para contrastaste de resultados se simula la cadena que se obtiene usando la pólitica *greedy*. Se finaliza este capítulo con un reporte de los resultados numéricos arrojados por la implementación y la simulación de la evolución del sistema.

En el capítulo 7 se dan las conclusiones y posibles extensiones de este trabajo.

Los resultados presentados en esta tesis estan basados en los trabajos [27] y [28].

Introduction

A wide range problems in economics, operations research, renovable and not renovable resources exploitation, etc, are modeled as optimization problems of stochastic systems that evolve during the time ([4], [6], [10], [11], [12], [19], [23], [25], [30]). These problems consist in the take of sequential decisions with the aim of drive the system to an optimal behavior. Is electing the best decision the among different alternatives (called actions or controls) requires to analyze the effects of each one of them. The immediate effects usually can be seen, but the long term effects not always are transparent, and some controls can have very poor immediate responses but they can be the best at the end. So that, the problem is to choose controls that give an adequate balance between immediate and future costs (or benefits). The difficulty of decisor is the uncertainty about the future because the effects of the controls are random.

Dynamic systems that involve making sequential decisions in an environment of uncertainty and whose evolution to future states just depends on the current state and control but not of the past history, are modeled by Markov control models (MCM). An MCM consists of an state space, a control space, a transition law and a cost function. At the beginning of each stage (time between consecutive decisions), the controller observes the state of the system and choose a feasible control. By doing this, the controlled system moves to a new state accordingly to the transition law, which generates a cost. If the observation of the states and application of control is done at discrete times, it is said the MCM in discrete-time evolves. The present work this kind of models.

The optimal control problem OCP the main is to find the decision rule (policy) which specifies the best control to apply to any state of the system. Each policy defines a controlled process and a cost accumulated given by an performance index or objective function. So then, the OCP to find a policy induces an optimal behaviour accordingly to the specified performance index (or objective function). The performance indexes more used are the discounted cost and the average cost criteria.

Currently, the theory of controlled Markov processes is a well-developed theory and provides a highly flexible framework for the analysis of many sequential optimization problems in a number of areas ([4], [6], [10], [11], [12], [19], [23], [25], [30]). However, is well known that for systems with large or continuous state spaces, the controlled processes theory suffers the so-called "*curse of dimensionality*" [21], which makes impossible to obtain numerically the optimal policies in a reasonable computational time. For this reason, many researchers have proposed methods of approximation with the hope of break down or mitigate the curse of dimensionality but obtaining suboptimal policies or at least "*good*" solutions ([1], [5], [8], [9], [15], [17], [20], [21], [22], [24], [26]). Nevertheless, the major part of these works are con-

centrated in discrete spaces, mainly in the finite case and with respect to the criteria of discounted costs criterion.

The main approach to solve a OCP is to find a solution to the optimality equation, which to the discounted cost criterion takes the form

$$V = TV \tag{11}$$

where T is the dynamic programming operator. A standard method to obtain a solution of the optimality equation is the value iteration (VI) algorithm, which applies iteratively the operator T in a specified space of functions ([4], [10], [11], [12], [19]). The VI algorithm generates a sequence of functions according to the recursive equation

$$V_{n+1} = TV_n, \quad n \in \mathbb{N}_0, \tag{12}$$

Under certain conditions T is a contraction operator in some space of functions; the Banach fixed point theorem guarantees the existence of a function V^* such that $V^* = TV^*$. However, the VI algorithm has two computational obstacles: (I) it has to stop after a finite number of iterations; (II) it has to compute TV_n for each element of the state space. For continuous state spaces or large finite spaces, the computational implementation TV_n is infeasible. Thus, an alternative is to approximate TV_n and to obtain suboptimal policies with two errors, one come for stopping the algorithm after a finite number of iterations, and the other one due to the approximation step,

It has been studied a wide variety of approximation schemes, as the approximate value iteration (AVI) algorithm ([8], [17], [20]), neuro-dynamic programming [5] or reinforcement learning [26], among others. The idea is to combine suitable function approximation schemes with the standard VI algorithm.

In this work we will study the AVI algorithm, in which the approximation scheme is represented by an operator L, which is applied between two consecutive applications of T. There are two AVI algorithms depending of which operator T or L acts first. These AVI algorithms are given by the composition $\widetilde{T} = LT$ y $\widehat{T} = TL$. These define the recursive equations

$$\widetilde{V}_{n+1} = \widetilde{T}\widetilde{V}_n, \quad n \in \mathbb{N}_0, \tag{13}$$

and

$$\widehat{V}_{n+1} = \widehat{T}\widehat{V}_n, \quad n \in \mathbb{N}_0. \tag{14}$$

The properties of the operator L strongly determine the goodness of the AVI algorithms. For instance, the convergence of the successions (13) and (14) are not guaranteed unless L is no-expansive with respect to an suitable norm ([8], [9], [21]).

When an AVI algorithms, for example (13) stops at the nth iteration, it is obtained a stationary policy f_n that optimizes $\widetilde{T}\widetilde{V}_n$ (this policy f_n is called $\widetilde{V}_n - greedy$) and the optimal value function is approximated by the value function V_{f_n} induced by the policy f_n. Thus, AVI algorithm rises the following important problems:

(D1) The first one is the convergence of sequence of the functions $(\widetilde{V}_n)_{n\in\mathbb{N}}$;

(D2) Once the convergence is guaranteed, say to a function $\widetilde{V}^*$, the second problem is to find bounds for the deviation $\widetilde{V}^* - V^*$ expressed in terms of the accuracy of operator L to approximate function.

(D3) The third one, perhaps the most important from the practical viewpoint, is to find bounds for the approximation error $e(f_n) := V_{f_n} - V^*$.

Obviously, the same problems are arise by the AVI algorithm (14) but replacing $\widetilde{T}$, $\widetilde{V}_n$ and $\widetilde{V}^*$ for $\widehat{T}$, $\widehat{V}_n$ y $\widehat{V}^*$; respectively.

For the discounted cost criterion, the references [1], [17] and [24] provide interesting bounds for performance algorithm. Their approaches and results differ with those of the present work in several ways. For instance, Almudevar [1] takes an abstract approach and first studies general approximate fixed point problems; he then applies his results to several types of Markov decision problems. However, he only obtains asymptotic results and his bounds are not tied to the accuracy with which the operator L approximates the elements of the control model. On the other hand, Munos [17] gives Lp bounds expressed in terms of some quantities he calls transition probabilities concentration coefficients, which seems very difficult to compute or to bound excepting in some simple cases; moreover, as in the Almudevar's paper [1], the bounds are not expressed in terms of the accuracy of the approximation schemes. Stachurski [24] removes this latter drawback under the additional assumption that the dynamic programming or Bellman operator preserves the monotonicity of the functions, which limits the general usefulness of his results. Additionally, Almudevar [1] and Munos [17] require for the analysis of the continuous state space case the transition law of the system admits a density, which is not needed in the present work.

The OCP for the average cost criterion is more complicated from the mathematical point of view; indeed some authors consider it is not a single problem, but as a collection problems [2]. The optimality equation in this case is

$$\rho^* + h^* = Th^*, \tag{15}$$

where ρ^* is a constant, h^* belongs to a suitable function space and T is the corresponding dynamic programming operator. If the pair (ρ^*, h^*) satisfies (15) and h^* is bounded, then it is proved using standard arguments that ρ^* is the optimal average cost and the stationary policy f^* obtained by optimizing the right-hand of the optimality equation is an optimal stationary policy.

An standard method to obtain a solution to the optimality equation (15) is VI, which consists in to apply iteratively the operator T in a space of functions [2]. The VI algorithm generates a sequence of functions that approximate the solution according to the recursive equation

$$J_{n+1} = TJ_n, \quad n \in \mathbb{N}_0, \tag{16}$$

with $J_0 \equiv 0$. Then is obtained a solution (ρ^*, h^*) as limit of the sequences

$$\rho_{n+1}(z) = J_{n+1}(z) - J_n(z), \quad n \in \mathbb{N}_0,$$
$$h_{n+1}(\cdot) = J_{n+1}(\cdot) - J_{n+1}(z), \quad n \in \mathbb{N}_0,$$

where z is an arbitrary but fixed state. Using the dynamic programming operator these sequences are related as follows:

$$\rho_{n+1}(z) + h_{n+1}(\cdot) = Th_n(\cdot), \quad n \in \mathbb{N}_0,$$

with $h_0 \equiv 0$. Note that $\rho_{n+1}(z) = Th_n(z)$ for every $n \in \mathbb{N}_0$ because $h_n(z) = 0$. Then, this lead to the equivalent form of the VI algorithm given by

$$h_{n+1}(\cdot) = T_z h_n(\cdot), \quad n \in \mathbb{N}_0, \tag{17}$$

where the operator T_z is defined as

$$T_z u(\cdot) := Tu(\cdot) - Tu(z), \tag{18}$$

for functions u to suitable function space. Note that the pair (ρ^*, h^*) with $h^*(z) = 0$ satisfies the optimality equation (15) if and only if it is a fixed point of T_z. More general and recent results for the average cost criterion, can be found in [11] and [12].

As in the discounted OCP, this procedure also has two computational obstacles: (I) it has to stop after a finite number of iterations; (II) it has to compute TJ_n for each element of the state space. For continuous state spaces or large finite spaces, the computational implementation TJ_n (16) ó $T_z h_n$ (17) is infeasible, so it is impossible to obtain an exact numeric representation of the sequence $\{(\rho_n, h_n)\}_{n\in\mathbb{N}}$.

To circumvent this obstacle, in a similar form to the discounted case, we consider the following algorithms:

$$\widetilde{h}_{n+1} = \widetilde{T}_z \widetilde{h}_n := LT\widetilde{h}_n - LT\widetilde{h}_n(z), \quad n \in \mathbb{N}_0 \tag{19}$$

and

$$\widehat{h}_{n+1} = \widehat{T}_z \widehat{h}_n := TL\widehat{h}_n - TL\widehat{h}_n(z), \quad n \in \mathbb{N}_0 \tag{20}$$

where T is the dynamic programming operator and L is an approximation operator.

Thus, if the AVI algorithm (19) is stopped at the nth iteration, it is computed a stationary policy f_n that optimize $\widetilde{T}\widetilde{h}_n$ and the optimal average cost ρ^* is approximated by the average cost $J(f_n)$ incurred by the f_n. The algorithm rises the following problems:

(P1) The first is the convergence of the successions $(\widetilde{\rho}_n, \widetilde{h}_n)_{n\in\mathbb{N}}$;

(P2) Once the convergence is guaranteed, say to $(\widetilde{\rho}^*, \widetilde{h}^*)$, the second problem is to find the bound of the deviations $\widetilde{\rho}^* - \rho^*$ and $\widetilde{h}^* - h^*$ expressed in terms of the quality of the approximation given by the operator L;

(P3) The third, perhaps the most important from the practical viewpoint, is to find bounds for the error of approximation $e(f_n) := J(f_n) - \rho^*$.

The same problems are presented for the AVI algorithm (20) replacing $(\widetilde{\rho}_n, \widetilde{h}_n)$ and $(\widetilde{\rho}^*, \widetilde{h}^*)$, for $\widehat{T}$, $(\widehat{\rho}_n, \widehat{h}_n)$ and $(\widehat{\rho}^*, \widehat{h}^*)$ respectively.

The aim of this thesis is to provide methods for computing approximations to the optimal control problem with respect to the discounted cost and average cost criteria for models with Borel space and bounded costs. Specifically, obtain bounds of performance AVI algorithm for these indexes in continuous space and bounded costs.

In the present work we addresses problems (D1)-(D3) and (P1)-(P3) for controlled Markov processes with Borel space and bounded costs. The analysis is done for a class of approximation operators with the following properties: (i) constant functions are fixed points of the operator; (ii) They are a positive linear operators; (iii) They have certain continuity properties (definition 3.1). Many approximation schemes are represented by operators with such properties for instance: piecewise constant approximations, linear and multilineal interpolation and kernel averages ([9], [24]). The key result of this work is that an operator with these properties defines a transition probability on the state space of the controlled system; this has the following important consequences: (a) one can see the approximation function as the average value target functions of the transition probability; (b) the approximation step in the AVI algorithm can be thought as a perturbation of the original model.

In the literature the above facts have passed unnoticed, except for the Gordon's paper [9]. He refers to the property that allows to see the averagers as Markov process as an *"intriguing property"* . In addition, Gordon does not deduce error bounds for the AVI algorithm using the properties of averagers.

The thesis is organized in seven chapters as follows:

The chapter 1 contains the formulation of the OCP, a brief description of the control model, the admissible control policies and the performance index, notation and terminology.

The chapter 2 is a summary of results that will be our departure point: compactness and continuity conditions, ergodicity conditions, and the value iteration algorithm for the discounted and average criteria.

In Chapter 3 the class of approximation operators called averager are introduced. It is proved that averagers define transition probabilities, which is a key result for this thesis. We introduce two perturbed Markov control models and some constants the measure the accuracy of the approximations.

Chapter 4 studies problems (D1) - (D3), whereas Chapter 5 studies problems (P1) - (P3). Chapter 6 illustrates our approach whit numerical computations for an inventory systems with respect to the discounted cost and the average cost criteria.

The results presented in this thesis on the AVI algorithm are based on the work [27] and [28].

Capítulo 1

Procesos de Markov controlados

1.1 Introducción

La formulación de un problema de control óptimo (PCO) requiere de la definición de tres elementos [10]: un modelo de control, un conjunto de políticas de control admisibles y un índice de funcionamiento. En este capítulo se presentan estos elementos y se discuten los índices de funcionamiento que serán considerados.

Notación. A lo largo de esta tesis usaremos la siguiente notación:

- $\mathbb{R}$ denota al conjunto de los números reales y $\mathbb{R}_+$ al subconjunto de los reales positivos.
- $\mathbb{N}$ denota al conjunto de los números naturales y $\mathbb{N}_0$ al subconjunto de los enteros no negativos.

Para un espacio topológico $(\mathbf{S}, \tau)$ usaremos la siguiente notación:

- $\mathcal{B}(\mathbf{S})$ es la $\sigma-$álgebra de Borel de $\mathbf{S}$, es decir, la mínima $\sigma-$álgebra que contiene a τ, y a sus elementos les llamamos subconjuntos de Borel.
- Diremos que $\mathbf{S}$ es un espacio de Borel si es un subconjunto de Borel de un espacio métrico separable y completo.
- $\mathbf{M}(\mathbf{S})$ es el espacio de las funciones $u : \mathbf{S} \to \mathbb{R}$ (Borel-) medibles.
- $\mathbf{C}(\mathbf{S})$ es el espacio de las funciones $u : \mathbf{S} \to \mathbb{R}$ (Borel-) continuas.
- $\mathbf{M}_a(\mathbf{S})$ es el subespacio de $\mathbf{M}(\mathbf{S})$ de las funciones $u : \mathbf{S} \to \mathbb{R}$ acotadas.
- $\mathbf{C}_a(\mathbf{S})$ es el subespacio de $\mathbf{C}(\mathbf{S})$ formado por las funciones $u : \mathbf{S} \to \mathbb{R}$ acotadas.
- $I_B : \mathbf{S} \to \mathbb{R}$ es la función indicadora del conjunto $B \in \mathcal{B}(\mathbf{S})$, es decir,

$$I_B(x) := \begin{cases} 1 & \text{si } x \in B, \\ 0 & \text{si } x \notin B. \end{cases}$$

- $P : \mathbf{S} \to \mathbb{R}$ es medida de probabilidad.

1.2 Modelo de control markoviano

Definición 1.1 *Un modelo de control markoviano (MCM) consta de los siguientes objetos*

$$\mathfrak{M} = (\mathbf{X}, \mathbf{A}, \{\mathbf{A}(x);\ x \in \mathbf{X}\}, C, Q),$$

donde

(a) $\mathbf{X}$ *es el espacio de estados del sistema, el cual es un espacio de Borel. A los elementos de* $\mathbf{X}$ *se les denomina estados.*

(b) $\mathbf{A}$ *representa el espacio de los controles, el cual también es un espacio de Borel.*

(c) Para cada estado $x \in \mathbf{X}$ *existe un subconjunto* $\mathbf{A}(x)$ *de* $\mathbf{A}$ *no-vacío y medible, el cual representa el conjunto de controles admisibles cuando el sistema se encuentra en el estado* x*. Al conjunto*

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

se le llama conjunto de pares admisibles. Suponemos que pertenece a la $\sigma-$*álgebra de Borel del producto cartesiano* $\mathbf{X} \times \mathbf{A}$*, ó equivalentemente que contiene la grafica de una función medible* $f : \mathbf{X} \to \mathbf{A}$*.*

(d) Q *es la ley de transición, es un kernel estocástico en* $\mathbf{X}$ *dado* $\mathbf{K}$*, es decir, satisface las siguientes propiedades:*

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

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

(e) $C : \mathbf{K} \to \mathbb{R}$ *es una función medible y representa la función de costo por etapa.*

Un MCM es la representación de un sistema estocástico que es observado y está sujeto a control en un conjunto discreto de tiempos, que sin pérdida de generalidad tomaremos como el conjunto $\mathbb{N}_0$. Denotaremos por x_t al estado del sistema y por a_t al control aplicado, en ambos casos en el tiempo $t \in \mathbb{N}_0$. Nos referiremos a x_0 como el estado inicial del sistema.

El PCO consiste en modificar o influir sobre el comportamiento del proceso de estados $\{x_t\}$ por medio de selecciones adecuadas del proceso de control $\{a_t\}$, de manera que dicho comportamiento sea óptimo de acuerdo a un criterio. Heurísticamente, podemos pensar que el proceso de control se efectúa secuencialmente de la siguiente forma. Supongamos que en el tiempo $t = 0$ el decisor observa el sistema en algún estado $x \in \mathbf{X}$, es decir, $x_0 = x$, y escoge una acción o control admisible $a_0 = a \in \mathbf{A}(x)$ generando un costo $C(x, a)$. Entonces el sistema se mueve en el tiempo $t = 1$ a un nuevo estado $x_1 = x' \in \mathbf{X}$ de acuerdo a la medida (o distribución) de probabilidad $Q(\cdot \mid x, a)$, es decir,

$$\Pr(x_1 \in B \mid x_0 = x, a_0 = a) = Q(x_1 \in B \mid x, a)$$

Una vez observado el nuevo estado del sistema $x_1 = x'$, el decisor elige un nuevo control admisible $a_1 = a' \in A(x')$ con un costo $C(x', a')$ y se repite el procedimiento anterior hasta cierto tiempo finito N ó indefinidamente. A las sucesiones $\{x_n\}$ y $\{a_n\}$ generadas se les denomina proceso de estados y proceso de control, respectivamente. Si el procedimiento de observación-control termina en un tiempo finito N, se dice que el problema de control es en horizonte finito y a N se le llama horizonte de planeación; por el contrario, si el procedimiento de observación-control se repite indefinidamente se habla de un problema de control en horizonte infinito.

1.3 Políticas de control

Definición 1.2 *Para cada $t \in \mathbb{N}_0$, $\mathbf{H}_t$ representa el espacio de las historias admisible hasta el tiempo t, el cual se define como*

$$\mathbf{H}_0 := \mathbf{X}, \quad \mathbf{H}_t := \mathbf{K}^t \times \mathbf{X}, \ t \in \mathbb{N}.$$

Observe que para cada $t \in \mathbb{N}_0$, el conjunto $\mathbf{H}_t$ contiene todas las formas posibles en que el sistema puede evolucionar hasta el tiempo t, razón por la cual a sus elementos denotados genéricamente por h_t, serán referidos como t-historias; más explícitamente, cada t-historia es un vector de la forma

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

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

Definición 1.3

(a) Por Φ denotaremos a la clase de todos los kernels estocásticos φ en $\mathbf{A}$ dado $\mathbf{X}$, que satisfacen la restricción $\varphi(\mathbf{A}(x) \mid x) = 1$ para cada $x \in \mathbf{X}$;

(b) Un selector o función de decisión es una función medible $f : \mathbf{X} \to \mathbf{A}$, tal que $f(x) \in \mathbf{A}(x)$ para todo $x \in \mathbf{X}$. Denotaremos por $\mathbf{F}$ a la familia de los selectores.

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

Definición 1.4

(a) Una política de control admisible es una sucesión $\pi = \{\pi_t\}$, tal que para cada $t \in \mathbb{N}_0$, π_t es un kernel estocástico en $\mathbf{A}$ dado $\mathbf{H}_t$ que satisface la restricción

$$\pi_t(\mathbf{A}(x) \mid h_t) = 1, \ \forall h_t \in \mathbf{H}_t.$$

(b) $\pi = \{\pi_t\}$ es markoviana aleatorizada si existe una sucesión $\{\varphi_t\}$, $\varphi_t \in \Phi$ tal que

$$\pi_t(\cdot \mid h_t) = \varphi_t(\cdot \mid x_t), \ \forall h_t \in \mathbf{H}_t, \quad t \in \mathbb{N}_0.$$

(c) $\pi = \{\pi_t\}$ *es estacionaria aleatorizada (ó relajada) si existe una función* $\varphi \in \Phi$, *tal que*

$$\pi_t(\cdot \mid h_t) = \varphi(\cdot \mid x_t),\ \forall h_t \in \mathbf{H}_t, \quad t \in \mathbb{N}_0.$$

(d) $\pi = \{\pi_t\}$ *es determinista si para cada* $t \in \mathbb{N}_0$ *existe una función medible* $f_t : \mathbf{H}_t \to \mathbf{A}$ *tal que,* $\pi_t(B \mid h_t) = I_B(f_t(h_t))$, *para todo* $B \in \mathcal{B}(\mathbf{A})$, $h_t \in \mathbf{H}_t$, *donde* I_B *es la función indicadora del conjunto* B.

(e) $\pi = \{\pi_t\}$ *es markoviana determinista si existe una sucesión* $\{f_t\} \subset \mathbf{F}$, *tal que para cada* t, $\pi_t(\cdot \mid h_t)$ *está concentrada en* $f_t(x_t)$, *es decir,* $\pi_t(B \mid h_t) = I_B(f_t(x_t))$, *para todo* $B \in \mathcal{B}(\mathbf{A})$, $h_t \in \mathbf{H}_t$.

(f) π *es estacionaria determinista si existe* $f \in \mathbf{F}$ *tal que* $\pi_t(\cdot \mid h_t)$ *está concentrada en* $f(x_t)$, *es decir,* $\pi_t(B \mid h_t) = I_B(f_t(x_t))$, *para todo* $B \in \mathcal{B}(\mathbf{A})$, $h_t \in \mathbf{H}_t$ *y* $t \in \mathbb{N}_0$.

Observe que cada elemento φ en Φ define una política estacionaria aleatorizada y viceversa; de manera que abusando de la notación identificaremos a φ con tal política y nos referiremos a Φ como la clase de las políticas estacionarias aleatorizadas. De manera similar, a la familia de los selectores $\mathbf{F}$ la identificaremos con la clase de las políticas estacionarias deterministas. Para los problemas de control en horizonte infinito son importantes ambas clases de políticas estacionarias.

La familia de todas las políticas de control admisibles será denotada por Π, mientras que las subclases de las políticas; markovianas aleatorizadas por Π_{MA}, estacionarias aleatorizadas por Π_{EA}, determínistas por Π_D, markovianas determínistas por Π_{MD} y las estacionarias determínistas Π_{EM}. De acuerdo a las definiciones anteriores se tiene que $\Pi_{EA} \subset \Pi_{MA} \subset \Pi$ y $\Pi_{ED} \subset \Pi_{MD} \subset \Pi$.

Observación 1.1 *Del teorema de Ionescu-Tulcea [3, teorema 2.7.2, p.109], tenemos que para cada política* $\pi \in \Pi$ *y cada medida* $\nu \in P(\mathbf{X})$ *existe una medida de probabilidad* P_ν^π *definida sobre el espacio canónico* $(\Omega, \mathcal{F})$, *donde* $\Omega := (\mathbf{X} \times \mathbf{A})^\infty$ *y* $\mathcal{F}$ *es la* σ*-álgebra producto correspondiente que satisface las siguientes propiedades para cada* $t \in \mathbb{N}_0$*:*

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

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

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

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

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

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

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

1.4 Índices de funcionamiento

Una vez especificados el modelo de control y el conjunto de políticas de control admisibles, para plantear un PCO sólo hace falta especificar el índice ó criterio de funcionamiento con el que se evaluarán los comportamientos posibles del sistema estocástico inducidos por la aplicación de las diferentes políticas de control. Generalmente, el índice de funcionamiento es una función

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

y la función de valor optimo correspondiente se define como

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

El PCO consiste en encontrar una política $\pi^* \in \Pi$ tal que

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

en cuyo caso se dice que π^* es óptima con respecto al índice $\mathbf{I}$.

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

Específicamente, nuestro trabajo está dirigido hacia el estudio de problemas de control óptimo en horizonte infinito con espacios de Borel y costos acotados con respecto a los índices de funcionamiento descontado y promedio. Estos índices se introducen a continuación.

1.4.1 Costo descontado

Para el PCO con costo descontado en cada etapa de control o decisión, el costo generado es "actualizado" por medio de un factor de descuento $\alpha \in (0, 1)$ y el problema de control consiste en minimizar el valor esperado de la suma de los costos descontados sobre la clase de todas las políticas admisibles. Esto es, el costo α-descontado generado por el uso de la política $\pi \in \Pi$, dado que el estado inicial es

$x_0 = x \in \mathbf{X}$, esta dado por

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

mientras que

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

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

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

Entre los problemas de control, el PCO con índice en costo descontado es el mejor comprendido y para el cual existe un cuerpo teórico muy completo, el cual incluye caracterizaciones de las políticas estacionarias óptimas y algoritmos de aproximación – por ejemplo, iteración de valores, iteración de políticas (vea, [4], [10], [11])–. En este trabajo se estudia el algoritmo iteración de valores.

1.4.2 Costo promedio

Otro criterio también estándar en la literatura es el índice en costo promedio. El costo (esperado) en n-etapas, $n \in \mathbb{N}$, cuando se aplica la política $\pi \in \Pi$, dado que el estado inicial es $x_0 = x \in \mathbf{X}$, se define como

$$J_n(\pi, x) := E_x^\pi \sum_{t=0}^{n-1} C(x_t, a_t),$$

y la función de **valor óptimo (esperado) en n-etapas** ($n \in \mathbb{N}_0$) por

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

El **costo promedio** se define como

$$J_\pi(x) := \limsup_{n \to \infty} \frac{1}{n} J_n(\pi, x), \quad \pi \in \Pi, \ x \in \mathbf{X},$$

mientras que la **función de valor óptimo** en costo promedio como

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

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

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

El índice costo promedio es adecuado para evaluar el funcionamiento de sistemas controlados cuya función de costo no tiene significado económico directo, y que por otro lado, realizan un número grande de transiciones en periodos cortos de tiempo real como, por ejemplo, los sistemas de comunicación (veáse [5] y sus referencias)

de manera que en tiempos razonables se espera que la sucesión $\left\{n^{-1}J_n(\pi, x)\right\}$ ya se haya estabilizado alrededor de su valor límite, lo que a su vez permite interpretar a $J_\pi(x)$ como la tasa de crecimiento del costo generado en horizontes finitos.

Un método muy popular para resolver este problema de control es el enfoque de aproximaciones por problemas descontados. Otro método imporante es el de iteración de valores o aproximaciones sucesivas, el cual es el objeto de estudio en este trabajo.

Capítulo 2

Resultados preliminares

2.1 Introducción

Este capítulo, presenta un resumen de resultados conocidos, los cuales garantizan que los PCO en costo descontado y costo promedio están bien definidos y también garantizan la existencia de sus soluciones. Dichos resultados seran nuestro punto de partida. Se analizan aquí, las condiciones de compacidad y continuidad, condiciones de ergodicidad, ecuaciónes de optimalidad y algoritmo iteración de valores para ambos criterios.

2.2 Condiciones de optimalidad

Para garantizar que los problemas de control óptimo en costo descontado y en costo promedio están bien definidos y que existen políticas óptimas, es necesario suponer que el modelo $\mathfrak{M} = (\mathbf{X}, \mathbf{A}, \{\mathbf{A}(x);\ x \in \mathbf{X}\}, C, Q)$ satisface ciertas condiciones de compacidad y continuidad. En este trabajo consideramos dos conjuntos de estas condiciones, las cuales se enuncian a continuación.

Hipótesis 2.1

(a) $C\colon \mathbf{K} \to \mathbb{R}$ *es una función medible y acotada por una constante* $M > 0$;

(b) $\mathbf{A}(x)$ *es un suconjunto de* $\mathbf{A}$ *no-vacío y compacto para cada* $x \in \mathbf{X}$ *y el mapeo* $x \to \mathbf{A}(x)$ *es continuo;*

(c) $C\colon \mathbf{K} \to \mathbb{R}$ *es una función continua;*

(d) $Q(\cdot \mid \cdot, \cdot)$ *es débilmente continua en* $\mathbf{K}$, *esto es, el mapeo*

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

es continuo para cada función $u \in \mathbf{C}_a(\mathbf{X})$

Hipótesis 2.2

(a) $C\colon \mathbf{K} \to \mathbb{R}$ *es una función medible y acotada por una constante* $M > 0$. *Además, para cada* $x \in \mathbf{X}$, *se cumplen las siguientes condiciones:*

(b) $\mathbf{A}(x)$ *es un suconjunto de* $\mathbf{A}$ *no-vacío y compacto;*

(c) $C(x, \cdot)$ *es una función continua en* $\mathbf{A}(x)$;

(d) $Q(\cdot \mid x, \cdot)$ *es fuertemente continua en* $\mathbf{A}(x)$, *esto es, el mapeo*

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

es continuo para cada función $u \in \mathbf{M}_a(\mathbf{X})$.

Observación 2.1 *En lo que resta de este trabajo asumiremos que* $\mathbf{CM}_a(\mathbf{X})$ *denota* $\mathbf{C}_a(\mathbf{X})$ *ó* $\mathbf{M}_a(\mathbf{X})$ *dependiendo si se usa la hipótesis 2.1 ó la hipótesis 2.2, respectivamente.*

Para estudiar el PCO en costo promedio, además de las condiciones de la hipótesis 2.1 ó las condiciones de la hipótesis 2.2 se necesita otra condición, denominada condición de ergocidad, en la cual se considera lo siguiente.

Sea ($\mathbf{S}$,$\mathcal{B}$) un espacio medible. Una función de valores reales μ definida en $\mathcal{B}$, se le llama medida signada si satisface que: $\mu(\emptyset) = 0$, μ es sigma-aditiva y μ a lo más alcanza uno de los valores $+\infty$, $-\infty$. Se define la **variación total** de μ como

$$\|\mu\|_{VT} := \sup_{B \in B} \mu(B) - \inf_{B \in B} \mu(B)$$

que es una norma en el espacio vectorial de todas las medidas signadas finitas en $\mathbf{S}$. Para más detalles sobre $\|\cdot\|_{VT}$ consulte [10, Apéndice B, p.124].

A continuación enunciamos la condición de ergodicidad.

Hipótesis 2.3 *Existe un número positivo* $\beta < 1$ *tal que*

$$\|Q(\cdot \mid x, a) - Q(\cdot \mid x', a')\|_{VT} \leqslant 2\beta, \quad \forall\ (x, a),\ (x', a') \in \mathbf{K}.$$

2.3 Iteración de valores en costo descontado

Recuerde que la función de costo descontado se define como

$$V_\pi(x) := E_x^\pi \sum_{t=0}^{\infty} \alpha^t C(x_t, a_t), \quad \pi \in \Pi,\ x \in \mathbf{X}, \tag{2.1}$$

y la función de valor óptimo descontado como

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

Asimismo, una política $\pi^* \in \Pi$ es α-óptima si satisface la relación

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

En el MCM con costo descontado se considera el espacio de Banach $(\mathbf{CM}_a(\mathbf{X}), \|\cdot\|_\infty)$, donde $\|\cdot\|_\infty$ representa la norma del supremo definida como

$$\|v\|_\infty := \sup_{x \in \mathbf{X}} |v(x)| \, , \ \forall \ v \in \mathbf{CM}_a(\mathbf{X}).$$

Bajo la hipótesis 2.1 ó la hipótesis 2.2 se define el **operador de programación dinámica** en costo descontado como

$$Tu(x) := \inf_{a \in \mathbf{A}(x)} \left\{ C(x,a) + \alpha \int_{\mathbf{X}} u(y) Q(dy \mid x, a) \right\}, \ x \in \mathbf{X}. \tag{2.4}$$

Por simplicidad en la notación para cada política estacionaria $f \in \mathbf{F}$ y función v medible en $\mathbf{K}$ se define

$$v_f(x) := v(x, f(x)) \quad x \in \mathbf{X}; \tag{2.5}$$

en particular, para la función de costo y probabilidad de transición correspondiente a una política estacionaria $f \in \mathbf{F}$, se definen

$$C_f(x) := C(x, f(x)), \ x \in \mathbf{X}, \tag{2.6}$$

y

$$Q_f(\cdot \mid x) = Q(\cdot \mid x, f(x)), \ x \in \mathbf{X}. \tag{2.7}$$

Además, definimos

$$Q_f u(x) := \int_{\mathbf{X}} u(y) Q_f(dy \mid x), \quad x \in \mathbf{X}, \tag{2.8}$$

para $u \in \mathbf{M}(\mathbf{X})$, siempre que la integral esté bien definida.

Además, para cada $f \in \mathbf{F}$, definimos el operador $T_f : \mathbf{M}_a(\mathbf{X}) \to \mathbf{M}_a(\mathbf{X})$ como

$$T_f u(x) := C_f(x) + \alpha \int_{\mathbf{X}} u(y) Q_f(dy \mid x) \ \ \forall \ x \in \mathbf{X}, \ v \in \mathbf{CM}_a(\mathbf{X}). \tag{2.9}$$

Con la notación anterior tenemos que

$$T_f u = C_f + \alpha Q_f u, f \in \mathbf{F}.$$

Observación 2.2 *Supongamos que se satisface la hipótesis 2.1 ó la hipótesis 2.2. Entonces, el operador T es un operador de contracción del espacio $(\mathbf{CM}_a(\mathbf{X}), \|\cdot\|_\infty)$ con módulo α (consulte [10, lema 2.5, p.20]), es decir,*

$$\|Tu - Tv\|_\infty \leq \alpha \|u - v\|_\infty \ \ \forall \ u, v \in \mathbf{CM}_a(\mathbf{X})$$

Además, por un teorema de selección medible (consulte [10, proposición D.3, p.130], para cada $u \in \mathbf{CM}_a(\mathbf{X})$ *existe un selector* $f_u \in \mathbf{F}$ *tal que,* $Tu = T_{f_u}u$*, es decir,*

$$Tu(x) := C(x, f_u(x)) + \alpha \int_{\mathbf{X}} u(y)Q(dy \mid x, f_u(x)),\ \forall\ x \in \mathbf{X}.$$

Al selector f_u *se le denomina política* $u-$*greedy.*

Observación 2.3 *El operador* T_f, $f \in \mathbf{F}$, *es un operador de contracción del espacio* $(\mathbf{CM}_a(\mathbf{X}), \|\cdot\|_\infty)$ *con módulo* α*, es decir,*

$$\|T_f u - T_f v\|_\infty \leq \alpha \|u - v\|_\infty\ \ \forall\ u, v \in \mathbf{CM}_a(\mathbf{X}).$$

El teorema de punto fijo de Banach y argumentos estándares de programación dinámica nos llevan al siguiente resultado (vease [10, teorema 2.2, p.19 y teorema 2.8, p.23]).

Teorema 2.1 *Supongamos que la hipótesis 2.1 ó la hipótesis 2.2 se cumple. Entonces:*

(a) la función valor óptimo descontado V^* *es el único punto fijo de* T *en* $\mathbf{CM}_a(\mathbf{X})$*, es decir, satisface la ecuación de optimalidad*

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

(b) una política estacionaria $f \in \mathbf{F}$ *es óptima si y sólo si* $V^* = T_f V^*$;

(c) existe una política estacionaria $f^* \in \mathbf{F}$ *tal que* $V^* = T_{f^*}V^*$*; por lo tanto,* f^* *es óptima;*

(d) $\|T^n u - V^*\|_\infty \to 0$ *geométricamente para cualquier* $u \in \mathbf{CM}_a(\mathbf{X})$.

Desde el punto de vista teórico, el teorema 2.1 (a)-(c) da solución al PCO para costo descontado. Sin embargo, para obtener una política estacionaria óptima se requiere que la función valor óptimo V^* sea conocida, lo cual ocurre en muy pocos casos. Por otra parte, el teorema 2.1 (d) proporciona un procedimiento recursivo para aproximarla. Fijemos una función $V_0 \equiv v_0 \in \mathbf{CM}_a(\mathbf{X})$ arbitraria y definamos

$$V_{k+1} = TV_k, \quad k \in \mathbb{N}_0. \tag{2.10}$$

A este procedimiento se le conoce como algoritmo iteración de valores (IV) o de aproximaciones sucesivas.

Puesto que el procedimiento recursivo (2.10) no se puede repetir indefinidamente, el algoritmo de IV se debe parar en un tiempo finito de acuerdo a alguna regla de paro establecida con anterioridad. Si la regla prescribe parar en la k-esima iteración, se obtiene la política estacionaria "f-*greedy*"con respecto a V_k y se aproxima a V^* con el costo descontado V_f.

A continuación se proporciona el pseudo-codigo para el algoritmo.

Algoritmo 2.1 *Obtiene una política $V_k-greedy$ y necesita como entrada una funcion $V_0 \in \mathbf{CM}_a(\mathbf{X})$ y el valor de la tolerancia de paro denotado como Tol.*

Paso 0

$V_0 \leftarrow$ *función dada*;

$Tol \leftarrow$ *valor dado*;

$k \leftarrow 0$;

Paso 1

Re*petir*

$k \leftarrow k+1$;

Calcular TV_{k-1};

$V_k \leftarrow TV_{k-1}$;

$ToleParo \leftarrow \|V_k - V_{k-1}\|_\infty$;

Hasta que ToleParo sea menor que Tol.

Paso 2

Obtener la política $V_k-greedy$;

Terminar.

El siguiente resultado da una cota del error cometido por parar el algoritmo en la k-esima iteración.

Teorema 2.2 *Supongamos que la hipótesis 2.1 ó la hipótesis 2.2 se satisface y sea $f \in \mathbf{F}$ una politica V_k $-greedy$, es decir $TV_k = T_f V_k$, entonces*

$$\|V^* - V_f\|_\infty \leq \frac{2\alpha}{1-\alpha} \|V_k - V_{k-1}\|_\infty . \tag{2.11}$$

La demostración de este teorema utiliza argumentos elementales y puede consultarse en [25, teorema 11.2.1, p.299].

Como ya se menciono anteriormente, el algoritmo IV da una solución aproximada al PCO con cota de error (2.11) cuando TV_k es computable. Sin embargo, si el espacio de estados es continuo o finito pero muy grande, TV_k no es computable ya que se requiere obtener su valor para cada elemento del espacio de estados. Este problema es uno de los motivos de nuestro trabajo.

2.4 Iteración de valores costo promedio

El costo esperado en n-etapas, $n \in \mathbb{N}$, cuando se aplica la política $\pi \in \Pi$ dado que el estado inicial es $x_0 = x \in \mathbf{X}$, se definió como

$$J_n(\pi, x) := E_x^\pi \sum_{t=0}^{n-1} C(x_t, a_t) \tag{2.12}$$

y la función valor óptimo en n-etapas ($n \in \mathbb{N}_0$) por

$$J_n^*(x) := \inf_{\pi \in \Pi} J_n(\pi, x), \quad \pi \in \Pi, \ x \in \mathbf{X}. \tag{2.13}$$

Por otra parte, la función de valor en costo promedio está definida como

$$J_\pi(x) := \limsup_{n \to \infty} \frac{1}{n} J_n(\pi, x), \quad \pi \in \Pi, \ x \in \mathbf{X} \tag{2.14}$$

y la función valor óptimo correspondiente por

$$J^*(x) := \inf_{\pi \in \Pi} J_\pi(x), \ x \in \mathbf{X}. \tag{2.15}$$

Además, el PCO en costo promedio consiste en encontrar una política $\pi^* \in \Pi$ tal que

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

Si tal política existe se le denomina política óptima en costo promedio.

Recuerde que para cada política estacionaria $f \in \mathbf{F}$ adoptamos la notación C_f para la función de costo (2.6) y Q_f para probabilidad de transición (2.7). Además denotaremos Q_f^n a la **probabidad de transición en $n-$pasos**.

Para cada política estacionariar $f \in \mathbf{F}$ se define para el PCO en costo promedio los operadores

$$T_f u(x) := C_f(x) + \int_{\mathbf{X}} u(y) Q_f(dy \mid x) \ \ \forall \ x \in \mathbf{X},$$

y el **operador de programación dinámica**

$$Tu(x) := \inf_{a \in \mathbf{A}(x)} \left\{ C(x, a) + \int_{\mathbf{X}} u(y) Q(dy \mid x, a) \right\} \ \ \forall \ x \in \mathbf{X}. \tag{2.16}$$

Obseve que los operadores T y T_f son los mismos que los de costo descontado con $\alpha = 1$. Por lo que de igual manera bajo la hipótesis 2.1 ó la hipótesis 2.2, estos operadores T y T_f mapean el espacio $\mathbf{CM}_a(\mathbf{X})$ en sí mismo y además para cada $u \in \mathbf{CM}_a(\mathbf{X})$ existe un selector $f_u \in \mathbf{F}$ (consulte [10, D.3, p.130]), tal que, $Tu = T_{f_u} u$, es decir,

$$Tu(x) = T_{f_u} u(x) := C_{f_u}(x) + \int_{\mathbf{X}} u(y) Q_{f_u}(dy \mid x), \ \ \forall \ x \in \mathbf{X}.$$

Al selector f_u se le denomina política $u-greedy$.

Utilizando la notación (2.8) la definición del operador T_f se simplifica a

$$T_f u = C_f + \alpha Q_f u, \ \ \forall \ f \in \mathbf{F}.$$

La ecuación de optimalidad en costo promedio es de la forma

$$\rho^* + h^*(x) = Th^*(x) \ \forall x \in \mathbf{X}, \tag{2.17}$$

donde ρ^* es una constante y h^* es función medible en $\mathbf{X}$. Si existe ρ^* y $h^* \in \mathbf{CM}_a(\mathbf{X})$ solución de (2.17) entonces existe una política $f^* \in \mathbf{F}$, denominada política $h^*-greedy$, esto es,

$$\rho^* + h^* = Th^* = T_{f^*}h^*$$

y a la terna (ρ^*, h^*, f^*) se le llama **terna canónica**. Entonces, con argumentos estándar de programación dinámica se demuestra que la política estacionaria f^* es una política óptima y que la constante ρ^* es la función valor óptimo, es decir,

$$J^*(\cdot) = J(f^*, \cdot) = \rho^*.$$

La existencia de solución de la ecuación de optimalidad está garantizada bajo la hipótesis 2.1 ó hipótesis 2.2 y la condición de ergocidad hipótesis 2.3 (vease [10, Corolario 3.6, p.61]).

Otra de las implicaciones importantes de la condición de ergocidad (hipótesis 2.3) es la siguiente. Para su demostración referimos al lector a los trabajos [7], [10, Ch.3], [13], [14], [16, Ch.16].

Observación 2.4 *La hipótesis 2.3 implica que la cadena de Markov $\{x_n\}$ inducida para cada política estacionaria $f \in \mathbf{F}$ es **uniformemente ergódica**, lo cual significa que*

$$\sup_{x\in\mathbf{X}} \left\| Q_f^n(\cdot \mid x) - \mu_f(\cdot) \right\|_{VT} \leqslant 2\beta^n, \tag{2.18}$$

*donde μ_f es la única medida de **probabilidad invariante** de la probabilidad de transición Q_f, es decir,*

$$\mu_f(B) = \int_{\mathbf{X}} Q_f(B \mid x)\mu_f(dy) \ \forall B \in \mathcal{B}(\mathbf{X}).$$

La propiedad (2.18) implica que

$$\sup_{x\in\mathbf{X}} \left| E_x^f Q_f^n u(x_n) - \mu_f(u) \right| \leq 2\beta^n \left\| u \right\|_\infty \quad \forall\ f \in \mathbf{F},\ u \in \mathbf{M}_a(\mathbf{X}), \tag{2.19}$$

donde

$$\mu_f(u) := \int_{\mathbf{X}} u(y)\mu_f(dy), \quad f \in \mathbf{F},\ u \in \mathbf{M}_a(\mathbf{X}).$$

Lo anterior implica que

$$\lim_{n\to\infty} \frac{1}{n} E_x^f \sum_{k=0}^{n} u(x_k) = \mu_f(u) \quad \forall\ x \in \mathbf{X},\ f \in \mathbf{F},\ u \in \mathbf{M}_a(\mathbf{X}). \tag{2.20}$$

En particular

$$J(f) := J_f(x) = \mu_f(C_f) \ \ \forall\ x \in \mathbf{X},\ f \in \mathbf{F}; \tag{2.21}$$

es decir, el costo promedio $J_f(x)$ es constante (consulte [10, lema 3.3 (b), p.57] y [16, teorema 16.0.2, p.384]).

En la prueba de los lemas 5.3 y 5.4, usamos el siguiente resultado de Nowak [18].

Teorema 2.3 *Sea $S(\cdot \mid \cdot)$ y $T(\cdot \mid \cdot)$ probabilidades de transición en $\boldsymbol{X}$. Se definen*

$$\theta := \sup_{x \in \mathbf{X}} \|S(\cdot \mid x) - T(\cdot \mid x)\|_{VT}$$

y

$$\gamma := \frac{1}{2} \sup_{x,y \in \mathbf{X}} \|S(\cdot \mid x) - S(\cdot \mid y)\|_{VT}\,.$$

Si la cadena de Markov con la probabilidad de transición T es uniformemente ergodica y $\gamma < 1$, entonces

$$\|\pi_S - \pi_T\|_{VT} \leqslant \frac{\theta}{1-\gamma},$$

donde π_S y π_T son medidas de probabilidad invariantes para S y T, respectivamente.

Por otra parte, debido a que en la solución de la ecuación de optimalidad (2.17), la función h^* es única salvo constantes aditivas, es necesario introducir el subespacio $\mathbf{M}_a^0(\mathbf{X})$ de las funciones $u \in \mathbf{M}_a(\mathbf{X})$ que satisfacen la condición $u(z) = 0$ donde $z \in \mathbf{X}$ es arbitrario pero fijo. Similarmente, se define el subespacio $\mathbf{C}_a^0(\mathbf{X})$. También asumiremos que $\mathbf{CM}_a^0(\mathbf{X})$ denota a $\mathbf{C}_a^0(\mathbf{X})$ ó $\mathbf{M}_a^0(\mathbf{X})$ dependiendo si se usa la hipótesis 2.1 ó la hipótesis 2.2, respectivamente.

Usualmente $\mathbf{CM}_a(\mathbf{X})$ es dotado con la norma del supremo $\|\cdot\|_\infty$, pero para el PCO en costo promedio, también consideramos la semi-norma

$$\|u\|_{sp} := \sup_{x \in \mathbf{X}} u(x) - \inf_{x \in \mathbf{X}} u(x), \quad \forall\, u \in \mathbf{CM}_a(\mathbf{X}). \tag{2.22}$$

Note que $\|\cdot\|_{sp}$ al ser restringida a $\mathbf{CM}_a^0(\mathbf{X})$ es una norma, por lo que $(\mathbf{CM}_a^0(\mathbf{X}), \|\cdot\|_{sp})$ es un espacio de Banach. Además se cumple la relación

$$\|u\|_\infty \leqslant \|u\|_{sp}\,, \quad u \in \mathbf{CM}_a^0(\mathbf{X}).$$

Sea (ρ^*, h^*) la solución de la ecuación de optimalidad (2.17) con $h^*(z) = 0$. En este caso, $\rho^* = Th^*(z)$ y la ecuación de optimalidad puede ser re-escrita como

$$h^*(x) = Th^*(x) - Th^*(z), \quad x \in \mathbf{X}.$$

Para simplificar la notación se define el operador

$$T_z u(x) := Tu(x) - Tu(z), \quad \forall\, x \in \mathbf{X}. \tag{2.23}$$

Note que T_z mapea el subespacio $\mathbf{CM}_a^0(\mathbf{X})$ en si mismo y que la función $h^* \in \mathbf{CM}_a^0(\mathbf{X})$ resuelve la ecuación de optimalidad si y sólo si, esta es un punto fijo de T_z, es decir,

$$h^*(x) = T_z h^*(x) = Th^*(x) - Th^*(z) \quad \forall\, x \in \mathbf{X}.$$

Por otro lado, se sabe por un resultado estandar de programación dinámica (vease [10]) que bajo la hipótesis 2.1 ó la hipótesis 2.2, la función valor óptimo en n-etapas (2.13) satisfacen la ecuación recursiva, es decir,

$$J_{n+1}^* = TJ_n^* \ \ \forall\ n \in \mathbb{N}_0, \tag{2.24}$$

donde $J_0^* \equiv 0$.

Esto último implica que

$$\rho_{n+1}(z) + h_{n+1} = Th_n, \tag{2.25}$$

donde

$$h_n = J_n^* - J_n^*(z), \quad n \in \mathbb{N}_0,$$

y

$$\rho_n = J_n^* - J_{n-1}^*, \quad n \in \mathbb{N}.$$

Entonces, se dice que el algoritmo IV converge si la sucesión $\{(\rho_n(z), h_n)\}_{n\in\mathbb{N}}$ converge a la solución (ρ^*, h^*) de la ecuación de optimalidad (2.17).

Ahora, observe que usando el operador T_z, la ecuación recursiva (2.25) se convierte en

$$h_{n+1} = T_z h_n = T_z^n h_0 \ \ \forall\ n \in \mathbb{N}_0, \tag{2.26}$$

con $h_0 = 0$, proporcionando otra forma equivalente del algoritmo de iteración de valores.

Observación 2.5 *Suponga que la hipótesis 2.3 y una de las hipótesis 2.1 ó 2.2 se cumplen. Entonces:*

(a) El operador T_z es una contracción del espacio de Banach $(\mathbf{CM}_a^0(\mathbf{X}), \|\cdot\|_{sp})$ en sí mismo con módulo β (vease [10, lema 3.5, p.59]), donde β es la constante de la condición de ergocidad en la hipótesis 2.3. Entonces, del teorema de punto fijo de Banach, existe una terna canónica (ρ^*, h^*, f^*) con $h^* \in \mathbf{CM}_a(\mathbf{X})$ y $\rho^* = h^*(z)$.

(b) Además,

$$\|T_z^n u - h^*\|_\infty \leqslant \|T_z^n u - h^*\|_{sp} \leqslant \beta^n \|h^*\|_{sp} \to 0,$$

para todo $u \in \mathbf{CM}_a(\mathbf{X})$. En particular, tomando $u \equiv 0$, tenemos

$$\|h_n - h^*\|_\infty \leqslant \|h_n - h^*\|_{sp} \leqslant \beta^n \|h^*\|_{sp} \to 0.$$

Para completar la prueba de la convergencia del algoritmo IV, sólo resta establecer la convergencia de la sucesión $\{\rho_n(z)\}_{n\in\mathbb{N}}$ a la constante ρ^*. Esto se afirma en el siguiente teorema cuya demostración puede consultarse en [10, teorema 4.8, p. 64].

Teorema 2.4 *Suponga que la hipótesis 2.3 y una de las hipótesis 2.1 ó 2.2 se cumplen. Sea h^* la función en la observación 2.5 y defina las sucesiones*

$$s_n := \inf_{x\in\mathbf{X}} \rho_n(x) \quad y \quad S_n := \sup_{x\in\mathbf{X}} \rho_n(x), \quad n \in \mathbb{N},$$

entonces:

(a) la sucesión $\{s_n\}$ es no-decreciente y $\{S_n\}$ es no-creciente. Además

$$-\beta^{n-1} \|h^*\|_{sp} \leqslant s_n - \rho^* \leqslant S_n - \rho^* \leqslant \beta^{n-1} \|h^*\|_{sp} \quad \forall\ n \in \mathbb{N},$$

por lo tanto, $\rho_n(z)$ también converge a ρ^*,

(b) si f es una política $h_n-greedy$, entonces

$$0 \leqslant J(f) - \rho^* \leqslant \|\rho_n\|_{sp} \leqslant \beta^{n-1} \|h^*\|_{sp} \quad \forall\ n \in \mathbb{N}.$$

Observación 2.6 *Debido a que*

$$\|h_n - h_{n-1}\|_{sp} = \|J_n - J_{n-1}\|_{sp} = \|\rho_n\|_{sp},$$

es indistinto trabajar con el algoritmo IV *(2.24) del operador T ó con el algoritmo* IV *(2.26) del operador T_z. Por simplicidad trabajaremos con el primero.*

Cuando se satisface alguna regla de paro, el algoritmo IV obtiene una política $f \in \mathbf{F}$, $J_n-greedy$ y el valor óptimo ρ^* se aproxima por medio de la función valor $J(f)$, con una cota del error de aproximación $\|\rho_n\|_{sp}$. A continuación se presenta el seudocodigo de este algoritmo.

Algoritmo 2.2 *Obtiene una política $J_n-greedy$, iniciando con la función constante $J_0 \equiv 0 \in \mathbf{CM}_a^0(\mathbf{X})$ y el valor de la tolerancia de paro Tol.*

Paso 0

$J_0 \equiv 0$;

$Tol \leftarrow$ *valor dado*;

$n \leftarrow 0$;

Paso 1

Re*petir*

$n \leftarrow n + 1$;

$J_n \leftarrow TJ_{n-1}$;

$\rho_n \leftarrow J_n - J_{n-1}$;

$s_n \leftarrow \inf_{x\in\mathbf{X}} \rho_n(x)$;

$S_n :\leftarrow \sup_{x\in\mathbf{X}} \rho_n(x)$;

$\|\rho_n\|_{sp} \leftarrow S_n - s_n$;

Hasta que $\|\rho_n\|_{sp}$ *sea menor que* Tol.

Paso 2

Obtener la política $J_n-greedy$;

Terminar.

En teoría, el algoritmo IV para el PCO en costo promedio bajo las suposiciones de las hipótesis 2.1 ó hipótesis 2.2 e hipótesis 2.3, debido al error cometido por parar el algoritmo en la k-esima iteración, obtiene una solución aproximada del PCO siempre y cuando TJ_n $(T_z h_n)$ sea computable en cada iteración. Al igual que para costo descontado, desafortunadamente esto es imposible cuando el espacio de estados es continuo o finito muy grande. Este es otro motivo de nuestro trabajo.

Capítulo 3

Operadores de aproximación

3.1 Introducción

En este capítulo se introduce la clase de operadores de aproximación que nosotros estamos considerando, se demuestra que dichos operadores definen una probabilidad de transición lo cual es fundamental en este trabajo y se demuestra que tienen una propiedad por la cual los denominaremos promediadores. Se introducen dos tipos MCM perturbados asociados a los operadores de esta clase y se finaliza este capítulo definiendo las constantes de aproximación de los MCM perturbados.

3.2 Promediadores y ejemplos

Definición 3.1 *Un operador* L:$\mathbf{M}_a(\mathbf{X})\to\mathbf{M}_a(\mathbf{X})$ *es un* ***promediador*** *si satisface las siguientes condiciones:*

(a) $LI_{\mathbf{X}}(x) = I_{\mathbf{X}}(x)$*, donde* $I_B(x)$ *denota la función indicadora de cualquier subconjunto* $B \in \mathcal{B}(\mathbf{X})$*;*

(b) L *es un operador lineal;*

(c) L *es un operador positivo, es decir,* $Lv \geqslant 0$ *para cada* $v \geqslant 0$ *en* $\mathbf{M}_a(\mathbf{X})$*;*

(d) L *satisface la siguiente propiedad de continuidad:*

$$v_n \downarrow 0\,,\ v_n \in \mathbf{M}_a(\mathbf{X}) \Rightarrow Lv_n \downarrow 0.$$

La Definición 3.1 puede parecer algo restrictiva a primera vista, pero algunos esquemas de aproximación de uso común dan lugar a operadores con estas propiedades, como veremos a continuación.

Para el análisis de algunos ejemplos unidimensionales, considérese $\mathbf{X} = [0, \theta]$ y una sucesión de puntos en $\mathbf{X}$, tales que $x_1 = 0 < x_2 < \cdots < x_k = \theta$ que generan la malla $\{x_i\}_{i=1}^{k}$ y la partición

$$\{D_i : D_1 = [x_1, x_2] \text{ y } D_i = (x_i, x_{i+1}]\,,\quad i = 2, 3, \cdots, k-1\}\,.$$

1. Aproximador constante por pedazos. Sea $v \in \mathbf{M}_a(\mathbf{X})$ y sea $y_i \in D_i$, se define como
$$Lv(x) = v(y_i), \quad x \in D_i.$$
A continuación verificaremos que L es un promediador, es decir, que se cumplen las condiciones de la definición 3.1:

   (a) Sea $x \in \mathbf{X}$ y supongase que $x \in D_i$. Entonces,
   $$LI_{\mathbf{X}}(x) = I_{\mathbf{X}}(y_i) = 1 = I_{\mathbf{X}}(x),$$
   por lo que L cumple la primera condición.

   (b) Sean $u, v \in \mathbf{M}_a(\mathbf{X})$ y a, b, escalares reales. Considere $x \in \mathbf{X}$ y sin pérdida de generalidad supongase que $x \in D_i$. Entonces,
   $$\begin{aligned} L(au + bv)(x) &= (au + bv)(y_i) \\ &= au(y_i) + bv(y_i) \\ &= aLu(x) + bLv(x), \end{aligned}$$
   por lo que L cumple la condición de linealidad.

   (c) Si $v \geq 0$ y sea $x \in \mathbf{X}$, tal que $x \in D_i$. Entonces,
   $$Lv(x) = v(y_i) \geq 0.$$
   Como esto sucede para cualquier x, entonces L es un operador positivo.

   (d) Sea $\{v_n\}_{n=1}^{\infty}$, tal que $v_n \in \mathbf{M}_a(\mathbf{X})$ y $v_n \downarrow 0$, considere un $x \in \mathbf{X}$ cualquiera y sin pérdida de generalidad supongase que $x \in D_i$. Entonces,
   $$Lv_n(x) = v_n(y_i) \downarrow 0,$$
   por lo que L satisface la condición de continuidad.

   Por lo todo lo anterior, el aproximador constante por pedazos es un promediador.

2. Aproximador de interpolación lineal. Para $v \in \mathbf{M}_a(\mathbf{X})$, el aproximador se define como
$$Lv(x) = b_i(x)v(x_i) + \overline{b_i}(x)v(x_{i+1}), \quad x \in D_i, \quad i = 1, 2, \cdots, k-1, \tag{3.1}$$
   donde
$$b_i(x) = (x_{i+1} - x)/(x_{i+1} - x_i), \quad x \in D_i, \quad i = 1, 2, \cdots, k-1$$
   y
$$\overline{b_i}(x) = 1 - b_i(x)$$

(a) Supongase que $x \in D_i$. Puesto que

$$\begin{aligned} LI_{\mathbf{X}}(x) &= b_i(x)I_{\mathbf{X}}(x_i) + \overline{b_i}(x)I_{\mathbf{X}}(x_{i+1}) \\ &= b_i(x)I_{\mathbf{X}}(x) + \overline{b_i}(x)I_{\mathbf{X}}(x) \\ &= I_{\mathbf{X}}(x), \end{aligned}$$

por lo que L cumple la primera condición de la definición de promediador.

(b) Sean $u, v \in \mathbf{M}_a(\mathbf{X})$, a, b, escalares reales. Considere $x \in \mathbf{X}$ y sin pérdida de generalidad supongase que $x \in D_i$. Entonces

$$\begin{aligned} L(au + bv)(x) &= b_i(x)(au + bv)(x_i) + \overline{b_i}(x)(au + bv)(x_{i+1}) \\ &= b_i(x)\left[au(x_i) + bv(x_i)\right] + \overline{b_i}(x)\left[au(x_{i+1}) + bv(x_{i+1})\right] \\ &= a\left[b_i(x)u(x_i) + \overline{b_i}(x)u(x_{i+1})\right] + b\left[b_i(x)v(x_i) + \overline{b_i}(x)v(x_{i+1})\right] \\ &= aLu(x) + bLv(x). \end{aligned}$$

L cumple la condición de linealidad.

(c) Sean $v \geq 0$ y $x \in \mathbf{X}$, tal que $x \in D_i$, entonces

$$Lv(x) = b_i(x)v(x_i) + \overline{b_i}(x)v(x_{i+1}) \geq 0,$$

ya que $0 \leq b_i(x) \leq 1$, $0 \leq \overline{b_i}(x) \leq 1$ y $v(x_i) \geq 0 \ \ \forall i = 1, 2, \cdots, k$. Entonces L es un operador positivo.

(d) Sea $\{v_n\}_{n=1}^{\infty}$, tal que $v_n \in \mathbf{M}_a(\mathbf{X})$ y $v_n \downarrow 0$, considere un $x \in \mathbf{X}$ cualquiera y sin pérdida de generalidad supongase que $x \in D_i$, entonces

$$Lv_n(x) = b_i(x)v_n(x_i) + \overline{b_i}(x)v_n(x_{i+1}) \downarrow 0,$$

ya que $v(x_i) \downarrow 0$ para todo $i = 1, 2, \cdots, k$, por lo que L satisface la condición de continuidad.

Por lo anterior, el interpolador lineal es un promediador.

3. Aproximador tipo kernel. Defina

$$Lv(x) = \sum_{i=1}^{k} \lambda(x, i)v(x_i),$$

donde

$$\lambda(x, i) = \frac{K_h(x_i - x)}{\sum_{j=1}^{k} K_h(x_j - x)},$$

$K_h : \mathbf{X} \to \mathbb{R}_+$ es un mapeo no negativo denominado kernel. Observe que $\lambda(x, i) \geq 0$, para todo $i = 1, 2, \cdots, k$ y todo $x \in \mathbf{X}$. Note, además que

$$\sum_{i=1}^{k} \lambda(x, i) = 1,$$

por lo que $Lv(x)$ es una combinación convexa de las observaciones $\{v(x_i)\}_{i=1}^k$. Un ejemplo de kernel es la función

$$K_h(x) = e^{-|x|/h},\ x \in \mathbb{R}.$$

Note que el valor del kernel decae a cero cuando x se aleja de los x_i's. El parámetro h es de suavizamiento y controla el peso asignado a la observación más distante. Se puede observar además que $Lv(x)$ asigna pesos grandes aquellas observaciones $v(x_i)$ cuyos x_i estan cercanos a x.

A continuación se analizan las condiciones para determinar si este aproximador es promediador:

(a) Sea $x \in \mathbf{X}$ y sin pérdida de generalidad supongase que $x \in D_i$. Entonces

$$LI_{\mathbf{X}}(x) = \sum_{i=1}^{k} \lambda(x,i) I_{\mathbf{X}}(x_i) = I_{\mathbf{X}}(x) \sum_{i=1}^{k} \lambda(x,i) = I_{\mathbf{X}}(x),$$

por lo que L cumple la primera condición para ser promediador.

(b) Sean $u, v \in \mathbf{M}_a(\mathbf{X})$ y a, b, escalares reales, considere $x \in \mathbf{X}$ y sin pérdida de generalidad supongase que $x \in D_i$. Entonces

$$\begin{aligned}
L(au+bv)(x) &= \sum_{i=1}^{k} \lambda(x,i)(au+bv)(x_i) \\
&= \sum_{i=1}^{k} \lambda(x,i)\left[au(x_i) + bv(x_i)\right] \\
&= \sum_{i=1}^{k} \left[a\lambda(x,i)u(x_i) + b\lambda(x,i)v(x_i)\right] \\
&= \sum_{i=1}^{k} \left[a\lambda(x,i)u(x_i) + b\lambda(x,i)v(x_i)\right] \\
&= a\sum_{i=1}^{k} \lambda(x,i)u(x_i) + b\sum_{i=1}^{k} \lambda(x,i)v(x_i) \\
&= aLu(x) + bLv(x),
\end{aligned}$$

lo cual prueba la linealidad de L.

(c) Sean $v \geq 0$ y $x \in \mathbf{X}$, tal que $x \in D_i$. Entonces

$$Lv(x) = \sum_{i=1}^{k} \lambda(x,i) v(x_i) \geq 0,$$

ya que $\lambda(x,i) \geq 0$ y $v(x_i) \geq 0$, para todo $i = 1, 2, \cdots, k$. Entonces L es un operador positivo.

(d) Sea $\{v_n\}_{n=1}^{\infty}$, tal que $v_n \in \mathbf{M}_a(\mathbf{X})$ y $v_n \downarrow 0$. Considere un $x \in \mathbf{X}$ cualquiera y sin pérdida de generalidad supongase que $x \in D_i$. Entonces,

$$Lv_n(x) = \sum_{i=1}^{k} \lambda(x,i) v_n(x_i) \downarrow 0,$$

ya que $v(x_i) \downarrow 0$, $\forall\ i = 1, 2, \cdots, k$, por lo que L satisface la condición de continuidad.

Por lo tanto, el aproximador tipo kernel es un promediador.

Observación 3.1 *Siguiendo pasos análogos se puede demostrar que el aproximador de interpolación multilineal también es un promediador.*

Definición 3.2 *Un promediador L es **continuo** si mapea funciones continuas en funciones continuas, es decir, si L:$\mathbf{C}_a(\mathbf{X}) \to \mathbf{C}_a(\mathbf{X})$.*

Por ejemplo, todos los promediadores anteriores son continuos exepto el aproximador constante por pedazos.

3.3 Propiedades de los promediadores

En el desarrollo del resto de este trabajo usaremos las siguientes propiedades de los promediadores.

Proposición 3.1 *Sea L un promediador. Defina el mapeo $L(\cdot \mid \cdot)$: $\mathcal{B}(\mathbf{X}) \times \mathbf{X} \to [0,1]$, como*

$$L(B \mid x) := LI_B(x), \quad x \in X, \quad B \in \mathcal{B}(\mathbf{X}). \tag{3.2}$$

Entonces, $L(\cdot \mid \cdot)$ es una probabilidad de transición en $\mathbf{X}$, esto es, $L(\cdot \mid x)$ es una medida de probabilidad en $\mathbf{X}$ para cada $x \in \mathbf{X}$, y $L(B \mid \cdot)$ es una función medible para cada $B \in \mathcal{B}(\mathbf{X})$.

Demostración.

(a) Se demostrará primero que $L(\cdot \mid x)$ es una medida de probabilidad en $\mathbf{X}$ para cada $x \in \mathbf{X}$.

(i) $L(\mathbf{X} \mid x) := LI_{\mathbf{X}}(x) = I_{\mathbf{X}}(x) \equiv 1$ para todo $x \in \mathbf{X}$.

(ii) $L(\cdot \mid x) \geqslant 0$, por la definición 3.1 (c).

(iii) Sean $\{B_i\}_{i=1}^n$, subconjuntos ajenos de $\mathcal{B}(\mathbf{X})$. Entonces,

$$\begin{aligned} L(\cup_{i=1}^n B_i \mid x) &:= LI_{\cup_{ii=1}^n B_i}(x) \\ &= L\left[\sum_{i=1}^n I_{B_i}(x)\right] \\ &= \sum_{i=1}^n \left[LI_{B_i}(x)\right], \quad \text{(por la definición 3.1 (b))} \\ &= \sum_{i=1}^n L(B_i \mid x), \quad \text{(por (3.2))}, \end{aligned}$$

por lo que $L(\cdot \mid x)$ es finitamente aditiva.

Para ver que es numerablemente aditiva, considerese una sucesón $\{B_i\}_{i=1}^{\infty}$, de subconjuntos de $\mathcal{B}(\mathbf{X})$ tales que $B_i \downarrow \emptyset$. Entonces,

$$L(B_i \mid x) = LI_{B_i}(x) \downarrow LI_{\emptyset}(x) = 0,$$

por las Definiciones 3.1(d) y 3.1 (a). Por lo tanto, $L(\cdot \mid x)$ es numerablemente aditiva.

Se concluye que $L(\cdot \mid x)$ es una medida de probabilidad en $\mathbf{X}$ para cada $x \in \mathbf{X}$.

(b) Por otra parte, I_B:$\mathbf{X} \to [0,1]$, $\forall\ B \in \mathcal{B}(\mathbf{X})$, es una función medible y además acotada, es decir, $I_B \in \mathbf{M}_a(\mathbf{X})$ y como por la Definición L:$\mathbf{M}_a(\mathbf{X})\to\mathbf{M}_a(\mathbf{X})$, entonces $LI_B \in \mathbf{M}_a(\mathbf{X})$, por lo cual $L(B \mid \cdot)$ es una función medible para cada $B \in \mathcal{B}(\mathbf{X})$.■

Proposición 3.2 *Sea L un promediador y $L(\cdot \mid \cdot)$ la probabilidad de transición definida en (3.2). Entonces,*

$$\int_{\mathbf{X}} v(y)L(dy \mid x) = Lv(x), \quad \forall\ x \in \mathbf{X},\ v \in \mathbf{M}_a(\mathbf{X}). \tag{3.3}$$

Demostración.

(a) Sea $v(x) = I_B(x)$, para algún $B \in \mathcal{B}(\mathbf{X})$. Entonces

$$\int_{\mathbf{X}} I_B(y)L(dy \mid x) = L(B \mid x) = LI_B(x), \quad \text{(por la definición (3.2))}.$$

Por lo tanto, la proposición se cumple para funciones indicadoras.

(b) Ahora supongase que v es una función simple, es decir, que tiene la forma

$v(x) = \sum_{i=1}^{n} I_{B_i}(x)$, donde $B_i \in \mathcal{B}(\mathbf{X})$, $i = 1, 2, \cdots, n$. Entonces,

$$\begin{aligned}\int_{\mathbf{X}} \left[\sum_{i=1}^{n} I_{B_i}(x)\right] L(dy \mid x) &= \sum_{i=1}^{n} \int_{\mathbf{X}} I_{B_i}(x) L(dy \mid x) \\ &= \sum_{i=1}^{n} L I_{Bi}(x), \text{ (por (a) anterior)}, \\ &= L(\sum_{i=1}^{n} I_{B_i})(x), \quad \text{(por la definición 3.1 (b))},\end{aligned}$$

por lo que la proposición se cumple para suma de indicadoras.

(c) Sea $v \in \mathbf{M}_a(\mathbf{X})$ una función no-negativa arbitraria. Existe una sucesión de funciones simples v_n que convergen crecientemente a v. Entonces, el teorema de convergencia monótona implica que

$$Lv_n(x) = \int_{\mathbf{X}} v_n(y) L(dy \mid x) \uparrow \int_{\mathbf{X}} v(y) L(dy \mid x)$$

para cada $x \in \mathbf{X}$.

Por otra parte, como $v_n \uparrow v$, entonces

$$\begin{aligned} v - v_n &\downarrow 0 \\ L(v - v_n) &\downarrow 0, \quad \text{(por la definición 3.1 (d))} \\ Lv - Lv_n &\downarrow 0, \text{ (por la definición 3.1 (b))} \\ Lv_n &\uparrow Lv. \end{aligned}$$

Por la unicidad del límite, concluimos que

$$Lv(x) = \int_{\mathbf{X}} v(y) L(dy \mid x), \ \forall\, v \geq 0, \ x \in \mathbf{X}.$$

Entonces, la proposición se cumple para funiones no-negativas.

(d) Finalmente considere el caso general. Sea $v \in \mathbf{M}_a(\mathbf{X})$ arbitrario y v^+, v^- la parte positiva y negativa de v, respectivamente. Entonces,

$$\begin{aligned}\int_{\mathbf{X}} v(y) L(dy \mid x) &= \int_{\mathbf{X}} \left[v^+(y) - v^-(y)\right] L(dy \mid x) \\ &= \int_{\mathbf{X}} v^+(y) L(dy \mid x) - \int_{\mathbf{X}} v^-(y) L(dy \mid x) \\ &= Lv^+(x) - Lv^-(x), \quad \text{(por (c) anterior)} \\ &= L\left[v^+(x) - v^-(x)\right], \quad \text{(por la definición 3.1 (b))} \\ &= Lv(x),\end{aligned}$$

con lo cual, queda demostrada la proposición.■

Observación 3.2 *Debido a que los operador de aproximación de la definición 3.1 cumplen (3.3), a estos operadores se les denominará promediadores.*

Proposición 3.3 *Sea L un promediador, entonces:*

(a) L es monótono;

(b) L es no-expansivo con respecto a la norma del supremo $\|\cdot\|_\infty$ y a la semi-norma $\|\cdot\|_{sp}$.

Demostración.

(a) Sean $u, v \in \mathbf{M}_a(\mathbf{X})$, tales que $u \geq v$. Entonces, $u - v \geq 0$, lo cual implica que

$$\begin{aligned} L(u-v) &\geq 0, \quad \text{(por la definición 3.1 (c))} \\ Lu - Lv &\geq 0, \quad \text{(por la definición 3.1 (b))} \\ Lu &\geq Lv. \end{aligned}$$

Por lo tanto, L es monótono.

(b) Sean $u, v \in \mathbf{M}_a(\mathbf{X})$. Usando (3.3) y la proposición 3.1, se tiene

$$\begin{aligned} Lu(x) - Lv(x) &= \int_{\mathbf{X}} u(y)L(dy \mid x) - \int_{\mathbf{X}} v(y)L(dy \mid x) \\ &= \int_{\mathbf{X}} [u(y) - v(y)]\, L(dy \mid x). \end{aligned}$$

Entonces,

$$|Lu(x) - Lv(x)| \leq \sup_{x \in \mathbf{X}} |u(x) - v(x)| = \|u - v\|_\infty ,$$

lo cual prueba que L es no-expansivo con respecto a $\|\cdot\|_\infty$.

Por otra parte, directamente de (2.22), se tiene

$$\begin{aligned} \|Lu - Lv\|_{sp} &= \sup_{x \in \mathbf{X}} [Lu(x) - Lv(x)] - \inf_{x \in \mathbf{X}} [Lu(x) - Lv(x)] \\ &\leq \sup_{x \in \mathbf{X}} [u(x) - v(x)] - \inf_{x \in \mathbf{X}} [u(x) - v(x)] \\ &= \|u - v\|_{sp} . \end{aligned}$$

Entonces, L es no-expansivo con respecto a $\|\cdot\|_{sp}$.■

Observación 3.3 *La composición de probabilidades de transición es una probabilidad de transición. En particular, si L es un promediador, se tiene que $\widetilde{Q}_f = LQ_f$ es una probabilidad de transición.*

3.4 Modelos de Markov perturbados

En el proceso de aproximación, al hacer la composición del promediador L con el operador de programación dinámica T, llegamos a dos esquemas diferentes de aproximación, dependiendo de cual operador actua primero, es decir, TL ó LT.

Estos dos esquemas, nos exiguen considerar dos MCM perturbados, los cuales surgen del modelo original $\mathfrak{M} = (\mathbf{X}, \mathbf{A}, \{\mathbf{A}(x);\ x \in \mathbf{X}\}, C, Q)$.

3.4.1 Modelo perturbado $\widehat{\mathfrak{M}} = (\mathbf{X},\mathbf{A}, \{\mathbf{A}(x);\ x \in \mathbf{X}\}, C, \widehat{Q})$

En este modelo el espacio de estados, el espacio de controles, el conjunto de controles admisibles y la función de costo en una etapa son iguales a los del modelo original $\mathfrak{M}$, la probabilidad de transición se define como

$$\widehat{Q}(B \mid x, a) := \int_{\mathbf{X}} L(B \mid y) Q(dy \mid x, a), \quad (x, a) \in \mathbf{K},\ B \in \mathcal{B}(\mathbf{X}), \tag{3.4}$$

la cual efectivamente es una probabilidad de transición por ser la composición de dos de ellas (proposición 3.1 y observación 3.3).

Entonces, dada una política $\pi \in \Pi$ y un estado inicial $\widehat{x}_0 = x \in \mathbf{X}$, se genera el proceso controlado $(\widehat{x}_k, \widehat{a}_k)$, con medida de probabilidad $\widehat{P}_x^\pi$ definidos en el espacio medible $(\Omega, \mathcal{F})$, donde $\widehat{E}_x^\pi$ es el operador esperanza con respecto a tal medida.

Naturalmente, el costo descontado y la función de valor óptimo descontado estan dadas como

$$\widehat{V}_\pi(x) := \widehat{E}_x^\pi \sum_{t=0}^{\infty} \alpha^t C(\widehat{x}_t, \widehat{a}_t), \quad \widehat{x}_0 = x \in \mathbf{X},\ \pi \in \Pi, \tag{3.5}$$

$$\widehat{V}^*(x) := \inf_{\pi \in \Pi} \widehat{V}_\pi(x),\ x \in \mathbf{X}. \tag{3.6}$$

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

$$\widehat{V}^*(x) = \widehat{V}_\pi(x),\ \ \forall\ x \in \mathbf{X}. \tag{3.7}$$

Análogamente,

$$\widehat{J}_\pi(x) := \limsup_{n \to \infty} \frac{1}{n} \widehat{E}_x^\pi \sum_{t=0}^{n-1} \alpha^t C(\widehat{x}_t, \widehat{a}_t), \quad \pi \in \Pi,\ x \in \mathbf{X}, \tag{3.8}$$

$$\widehat{J}^*(x) := \inf_{\pi \in \Pi} \widehat{J}_\pi(x), \quad \pi \in \Pi,\ x \in \mathbf{X}, \tag{3.9}$$

es el costo promedio y la función de costo promedio óptimo.

Igualmente, se dice que $\pi^* \in \Pi$ es óptima en costo promedio si

$$\widehat{J}^*(x) = \widehat{J}_{\pi^*}(x),\ \forall\ x \in \mathbf{X}. \tag{3.10}$$

3.4.2 Modelo perturbado $\widetilde{\mathfrak{M}} = \left(\mathbf{X}, \mathbf{A}, \left\{\widetilde{C}_f, \widetilde{Q}_f,\ f \in \mathbf{F}\right\}\right)$

Para este modelo, el espacio de estados, el espacio de controles, asi como el conjunto de controles admisible, son los del modelo original $\mathfrak{M}$, mientras que el costo en una etapa y la probabilidad de transición, sólo están definidos para la clase de políticas estacionarias $f \in \mathbf{F}$, de la siguiente manera en notación simplificada (2.5):

$$\widetilde{C}_f := LC_f, \tag{3.11}$$

$$\widetilde{Q}_f(B \mid \cdot) := LQ_f(B \mid \cdot), \tag{3.12}$$

para cada $B \in \mathcal{B}(\mathbf{X})$ y $\ f \in \mathbf{F}$. Note que por la proposición 3.1 y la observación 3.3, $\widetilde{Q}_f$ es una probabilidad de transición en $\mathbf{X}$ para cada política estacionaria $f \in \mathbf{F}$.

Entonces, para cada política estacionaria $f \in \mathbf{F}$ y estado inicial $\widetilde{x}_0 = x \in \mathbf{X}$, existe una cadena de Markov $\{\widetilde{x}_n\}$ y una medida de probabilidad $\widetilde{P}_x^f$, definidas ambas en el espacio medible $(\Omega, \mathcal{F})$, tal que, $\widetilde{Q}_f$ es la probabilidad de transición en un paso de $\{\widetilde{x}_n\}$. El operador esperanza con respecto a $\widetilde{P}_x^f$ esta denotado por $\widetilde{E}_x^f$.

En este modelo, el costo descontado y la función de valor óptimo descontado están dadas como

$$\widetilde{V}_f(x) := \widetilde{E}_x^f \sum_{t=0}^{\infty} \alpha^t \widetilde{C}_f(\widetilde{x}_t), \quad \widetilde{x}_0 = x \in \mathbf{X},\ \ f \in \mathbf{F}, \tag{3.13}$$

$$\widetilde{V}^*(x) := \inf_{f \in \mathbf{F}} \widetilde{V}_f(x),\ x \in \mathbf{X}. \tag{3.14}$$

Una política $f^* \in \mathbf{F}$ es α-óptima si

$$\widetilde{V}^*(x) = \widetilde{V}_{f^*}(x)\ \ \forall\ x \in \mathbf{X}. \tag{3.15}$$

Análogamente, el costo promedio y la función de costo promedio óptimo se definen respectivamente como

$$\widetilde{J}_f(x) := \limsup_{n \to \infty} \frac{1}{n} \widetilde{E}_x^f \sum_{t=0}^{n-1} \widetilde{C}_f(\widetilde{x}_t), \quad f \in \mathbf{F},\ x \in \mathbf{X}, \tag{3.16}$$

$$\widetilde{J}^*(x) := \inf_{f \in \mathbf{F}} \widetilde{J}_f(x),\ x \in \mathbf{X}. \tag{3.17}$$

Y una política $f^* \in \mathbf{F}$, es óptima en costo promedio si

$$\widetilde{J}^*(x) = \widetilde{J}_{f^*}(x),\ \forall\ x \in \mathbf{X}. \tag{3.18}$$

De esta manera quedan definidos los dos MCM con los índices de funcionamiento que trabajaremos.

3.5 Constantes de aproximación

Con el objeto de medir la precisión de aproximación de un promediador, tanto en la funcion de costo en un paso C como en la probabilidad de transición Q, se considera una subclase de políticas estacionarias $\widehat{\mathbf{F}}_0$:

(a) para el índice en costo descontado, contiene las políticas estacionarias óptimas del modelo original $\mathfrak{M}$, las del modelo perturbado $\widehat{\mathfrak{M}}$ y las políticas $\widehat{V}_n-greedy$, para todo $n \in \mathbb{N}$;

(b) para el índice en costo promedio, contiene las políticas $\widehat{h}_n-greedy$, $\forall\ n \in \mathbb{N}$ y las políticas $\widehat{h}-greedy$.

Similarmente, se considera una subclase de políticas estacionarias $\widetilde{\mathbf{F}}_0$:

(a) para el índice en costo descontado, contiene las políticas óptimas de $\mathfrak{M}$, las de $\widetilde{\mathfrak{M}}$ y las políticas $\widetilde{V}_n-greedy$;

(b) para el índice en costo promedio, contiene las políticas $\widetilde{h}_n-greedy$, para todo $n \in \mathbb{N}$ y las políticas $\widetilde{h}-greedy$.

Con base en estas subclases de políticas estacionarias $\widehat{\mathbf{F}}_0$ y $\widetilde{\mathbf{F}}_0$, se definen a continuación las **constantes de aproximación**, las cuales miden el grado de exactitud del promediador. Para el modelo $\widehat{\mathfrak{M}}$,

$$\delta_Q(\widehat{\mathbf{F}}_0) = \sup_{x\in\mathbf{X},\ f\in\widehat{\mathbf{F}}_0} \left\| \widehat{Q}_f(\cdot \mid x) - Q_f(\cdot \mid x) \right\|_{VT}, \tag{3.19}$$

y para el modelo $\widetilde{\mathfrak{M}}$,

$$\delta_C(\widetilde{\mathbf{F}}_0) = \sup_{f\in\widetilde{\mathbf{F}}_0} \left\| \widetilde{C}_f - C_f \right\|_{\infty}, \tag{3.20}$$

$$\delta_Q(\widetilde{\mathbf{F}}_0) = \sup_{x\in\mathbf{X},\ f\in\widetilde{\mathbf{F}}_0} \left\| \widetilde{Q}_f(\cdot \mid x) - Q_f(\cdot \mid x) \right\|_{VT}. \tag{3.21}$$

Las siguientes caracterizaciones de la norma de variación total (veáse, [10, Apéndice B, p.125]) seran necesarias para hacer algunos comentarios respecto a las constantes de aproximación y demostrar la proposición 3.4.

Sea λ una medida con signo finita. Entonces,

$$\|\lambda\|_{VT} := \sup\left\{ \left| \int_{\mathbf{X}} v(y)\lambda(dy) \right| : v \in \mathbf{M}_a(\mathbf{X}),\ \|v\|_{\infty} \leq 1 \right\}. \tag{3.22}$$

Entonces,

$$\left| \int_{\mathbf{X}} v(y)\lambda(dy) \right| \leq \|\lambda\|_{VT} \|v\|_{\infty}, \quad v \in \mathbf{M}_a(\mathbf{X}). \tag{3.23}$$

Se puede probar que si P_1 y P_2 son medidas de probabilidad, entonces

$$\begin{aligned}\|P_1 - P_2\|_{VT} &= \sup_{\|v\|_\infty \leq 1} \left| \int_{\mathbf{X}} v(y) P_1(dy) - \int_{\mathbf{X}} v(y) P_2(dy) \right| \\ &= 2 \sup_{B \in \mathcal{B}(\mathbf{X})} |P_1(B) - P_2(B)| .\end{aligned}$$

Si además, P_1 y P_2 tienen densidades p_1 y p_2 respectivamente, con respecto a una misma medida signada, finita λ, entonces, por el teorema de Scheffe, se tiene

$$\|P_1 - P_2\|_{VT} = \int_{\mathbf{X}} |p_1 - p_2| \, d\lambda. \tag{3.24}$$

Respecto a $\delta_Q(\widehat{\mathbf{F}}_0)$ y $\delta_Q(\widetilde{\mathbf{F}}_0)$, note que son menores o igual a 2. Sin enbargo, en general es muy difícil obtener cotas más ajustadas a menos que se impongan algunas condiciones adicionales en la probabilidad de transición. Por ejemplo, los artículos [1] y [17] estudian una aproximacion del algoritmo IV para sistemas con espacios de estados continuos, asumiendo que la probabilidad de transición es absolutamente continua con respecto a una medida $\sigma-$ finita m, lo cual significa que existe una funcion de densidad $q(y \mid x, a)$ tal que

$$Q(B \mid x, a) = \int_B q(y \mid x, a) dm(y), \ \forall \ B \in \mathcal{B}(\mathbf{X}), \ (x, a) \in \mathbf{K}. \tag{3.25}$$

Si este es el caso $\delta_Q(\widetilde{\mathbf{F}}_0)$ toma una forma mas concreta y menos restrictiva, como se demuestra en la siguiente proposición.

Proposición 3.4 *Supongase que la probabilidad de transición Q satisface (3.25). Entonces, la probabilidad de transición perturbada $\widetilde{Q}_f$ es absolutamente continua con respecto a la medida m, y la función de densidad condicional es*

$$\widetilde{q}_f(\cdot \mid x) := L q_f(\cdot \mid x) = \int_{\mathbf{X}} q_f(\cdot \mid z) L(dz \mid x), \ f \in \mathbf{F}, \tag{3.26}$$

donde $q_f(\cdot \mid x) := q(\cdot \mid x, f(x))$. Además,

$$\delta_Q(\widetilde{\mathbf{F}}_0) = \sup_{x \in \mathbf{X}, \ f \in \widetilde{\mathbf{F}}_0} \int_{\mathbf{X}} |\widetilde{q}_f(y \mid x) - q_f(y \mid x)| \, dm(y).$$

Si adicionalmente se cumple que $q(y \mid x, a) \leqslant M'$, para todo $y \in \mathbf{X}$, $(x, a) \in \mathbf{K}$ y m es una medida finita, entonces

$$\delta_Q(\widetilde{\mathbf{F}}_0) \leqslant \sup_{x \in \mathbf{X}, \ f \in \widetilde{\mathbf{F}}_0} \|\widetilde{q}_f(\cdot \mid x) - q_f(\cdot \mid x)\|_\infty \, m(\mathbf{X}). \tag{3.27}$$

Demostración.

De la definición de $\widetilde{Q}_f$ dada en (3.12) y por la proposición 3.2, se tiene

$$\begin{aligned}\widetilde{Q}_f(B \mid x) &= \int_{\mathbf{X}} Q_f(B \mid z) L(dz \mid x) \ \ \forall\ x \in \mathbf{X} \text{ y } B \in \mathcal{B}(\mathbf{X}) \\ &= \int_{\mathbf{X}} \left[\int_B q_f(y \mid z) dm(y) \right] L(dz \mid x), \text{ por (3.25)} \\ &= \int_B \left[\int_{\mathbf{X}} q_f(y \mid z) L(dz \mid x) \right] dm(y),\end{aligned}$$

por lo que

$$\widetilde{q}_f(y \mid x) := \int_{\mathbf{X}} q_f(y \mid z) L(dz \mid x),\ x, y \in \mathbf{X},$$

es la función de densidad condicional con respecto a la medida m Además por (3.3) se tiene que

$$\widetilde{q}_f(y \mid x) = L q_f(y \mid x).$$

Por otra parte, por el teorema de Scheffe (3.24), vemos que

$$\left\| \widetilde{Q}_f(\cdot \mid x) - Q_f(\cdot \mid x) \right\|_{VT} = \int_{\mathbf{X}} |\widetilde{q}_f(y \mid x) - q_f(y \mid x)| \, dm(y),$$

por lo cual (3.21) implica que

$$\delta_Q(\widetilde{\mathbf{F}}_0) = \sup_{x \in \mathbf{X},\ f \in \widetilde{\mathbf{F}}_0} \int_{\mathbf{X}} |\widetilde{q}_f(y \mid x) - q_f(y \mid x)| \, dm(y).$$

De aquí, (3.27) se sigue directamente.■

Capítulo 4

Aproximación del PCO en costo descontado

4.1 Introducción

En este capítulo, se estudia el PCO con índice en costo descontado y se resulven los problemas (D1)-(D3) planteados en la introducción. Se desarrolla el algoritmo iteración de valores aproximados. Se obtienen las cotas de desempeño del algoritmo iteración de valores aproximados, que constituye la primera parte central de esta tesis ([27] en revisión para publicación).

4.2 Iteración de valores aproximados (IVA)

El obstáculo del algoritmo IV para obtener el valor de la función Tv en espacios de estados continuos, comentado al final de la sección 2.3, es superado al introducir una segunda aproximación en cada iteración del algoritmo IV mediante un promediador L (Definición 3.1). Generando los dos MCM perturbados $\widehat{\mathfrak{M}}$ y $\widetilde{\mathfrak{M}}$ vistos en el capítulo anterior y algoritmos conocidos como algoritmos de iteración de valores aproximados, los cuales veremos a continuación.

4.2.1 Modelo $\widehat{\mathfrak{M}} = (\mathbf{X}, \mathbf{A}, \{\mathbf{A}(x);\ x \in \mathbf{X}\}, C, \widehat{Q})$,

Recuerde que en el modelo $\widehat{\mathfrak{M}}$, el espacio de estados, el espacio de controles, el conjunto de controles admisibles y la función de costo en una etapa son las mismas del modelo original $\mathfrak{M}$, mientras que la probabilidad de transición fué definida en (3.4) como

$$\widehat{Q}(B \mid x, a) := \int_{\mathbf{X}} L(B \mid y) Q(dy \mid x, a), \quad (x, a) \in \mathbf{K},\ B \in \mathcal{B}(\mathbf{X}).$$

El operador de programación dinámica para este modelo perturbado $\widehat{\mathfrak{M}}$, se define como

$$\widehat{T}u(x) := \inf_{a \in \mathbf{A}(x)} \left\{ C(x, a) + \alpha \int_{\mathbf{X}} u(y) \widehat{Q}(dy \mid x, a) \right\}, \quad (x, a) \in \mathbf{K},\ u \in \mathbf{C}_a(\mathbf{X}). \tag{4.1}$$

Además, para cada $f \in \mathbf{F}$ definimos el operador

$$\widehat{T}_f u(x) := C_f(x) + \alpha \int_{\mathbf{X}} u(y) \widehat{Q}_f(dy \mid x), \quad x \in \mathbf{X},\ u \in \mathbf{C}_a(\mathbf{X}), \tag{4.2}$$

donde $\widehat{Q}_f(dy \mid x) := \widehat{Q}(dy \mid x, f(x))$. Además, definimos

$$\widehat{Q}_f u(x) := \int_{\mathbf{X}} u(y)\widehat{Q}_f(dy \mid x), \quad x \in \mathbf{X},\, u \in \mathbf{M}(\mathbf{X}), \tag{4.3}$$

siempre que la integral esté bien definida.

La definición del operador anterior en forma simplificada queda

$$\widehat{T}_f u := C_f + \alpha \widehat{Q}_f u = C_f + \alpha Q_f L u, \quad f \in \mathbf{F},\, u \in \mathbf{C}_a(\mathbf{X}).$$

Proposición 4.1 *Suponemos que el modelo original $\mathfrak{M}$ satisface las condiciones de la hipótesis 2.1. Sea T el operador de programación dinámica (2.4) y L un promediador. Entonces $\widehat{T} = TL$.*

Demostración.

Para verificar lo anterior, note que de la definición de $\widehat{T}$ dada en (4.1) y usando (3.4) se tiene que

$$\begin{aligned}
\widehat{T}u(x) &= \inf_{a \in \mathbf{A}(x)} \left\{ C(x,a) + \alpha \int_{\mathbf{X}} \int_{\mathbf{X}} u(y) L(dy \mid z) Q(dz \mid x, a) \right\}, \\
&= \inf_{a \in \mathbf{A}(x)} \left\{ C(x,a) + \alpha \int_{\mathbf{X}} Lu(z) Q(dz \mid x, a) \right\}, \quad \text{(por la proposición (3.2))} \\
&= T(Lu(x)), \quad \forall\, x \in \mathbf{X}.
\end{aligned}$$

por lo tanto, $\widehat{T} = TL$.■

En el siguiente lema, se dan las condiciones para que $\widehat{\mathfrak{M}}$ satisfaga el teorema 2.1 y el teorema 2.2.

Lema 4.1 *Suponemos que el modelo original $\mathfrak{M}$ satisface las condiciones de la hipótesis 2.1 y que L es un promediador continuo. Entonces el modelo perturbado $\widehat{\mathfrak{M}}$, también satisface las condiciones de la hipótesis 2.1 y por lo tanto, en el modelo $\widehat{\mathfrak{M}}$ se cumplen todas las conclusiones del teorema 2.1 y del teorema 2.2.*

Demostración.

Como se vio en la sección 3.4.1, todos los objetos del modelo $\widehat{\mathfrak{M}}$ son los mismos del modelo original $\mathfrak{M}$, exepto la probabilidad de transición $\widehat{Q}$ definido en (3.4).

Para que se satisfagan todas las condiciones de la hipótesis 2.1, solo falta demostrar que $\widehat{Q}$ sea débilmente continua en $\mathbf{K}$. La continuidad debil se deduce de la definición de $\widehat{Q}$ y la proposición 3.2, ya que

$$\int_{\mathbf{X}} u(y)\widehat{Q}(dy \mid x, a) = \int_{\mathbf{X}} \int_{\mathbf{X}} u(y) L(dy \mid z) Q(dz \mid x, a) \quad = \int_{\mathbf{X}} Lu(y) Q(dy \mid x, a).$$

Si $u \in \mathbf{C}_a(\mathbf{X})$, como L es continuo entonces $Lu \in \mathbf{C}_a(\mathbf{X})$. Además, como Q es debilmente continua en $\mathbf{K}$. Entonces, el mapeo

$$(x,a) \to \int_{\mathbf{X}} Lu(y)Q(dy \mid x,a),$$

es continuo para cada función $u \in \mathbf{C}_a(\mathbf{X})$. Por lo tanto, $\widehat{Q}$ es debilmente continua en $\mathbf{K}$.
Por otra parte, sabemos que el operador de programación dinámica T es de contracción en el espacio de Banach $(\mathbf{C}_a(\mathbf{X}), \|\cdot\|_\infty)$ con factor de contracción α y L es no-expansivo en el mismo espacio (proposición 3.3 (b)), por lo cual el operador $\widehat{T} := TL$ también es de contracción en $(\mathbf{C}_a(\mathbf{X}), \|\cdot\|_\infty)$ con factor de contracción α y se concluye que en $\widehat{\mathfrak{M}}$ se cumplen todas las conclusiones del teorema 2.1 y del teorema 2.2.■

Con base al lema anterior y del teorema 2.1 (d) para el modelo $\widehat{\mathfrak{M}}$, se puede construir un procedimiento para encontrar una aproximación a la solución del PCO con costo descontado, llamado algoritmo iteración de valores aproximado (IVA), el cual tiene la siguiente estructura

$$\widehat{V}_{k+1} = \widehat{T}\widehat{V}_k, \quad k \in \mathbb{N}_0, \tag{4.4}$$

donde $\widehat{V}_0$ es una funcion arbitraria fija en $\mathbf{C}_a(\mathbf{X})$. Además, el teorema 2.1 (d) asegura que $\widehat{V}_k \to \widehat{V}^*$ geométricamente, donde $\widehat{V}^*$ es la función valor óptimo para $\widehat{\mathfrak{M}}$. Con esto, queda resulto el problema (D1).

El algoritmo IVA también se detiene como el algoritmo estándar IV, en una etapa $k \in \mathbb{N}$ mediante una regla de paro. Luego se calcula una política $f \in \mathbf{F}$, $\widehat{V}_k - greedy$ y la función valor óptimo $\widehat{V}^*$ es aproximada por la función valor descontado $\widehat{V}_f$.

Observación 4.1 *Las instrucciones del Algoritmo* IVA *para el modelo perturbado* $\widehat{\mathfrak{M}}$ *son exactamente las mismas instrucciones que las del algoritmo* IV *en costo descontado (2.1) para el modelo original* $\mathfrak{M}$, *unicamente hay que sustituir el operador* T *por el operador* $\widehat{T}$.

Con base en el lema 4.1, se establece el análogo al teorema 2.2 para el modelo $\widehat{\mathfrak{M}}$ en el siguiente lema, en el cual se establece directamente una cota para el error de aproximación $\widehat{V}^* - \widehat{V}_f$.

Lema 4.2 *Suponemos que el modelo original* $\mathfrak{M}$ *satisface las condiciones de la hipótesis 2.1 y* L *es un promediador continuo. Entonces,*

$$\left\|\widehat{V}^* - \widehat{V}_f\right\|_\infty \leq \frac{2\alpha}{1-\alpha}\left\|\widehat{V}_k - \widehat{V}_{k-1}\right\|_\infty, \tag{4.5}$$

donde $f \in \mathbf{F}$ *es la politica* $\widehat{V}_k$ $-greedy$.

4.2.2 Modelo $\widetilde{\mathfrak{M}} = (\mathbf{X}, \mathbf{A}, \left\{\widetilde{C}_f, \widetilde{Q}_f,\ f \in \mathbf{F}\right\})$

El modelo $\widetilde{\mathfrak{M}}$ visto en el capítulo anterior, considera que el espacio de estados, el espacio de controles, asi como el conjunto de controles admisible permanecen iguales a los del modelo original $\mathfrak{M}$, mientras que los costos en una etapa y la probabilidad de transición perturbada están definidos en (3.11) y (3.12) para la clase de políticas estacionarias $f \in \mathbf{F}$ como

$$\widetilde{C}_f := LC_f,$$

$$\widetilde{Q}_f(B \mid \cdot) := LQ_f(B \mid \cdot) = \int_{\mathbf{X}} Q_f(B \mid y)L(dy \mid \cdot),$$

respectivamente, para cada $B \in \mathcal{B}(\mathbf{X})$, $f \in \mathbf{F}$.

El operador de programación dinámica para $\widetilde{\mathfrak{M}}$, se define como

$$\widetilde{T}u(x) := \inf_{f\in\mathbf{F}} \left\{\widetilde{C}_f(x) + \alpha\int_{\mathbf{X}} u(y)\widetilde{Q}_f(dy \mid x)\right\}, \quad x \in \mathbf{X},\ u \in \mathbf{M}_a(\mathbf{X}). \tag{4.6}$$

También, para cada $f \in \mathbf{F}$ definimos el operador

$$\widetilde{T}_f u(\cdot) := \widetilde{C}_f(\cdot) + \alpha\int_{\mathbf{X}} u(y)\widetilde{Q}_f(dy \mid \cdot), \quad f \in \mathbf{F},\ u \in \mathbf{M}_a(\mathbf{X}). \tag{4.7}$$

Además definimos

$$\widetilde{Q}_f u(x) := \int_{\mathbf{X}} u(y)\widetilde{Q}_f(dy \mid x), \quad x \in \mathbf{X},\ \ f \in \mathbf{F}, \tag{4.8}$$

donde $u \in \mathbf{M}(\mathbf{X})$, siempre que la integral esté bien definida. Las definiciones de los operador anteriores en forma simplificada quedan

$$\widetilde{T}u = \inf_{f\in\mathbf{F}} \left\{\widetilde{C}_f + \alpha\widetilde{Q}_f u\right\} \text{ y } \widetilde{T}_f u := \widetilde{C}_f + \alpha\widetilde{Q}_f u(x), \quad f \in \mathbf{F},\ u \in \mathbf{M}_a(\mathbf{X}).$$

Proposición 4.2 *Suponemos que el modelo original $\mathfrak{M}$ satisface las condiciones de la hipótesis 2.2 y que L es un promediador en $\mathbf{M}_a(\mathbf{X})$. Entonces $\widetilde{T} = LT$.*

Demostración.

Para verificarlo, considere una función fija arbitraria $u \in \mathbf{M}_a(\mathbf{X})$. Note que

$$Tu(\cdot) \le C_f(\cdot) + \alpha\int_{\mathbf{X}} u(y)Q_f(dy \mid \cdot), \quad \forall\ f \in \mathbf{F}.$$

Por la linealidad y monotonia de L (proposición 3.3 (a)) y por la observación 3.3 se sigue que

$$LTu(\cdot) \le \widetilde{C}_f(x) + \alpha\int_{\mathbf{X}} u(y)\widetilde{Q}_f(dy \mid \cdot) \quad \forall\ f \in \mathbf{F}.$$

Por otra parte, recordemos que existe f_u, tal que

$$Tu(\cdot) = C_{f_u}(\cdot) + \alpha \int_{\mathbf{X}} u(y) Q_{f_u}(dy \mid \cdot);$$

de nuevo, usando la linealidad y monotonia de L (proposición 3.3 (a)) y la observación 3.3, vemos que

$$\begin{aligned} LTu(\cdot) &= \widetilde{C}_{f_u}(x) + \alpha \int_{\mathbf{X}} u(y) \widetilde{Q}_{f_u}(dy \mid \cdot) \\ &\leq \widetilde{C}_f(x) + \alpha \int_{\mathbf{X}} u(y) \widetilde{Q}_f(dy \mid \cdot), \quad \forall\ f \in \mathbf{F}. \end{aligned}$$

De la definición de $\widetilde{T}$, resulta que

$$\widetilde{T}u = LTu \quad \forall\ u \in \mathbf{M}_a(\mathbf{X}).$$

Por lo tanto, $\widetilde{T} = LT$.■

Observación 4.2 *Por la proposición 3.3 (b) sabemos que L es operador no-expansivo en $(\mathbf{M}_a(\mathbf{X}), \|\cdot\|_\infty)$ y además que T es operador de contracción módulo α. Entonces, bajo la hipótesis 2.2 también $\widetilde{T}$ es operador de contracción módulo α en $(\mathbf{M}_a(\mathbf{X}), \|\cdot\|_\infty)$.*

Por lo cual, resulta natural el siguiente lema análogo al teorema 2.1.

Lema 4.3 *Suponemos que el modelo original $\mathfrak{M}$ satisface las condiciones de la hipótesis 2.2 y que L es un promediador. Entonces, la función valor óptimo descontado $\widetilde{V}^*(x)$ es la única función en $\mathbf{M}_a(\mathbf{X})$ que satisface la* ***ecuación de optimalidad***

$$\widetilde{V}^*(x) = \inf_{f \in \mathbf{F}} \left\{ \widetilde{C}_f(x) + \alpha \int_{\mathbf{X}} \widetilde{V}^*(y) \widetilde{Q}_f(dy \mid x) \right\} \quad \forall\ x \in \mathbf{X},$$

y existe una política estacionaria $\widetilde{f} \in \mathbf{F}$ tal que

$$\widetilde{V}^*(x) = \widetilde{C}_{\widetilde{f}}(x) + \alpha \int_{\mathbf{X}} \widetilde{V}^*(y) \widetilde{Q}_{\widetilde{f}}(dy \mid x) \quad \forall\ x \in \mathbf{X}.$$

Por lo tanto, $\widetilde{f}$ es óptima. Además, el resto de las conclusiones del teorema 2.1 (b)-(d) se cumplen para el modelo $\widetilde{\mathfrak{M}}$.

Demostración.

De la observación 4.2 y el teorema de punto fijo de Banach, existe una única función $\widetilde{W}$ en $\mathbf{M}_a(\mathbf{X})$ tal que $\widetilde{W} = \widetilde{T}\widetilde{W}$, es decir,

$$\widetilde{W}(x) = \inf_{f \in \mathbf{F}} \left\{ \widetilde{C}_f(x) + \alpha \int_{\mathbf{X}} \widetilde{W}(x) \widetilde{Q}_f(dy \mid x) \right\}.$$

Por otra parte, para todo $f \in \mathbf{F}$, se tiene que

$$\begin{aligned}\widetilde{W}(x) &\leq LT_f\widetilde{W}(x)\\ &= L\left[C_f(x) + \alpha\int_{\mathbf{X}} \widetilde{W}(y)Q_f(dy \mid x)\right], \quad \text{(por la Definición 3.1 (b))}\\ &= \widetilde{C}_f(x) + \alpha\int_{\mathbf{X}} \widetilde{W}(y)\widetilde{Q}_f(dy \mid x), \quad \text{(por la proposición 3.2).}\end{aligned}$$

Además, la hipótesis 2.2 garantiza la existencia de un selector $\widetilde{f} \in \mathbf{F}$, tal que

$$T\widetilde{W}(x) = C_{\widetilde{f}}(x) + \alpha\int_{\mathbf{X}} \widetilde{W}(y)Q_{\widetilde{f}}(dy \mid x), \quad x \in \mathbf{X}.$$

Aplicando el operador L en ambos miembros de la ecuación anterior y considerando la linealidad de L y la proposición 3.2, se obtiene que

$$\widetilde{T}\widetilde{W}(x) = \widetilde{C}_{\widetilde{f}}(x) + \alpha\int_{\mathbf{X}} \widetilde{W}(y)\widetilde{Q}_{\widetilde{f}}(dy \mid x).$$

Como $\widetilde{W} = \widetilde{T}\ \widetilde{W}$, entonces

$$\widetilde{W}(x) = \widetilde{C}_{\widetilde{f}}(x) + \alpha\int_{\mathbf{X}} \widetilde{W}(y)\widetilde{Q}_{\widetilde{f}}(dy \mid x).$$

Usando argumentos estandar de programación dinámica, es fácil ver que la ecuación anterior y la desigualdad implican

$$\widetilde{W} = \widetilde{V}^* = V_{\widetilde{f}},$$

lo cual demuestra la primera parte del teorema. La prueba del resto de las partes se siguen de argumentos de programación dinámica después de notar que el operador $\widetilde{T}_f$ en (4.7) es un operador de contracción de $\mathbf{M}_a(\mathbf{X})$ en si mismo con modulo α y que su único punto fijo es la función $\widetilde{V}_f$.■

Observación 4.3 *Note que en este modelo $\widetilde{\mathfrak{M}}$ no es necesario que el promediador L sea continuo, ya que aquí, L se aplica a T por la izquierda y se distribuye por linealidad, además de que $\mathfrak{M}$ satisface las condiciones de la hipótesis 2.2, por lo que se garantiza el resultado del lema anterior.*

Con base a la parte (d) del teorema 2.1 para el modelo $\widetilde{\mathfrak{M}}$, se puede construir el algoritmo IVA para $\widetilde{\mathfrak{M}}$, el cual tiene la siguiente estructura

$$\widetilde{V}_{k+1} = \widetilde{T}\widetilde{V}_k, \quad k \in \mathbb{N}_0, \tag{4.9}$$

donde $\widetilde{V}_0$ es una funcion arbitraria fija en $\mathbf{M}_a(\mathbf{X})$. Además, este teorema asegura que $\widetilde{V}_k \to \widetilde{V}^*$ geométricamente, donde $\widetilde{V}^*$ es la función valor óptimo para $\widetilde{\mathfrak{M}}$. Con esto, queda resuelto el problema (D1) para este modelo $\widetilde{\mathfrak{M}}$.

Este otro algoritmo IVA también se detiene en alguna etapa $k \in \mathbb{N}$ mediante una regla de paro, se calcula una política $f \in \mathbf{F}$, $\widetilde{V}_k{-}greedy$ y la función valor óptimo $\widetilde{V}^*$ es aproximada por la función $\widetilde{V}_f$.

Observación 4.4 *Las instrucciones del Algoritmo* IVA *para el modelo perturbado* $\widetilde{\mathfrak{M}}$ *son exactamente las mismas instrucciones que las del algoritmo* IV *en costo descontado (2.1) para el modelo original* $\mathfrak{M}$, *unicamente hay que sustituir el operador* T *por el operador* $\widetilde{T}$.

Por otra parte, también resulta natural que bajo las condiciones del lema 4.3, se cumplan las conclusión del teorema 2.2 para el modelo $\widetilde{\mathfrak{M}}$. El siguiente lema establece una cota para el error de aproximación $\widetilde{V}^* - \widetilde{V}_f$.

Lema 4.4 *Suponemos que el modelo original* $\mathfrak{M}$ *satisface las condiciones de la hipótesis 2.2 y que* L *es un promediador. Entonces,*

$$\left\|\widetilde{V}^* - \widetilde{V}_f\right\|_\infty \leq \frac{2\alpha}{1-\alpha}\left\|\widetilde{V}_k - \widetilde{V}_{k-1}\right\|_\infty,$$

donde $f \in \mathbf{F}$ *es una política* $\widetilde{V}_k-greedy$.

En la siguiente proposición, se determinan las diferencias de los algoritmos IVA de ambos modelos.

Proposición 4.3 *Suponemos que el modelo original* $\mathfrak{M}$ *satisface las condiciones de la hipótesis 2.1 y que* L *es un promediador continuo. Entonces:*

(a) $\widetilde{V}^* = L\ \widehat{V}^*$ *y* $\widehat{V}^* = T\widetilde{V}^*$;

(b) $f \in \mathbf{F}$ *es óptima para el modelo* $\widehat{\mathfrak{M}}$ *si y sólo si es óptima para el modelo* $\widetilde{\mathfrak{M}}$.

Demostración.

(a) Del lema 4.1, $\widehat{V}^*$ es el único punto fijo de $\widehat{T}$ en $\mathbf{C}_a(\mathbf{X})$, es decir,

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

Entonces,

$$L\widehat{V}^* = LT(L\widehat{V}^*) = \widetilde{T}(L\widehat{V}^*),$$

lo cual significa que $L\widehat{V}^* \in \mathbf{C}_a(\mathbf{X})$ es punto fijo de $\widetilde{T}$ y por el lema 4.3, $\widetilde{V}^*$ es el único punto fijo de $\widetilde{T}$, por lo que se concluye que $\widetilde{V}^* = L\ \widehat{V}^*$.

Con argumentos similares se llega a que $\widehat{V}^* = T\widetilde{V}^*$.

(b) La demostración de esta parte es practicamente la misma, pero se considera que los operadores $\widehat{T}_f$ en (4.2) y $\widehat{T}_f$ en (4.7) son operador de contracción de $\mathbf{C}_a(\mathbf{X})$ en si mismo con modulo α.■

En la siguiente sección a partir de la política f estacionaria $\widehat{V}_k-greedy$ o de la política f estacionaria $\widetilde{V}_k-greedy$ se analiza como la función valor óptimo V^* del modelo original $\mathfrak{M}$ es aproximada por la función valor descontado V_f del modelo original $\mathfrak{M}$.

4.3 Cotas de aproximación del algoritmo IVA

En esta sección se resuelven los problemas (D2) y (D3) planteados desde la introducción para ambos modelos, y se obtienen las cotas de aproximación de los algoritmos IVA para costo descontado.

La exactitud de los algoritmos IVA utilizando promediadores es expresada en términos de la constate de aproximación $\delta_Q(\widehat{\mathbf{F}}_0)$ para $\widehat{\mathfrak{M}}$ dada en (3.19) y de las constates de aproximación $\delta_C(\widetilde{\mathbf{F}}_0)$ y $\delta_Q(\widetilde{\mathbf{F}}_0)$ para $\widetilde{\mathfrak{M}}$ dadas en (3.20) y (3.21), respectivamente. Dicha exactitud es expresada en el siguiente teorema, el cual es el primer resultado fundamental de esta tesis.

Teorema 4.1 *Suponemos que el modelo original $\mathfrak{M}$ satisface las condiciones de la hipótesis 2.1 y que L es un promediador continuo. Entonces, para el modelo perturbado $\widehat{\mathfrak{M}}$ se cumple lo siguiente:*

(a) $\left\|V_f - \widehat{V}_f\right\|_\infty \leqslant \frac{\alpha M}{(1-\alpha)^2} \sup_{x\in\mathbf{X}} \left\|\widehat{Q}_f(\cdot \mid x) - Q_f(\cdot \mid x)\right\|_{VT}, \quad \forall\ f \in \mathbf{F};$
donde M es cota de la función de costo C.

(b) $\left\|V^* - \widehat{V}^*\right\|_\infty \leqslant \frac{\alpha M}{(1-\alpha)^2}\delta_Q(\widehat{\mathbf{F}}_0);$

(c) Si $f \in \mathbf{F}$ es una politica $\widehat{V}_k$ $-greedy$ entonces la cota de error total es

$$\|V^* - V_f\|_\infty \leq \frac{2\alpha}{1-\alpha}\left\|\widehat{V}_k - \widehat{V}_{k-1}\right\|_\infty + \frac{2\alpha M}{(1-\alpha)^2}\delta_Q(\widehat{\mathbf{F}}_0), \tag{4.10}$$

Demostración.

(a) Del teorema 2.1 y del lema 4.1, consideremos que

$$V_f = C_f + \alpha Q_f V_f,$$
$$\widehat{V}_f = C_f + \alpha \widehat{Q}_f \widehat{V}_f.$$

Entonces,

$$\begin{aligned}\left\|\widehat{V}_f - V_f\right\|_\infty &= \alpha\left\|\widehat{Q}_f\widehat{V}_f - Q_f V_f\right\|_\infty \\ &\leq \alpha\left\|\widehat{Q}_f\widehat{V}_f - Q_f\widehat{V}_f\right\|_\infty + \alpha\left\|Q_f\widehat{V}_f - Q_f V_f\right\|_\infty \\ &\leq \alpha \sup_{x\in\mathbf{X}}\left\|\widehat{Q}_f(\cdot \mid x) - Q_f(\cdot \mid x)\right\|_{VT}\left\|\widehat{V}_f\right\|_\infty + \alpha\left\|\widehat{V}_f - V_f\right\|_\infty.\end{aligned}$$

Como $\left\|\widehat{V}_f\right\|_\infty \leq M(1-\alpha)^{-1}$, se tiene

$$(1-\alpha)\left\|\widehat{V}_f - V_f\right\|_\infty \leq \alpha M(1-\alpha)^{-1}\sup_{x\in\mathbf{X}}\left\|\widehat{Q}_f(\cdot \mid x) - Q_f(\cdot \mid x)\right\|_{VT},$$

de donde se deduce (a)

(b) De (a) se sigue que

$$-K + \widehat{V}_f \leq V_f \leq \widehat{V}_f + K, \quad \forall\ f \in \widehat{\mathbf{F}}_{\mathbf{0}},$$

donde

$$K := \frac{\alpha M}{(1-\alpha)^2} \sup_{x \in \mathbf{X}} \left\| \widehat{Q}_f(\cdot \mid x) - Q_f(\cdot \mid x) \right\|_{VT}.$$

Puesto que $\widehat{\mathbf{F}}_{\mathbf{0}}$ contiene las políticas estacionarias óptimas tanto de $\mathfrak{M}$ como de $\widehat{\mathfrak{M}}$, tomando el supremo sobre toda esta clase, se tiene que

$$-\frac{\alpha M}{(1-\alpha)^2}\delta_Q(\widehat{\mathbf{F}}_0) + \widehat{V}^* \leq V^* \leq \widehat{V}^* + \frac{\alpha M}{(1-\alpha)^2}\delta_Q(\widehat{\mathbf{F}}_0), \quad \forall\ f \in \widehat{\mathbf{F}}_{\mathbf{0}},$$

lo cual prueba (b).

(c) Ahora, observe que

$$\|V^* - V_f\|_\infty \leq \left\|V^* - \widehat{V}^*\right\|_\infty + \left\|\widehat{V}^* - \widehat{V}_f\right\|_\infty + \left\|\widehat{V}_f - V_f\right\|_\infty.$$

Entonces, la parte (c) se sigue de (a)-(b) y el lema 4.2. .■

A continuación damos el resultado análogo para el modelo perturbado $\widetilde{\mathfrak{M}}$.

Teorema 4.2 *Suponemos que el modelo original $\mathfrak{M}$ satisface las condiciones de la hipótesis 2.2 y que L es un promediador. Entonces, para el modelo perturbado $\widetilde{\mathfrak{M}}$ se cumplen las siguientes afirmaciones:*

(a) $\left\|V_f - \widetilde{V}_f\right\|_\infty \leqslant \frac{1}{1-\alpha}\left\|C_f - \widetilde{C}_f\right\|_\infty + \frac{\alpha M}{(1-\alpha)^2}\sup_{x\in\mathbf{X}}\left\|Q_f(\cdot \mid x) - \widetilde{Q}_f(\cdot \mid x)\right\|_{VT}$,
donde M es cota de la función de costo C y $c_0 = M/(1-\alpha)$.

(b) $\left\|V^* - \widetilde{V}^*\right\|_\infty \leqslant \frac{1}{1-\alpha}\delta_C(\widetilde{\mathbf{F}}_0) + \frac{\alpha M}{(1-\alpha)^2}\delta_Q(\widetilde{\mathbf{F}}_0)$;

(c) Si $f \in \mathbf{F}$ es una politica $\widetilde{V}_k$ $-greedy$ entonces la cota de error total es

$$\|V^* - V_f\|_\infty \leq \frac{2\alpha}{1-\alpha}\left\|\widetilde{V}_k - \widetilde{V}_{k-1}\right\|_\infty + \frac{2}{1-\alpha}\delta_C(\widetilde{\mathbf{F}}_0) + \frac{2\alpha M}{(1-\alpha)^2}\delta_Q(\widetilde{\mathbf{F}}_0). \quad (4.11)$$

Demostración
Como en la demostración del teorema 4.1, es suficiente demostrar (a)

(a) Para hacerlo, note que del teorema 2.2 y lema 4.3

$$V_f = C_f + \alpha Q_f V_f,$$
$$\widetilde{V}_f = \widetilde{C}_f + \alpha \widetilde{Q}_f \widetilde{V}_f.$$

De aqui, tenemos que

$$\begin{aligned}\left\|V_f - \widetilde{V}_f\right\|_\infty &\leqslant \left\|C_f - \widetilde{C}_f\right\|_\infty + \alpha \left\|Q_f V_f - \widetilde{Q}_f \widetilde{V}_f\right\|_\infty \\ &\leqslant \left\|C_f - \widetilde{C}_f\right\|_\infty + \alpha \left\|Q_f V_f - Q_f \widetilde{V}_f\right\|_\infty + \alpha \left\|Q_f \widetilde{V}_f - \widetilde{Q}_f \widetilde{V}_f\right\|_\infty \\ &\leqslant \left\|C_f - \widetilde{C}_f\right\|_\infty + \alpha \left\|V_f - \widetilde{V}_f\right\|_\infty \\ &\quad + \alpha \sup_{x \in \mathbf{X}} \left\|Q_f(\cdot \mid x) - \widetilde{Q}_f(\cdot \mid x)\right\|_{VT} \left\|\widetilde{V}_f\right\|_\infty .\end{aligned}$$

Como $\left\|\widetilde{V}_f\right\|_\infty \leq M(1-\alpha)^{-1}$, se tiene

$$\left\|V_f - \widetilde{V}_f\right\|_\infty \leqslant \frac{1}{1-\alpha} \left\|C_f - \widetilde{C}_f\right\|_\infty + \frac{\alpha M}{(1-\alpha)^2} \sup_{x \in \mathbf{X}} \left\|Q_f(\cdot \mid x) - \widetilde{Q}_f(\cdot \mid x)\right\|_{VT} . \blacksquare$$

Los teoremas anteriores para los respectivos modelos con (b) resuelven el problema (D2) y con (c) resuelven el problema (D3). En (c) de ambos teoremas, al primer término del lado derecho le llamaremos **cota del error de paro** y al resto **cota de error de aproximación,** para los respectivos modelos perturbados.

Capítulo 5

Aproximación del PCO en costo promedio

5.1 Introducción

En este capítulo se estudia el PCO con respecto al índice en costo promedio. Se resuelven los problemas (P1)-(P3) planteados en la introducción. Para lo anterior, se desarrolla el algoritmo de IVA y las suposiciones bajo las cuales este algoritmo funciona. Se desarrolla la otra parte central de esta tesis, las cotas de desempeño del algoritmo de IVA.

5.2 Iteracion de valores aproximado (IVA)

Comenzaremos la construcción del algoritmo IVA en costo promedio, recordando que en el Capítulo 2, la observación 2.5 y el teorema 2.4 garantizan que el algoritmo IV (2.24) ó el equivalente (2.26), obtienen una solución aproximada del PCO siempre y cuando TJ_n ó $T_z h_n$ sean computables en cada iteración. Sin embargo, como se comentó al final de la sección 2.4, al igual que en el PCO en costo descontado, este procedimiento tampoco es computable para los sistemas con espacios continuos o discretos muy grandes, ya que la implementación computacional de TJ_n ó de $T_z h_n$ es imposible.

De nuevo, este obstáculo del algoritmo IV para el PCO en costo promedio es superado, al introducir una segunda aproximación mediante un promediador L en cada iteración del algoritmo IV, el cual es aplicado entre dos ejecuciones consecutivas de T en la ecuación recursiva (2.24). A continuación se desarrolla el algoritmos IVA para cada uno de los MCM perturbados $\widehat{\mathfrak{M}}$ y $\widetilde{\mathfrak{M}}$. La diferencia sustancial entre los dos desarrollos son: En $\widehat{\mathfrak{M}}$ se trabaja en el espacio $(\mathbf{C}_a^0(\mathbf{X}), \|\cdot\|_{sp})$, bajo la suposición de que en el modelo original $\mathfrak{M}$ se satisface las hipótesis 2.1 y 2.3, mientras que en $\widetilde{\mathfrak{M}}$ se trabaja en $(\mathbf{M}_a^0(\mathbf{X}), \|\cdot\|_{sp})$, bajo la suposición de que en $\mathfrak{M}$ se satisface las hipótesis 2.2 y 2.3.

5.2.1 Modelo $\widehat{\mathfrak{M}} = (\mathbf{X}, \mathbf{A}, \{\mathbf{A}(x);\ x \in \mathbf{X}\}, C, \widehat{Q})$,

A manera de recordatorio en este modelo el espacio de estados, el espacio de controles, el conjunto de controles admisibles y la función de costo en una etapa son las mismas del modelo original $\mathfrak{M}$, mientras que la probabilidad de transición fué definida en (3.4) como

$$\widehat{Q}(B \mid x, a) := \int_{\mathbf{X}} L(B \mid y) Q(dy \mid x, a), \quad (x, a) \in \mathbf{K},\ B \in \mathcal{B}(\mathbf{X}).$$

Se define el operador de programación dinámica

$$\widehat{T}u(x) := \inf_{a \in \mathbf{A}(x)} \left\{ C(x, a) + \int_{\mathbf{X}} u(y) \widehat{Q}(dy \mid x, a) \right\}, \quad (x, a) \in \mathbf{K},\ u \in \mathbf{C}_a(\mathbf{X}). \quad (5.1)$$

También, para cada $f \in \mathbf{F}$, definimos el operador

$$\widehat{T}_f u(x) := C_f(x) + \int_{\mathbf{X}} u(y) \widehat{Q}_f(dy \mid x), \quad x \in \mathbf{X},\ u \in \mathbf{C}_a(\mathbf{X}), \quad (5.2)$$

donde $\widehat{Q}_f(dy \mid x) := \widehat{Q}(dy \mid x, f(x))$. Utilizando las definiciones (4.3) este operador en notación simplificada queda

$$\widehat{T}_f u = C_f + \widehat{Q}_f u, \quad \forall \quad f \in \mathbf{F} \text{ y } u \in \mathbf{C}_a(\mathbf{X}).$$

Observe que los operadores $\widehat{T}$ y $\widehat{T}_f$ son iguales a los de costo descontado con $\alpha = 1$. Entonces por los mismos argumentos de la proposición 4.1 tomando $\alpha = 1$, se tiene la siguiente observación.

Observación 5.1 *Suponemos que en el modelo original $\mathfrak{M}$ se satisfacen las condiciones de la hipótesis 2.1. Sea T el operador de programación dinámica (2.16) y L un promediador. Entonces, $\widehat{T} = TL$.*

Ahora con el operador $\widehat{T}$ definimos las funciones análogas a las del algoritmo IV (2.24).

$$\widehat{J}_0 := 0 \text{ y } \widehat{J}_{n+1} = \widehat{T}\widehat{J}_n, \quad n \in \mathbb{N}, \quad (5.3)$$

La ecuacion de optimalidad para costo promedio queda de la forma

$$\widehat{\rho} + \widehat{h} = \widehat{T}\widehat{h}, \quad (5.4)$$

donde $\widehat{\rho}$ es una constante y $\widehat{h} \in \mathbf{C}_a(\mathbf{X})$. Una solución de la ecuación de optimalidad consiste en un par formado por una constante y una función. Si tal solución existe, digamos $(\widehat{\rho}^*, \widehat{h}^*)$, entonces existe la política $\widehat{f}^* \in \mathbf{F}$, denominada política $\widehat{h}^*-greed$ (lo cual, se demuestra más adelante), esto es,

$$\widehat{\rho}^* + \widehat{h}^* = \widehat{T}_{\widehat{f}^*}\widehat{h}^*.$$

Y a la terna $(\widehat{\rho}^*, \widehat{h}^*, \widehat{f}^*)$ se le llama terna canónica.

Se obtiene la solución de la ecuación de optimalidad como límite las sucesiones

$$\widehat{\rho}_{n+1} = \widehat{J}_{n+1}(z) - \widehat{J}_n(z), \quad n \in \mathbb{N}_0, \tag{5.5}$$

$$\widehat{h}_{n+1} = \widehat{J}_{n+1} - \widehat{J}_{n+1}(z), \quad n \in \mathbb{N}_0, \tag{5.6}$$

con z un estado arbitrario pero fijo. Estas sucesiones están relacionadas como sigue

$$\widehat{\rho}_{n+1} + \widehat{h}_{n+1}(\cdot) = \widehat{T}\widehat{h}_n(\cdot), \quad n \in \mathbb{N}_0,$$

con $\widehat{h}_0 \equiv 0$.

Note que, $\widehat{\rho}_{n+1} = \widehat{T}\widehat{h}_n(z)$ para todo $n \in \mathbb{N}_0$, porque $\widehat{h}_{n+1}(z) = 0$. Entonces

$$\widehat{h}_{n+1} = \widehat{T}\widehat{h}_n - \widehat{T}\widehat{h}_n(z), \quad n \in \mathbb{N}_0,$$

definiendo el operador $\widehat{T}_z$ para cada $u \in \mathbf{C}_a(\mathbf{X})$ como

$$\widehat{T}_z u := \widehat{T}u - \widehat{T}u(z), \tag{5.7}$$

entonces, se obtiene la otra ecuación recursiva del algoritmo IVA.

$$\widehat{h}_{n+1} = \widehat{T}_z \widehat{h}_n, \quad n \in \mathbb{N}_0, \tag{5.8}$$

que al considerar que $\widehat{T} = TL$, se tiene explícitamente

$$\widehat{h}_{n+1} := TL\widehat{h}_n - TL\widehat{h}_n(z), \quad n \in \mathbb{N}_0,$$

que es la referida por la ecuación (10) en la introducción.

A continuación se dan las condiciones para que $\widehat{\mathfrak{M}}$ satisfaga las conclusiones de la observación 2.5 y el teorema 2.4, con lo cual se garantizará la existencia de solución de la ecuación de optimalidad (5.4).

Por el lema 4.1, se sabe que si en el modelo original $\mathfrak{M}$ se satisface las condiciones de la hipótesis 2.1 y que L es un promediador continuo, entonces en el modelo perturbado $\widehat{\mathfrak{M}}$ también se satisface las condiciones de la hipótesis 2.1, sólo falta demostrar la condición de ergodicidad para en este modelo. Para lo anterior es necesario de la siguiente definición

$$\widehat{Q}u(x) := \int_{\mathbf{X}} u(y)\widehat{Q}(dy \mid x, a), \quad (x, a) \in \mathbf{K},\, u \in \mathbf{M}(\mathbf{X}).$$

Note que de la definición de $\widehat{Q}$ y la proposición 3.2, se tiene que

$$\widehat{Q}u(x) = \int_{\mathbf{X}} Lu(y)Q(dy \mid x, a). \tag{5.9}$$

En el siguiente lema se demuestra la condicion de ergodicidad.

Lema 5.1 *Suponemos que en el modelo original se satisface la condición de ergodicidad (hipótesis 2.3). Entonces, en $\widehat{\mathfrak{M}}$ también se satisface, es decir:*

$$\left\|\widehat{Q}(\cdot \mid x,a) - \widehat{Q}(\cdot \mid y,b)\right\|_{VT} \leqslant 2\beta, \quad \forall\ (x,a),(y,b) \in \mathbf{K},$$

donde β es la constante de la condición de ergodicidad (hipótesis 2.3) en el modelo original $\mathfrak{M}$.

Demostración.

Considerando $u \in \mathbf{C}_a(\mathbf{X})$ tal que $\|u\|_\infty \leq 1$ y (5.9) se tiene

$$\begin{aligned}
\left|\widehat{Q}u(x) - \widehat{Q}u(y)\right| &= \left|\int_{\mathbf{X}} Lu(z)Q(dz \mid x,a) - \int_{\mathbf{X}} Lu(z)Q(dz \mid y,b)\right| \\
&= \left|\int_{\mathbf{X}} Lu(z)\left[Q(dz \mid x,a) - Q(dz \mid y,b)\right]\right| \\
&\leq \|Lu\|_\infty \left\|Q(\cdot \mid x,a) - Q(\cdot \mid y,b)\right\|_{VT}, \\
&\leq 2\beta \|u\|_\infty, \quad \text{(monotonia de } L \text{ y ergodicidad de } Q) \\
&\leq 2\beta, \quad \text{(considerando } \|u\|_\infty \leq 1).
\end{aligned}$$

Entonces,

$$\begin{aligned}
\sup_{\|v\|\leq 1} \left|\widehat{Q}u(x) - \widehat{Q}u(y)\right| &\leq 2\beta, \\
\left\|\widehat{Q}(\cdot \mid x) - \widehat{Q}(\cdot \mid y)\right\|_{VT} &\leq 2\beta,
\end{aligned}$$

se llega al resultado deseado. ■

Por el lema anterior, en el modelo $\widehat{\mathfrak{M}}$ también se cumple con las propiedades (2.18)-(2.21) para cada política estacionaria $f \in \mathbf{F}$, donde $\widehat{\mu}_f$ es la medida de probabilidad invariante. En particular, se tiene que

$$\widehat{J}(f) := \widehat{J}_f(x) = \widehat{\mu}_f(C_f) \ \ \forall\ x \in \mathbf{X} \text{ y } f \in \mathbf{F}, \tag{5.10}$$

Proposición 5.1 *Suponemos que en el modelo original $\mathfrak{M}$ se satisfacen las condiciones de las hipótesis 2.1 y 2.3. Entonces, el operador $\widehat{T}_z$ es una contracción del espacio de Banach $(\mathbf{C}_a^0(\mathbf{X}), \|\cdot\|_{sp})$ en sí mismo, con módulo β igual al de T_z.*

Demostración.

Para verificar lo anterior recuerde que

$$\widehat{T}_z u := \widehat{T}u - \widehat{T}u(z) = TLu - TLu(z) = T_z Lu$$

para toda $u \in \mathbf{C}_a(\mathbf{X})$. Entonces, para toda $u, v \in \mathbf{C}_a(\mathbf{X})$ se cumple que

$$\left\|\widehat{T}_z u - \widehat{T}_z v\right\|_{sp} = \|T_z Lu - T_z Lv\|_{sp} \leq \beta \|Lu - Lv\|_{sp} \leq \beta \|u - v\|_{sp}$$

puesto que T_z es una contracción del espacio de Banach $(\mathbf{C}_a^0(\mathbf{X}), \|\cdot\|_{sp})$ y L es no-expansivo en $(\mathbf{C}_a^0(\mathbf{X}), \|\cdot\|_{sp})$. ■

Con lo anterior se genera la siguiente proposición

Proposición 5.2 *Suponemos que en el modelo original $\mathfrak{M}$ se satisfacen las condiciones de las hipótesis 2.1 y 2.3 y L es un promediador continuo. Entonces, existe una terna canónica $(\widehat{\rho}^*, \widehat{h}^*, \widehat{f}^*)$ para $\widehat{\mathfrak{M}}$ donde $\widehat{h}^*$ pertenece a $\mathbf{C}_a(\mathbf{X})$, $\widehat{\rho}^* = \widehat{h}^*(z)$ es el valor óptimo y $\widehat{f}^*$ es una política óptima. Además,*

$$\left\|\widehat{h}_n - \widehat{h}^*\right\|_\infty \leqslant \left\|\widehat{h}_n - \widehat{h}^*\right\|_{sp} \leqslant \beta^n \left\|\widehat{h}^*\right\|_{sp}, \quad n \in \mathbb{N}.$$

Demostración.

Lo anterior se debe a que $\widehat{T}_z$ es contracción de $(\mathbf{M}_a^0(\mathbf{X}), \|\cdot\|_{sp})$ módulo β, entonces por el teorema de punto fijo de Banach existe una única función $\widehat{h}^* \in \mathbf{M}_a^0(\mathbf{X})$ que satisface la ecuación de optimalidad perturbada

$$\widehat{h}^*(x) = \widehat{T}_z\widehat{h}^*(x) = \widehat{T}\widehat{h}^*(x) - \widehat{\rho}^*, \ \forall\, x \in \mathbf{X},$$

donde $\widehat{\rho}^* = \widehat{T}\widehat{h}^*(z)$. Como $T\widehat{h}^* \leqslant T_f\widehat{h}^*$ para toda política estacionaria $f \in \mathbf{F}$, en particular si $u = L\widehat{h}^*$ se tiene $TL\widehat{h}^* \leqslant T_fL\widehat{h}$, es decir, $\widehat{T}\widehat{h}^* \leqslant \widehat{T}_f\widehat{h}^*$ para toda política estacionaria $f \in \mathbf{F}$, entonces

$$\widehat{\rho}^* + \widehat{h}^*(x) \leqslant \widehat{T}_f\widehat{h}^*, \ f \in \mathbf{F}.$$

y con argumentos de programación dinámica estándar se llega a que

$$\widehat{\rho}^* \leqslant \widetilde{J}(f, x), \ \forall\, x \in \mathbf{X}, \ f \in \mathbf{F}.$$

Por otro lado, las condiciones de la hipótesis 2.1 o de la hipótesis 2.2. garantizan la existencia de un selector medible $\widehat{f}^*$ el cual es $\widehat{h}^*-greedy$, es decir,

$$\widehat{\rho}^* + \widehat{h}^*(x) = \widehat{T}\widehat{h}^*(x) = \widehat{T}_{\widehat{f}^*}\ \widehat{h}^*(x)$$

lo cual implica que

$$\widehat{\rho}^* = \widehat{J}_{\widehat{f}^*} = \widehat{J}^*$$

por lo que $(\widehat{\rho}^*, \widehat{h}^*, \widehat{f}^*)$ es una terna canónica.■

La siguiente proposición análoga al teorema (2.4) para $\mathfrak{M}$, se refieren a la convergencia del algoritmo IVA y resuelve el problemas (P1) para $\widehat{\mathfrak{M}}$. En vista del lema 5.1 y de la proposiciones 5.2 la prueba es prácticamente la misma que la prueba del teorema 2.4, por lo tanto será omitida.

Proposición 5.3 *Suponemos que en el modelo original $\mathfrak{M}$ se satisface las condiciones de las hipótesis 2.1 y 2.3 y L es un promediador continuo. Sea $(\widehat{\rho}^*, \widehat{h}^*, \widehat{f}^*)$ una terna canónica para $\widehat{\mathfrak{M}}$ (como en la proposición 5.2) y defina las sucesiones*

$$\widehat{s}_n := \inf_{x\in\mathbf{X}} \widehat{\rho}_n(x) \quad y \quad \widehat{S}_n := \sup_{x\in\mathbf{X}} \widehat{\rho}_n(x), \quad n \in \mathbb{N}.$$

Entonces:

(a) La sucesión $\{\widehat{s}_n\}$ es no-decreciente y $\left\{\widehat{S}_n\right\}$ es no-creciente. Además,

$$-\beta^{n-1}\left\|\widehat{h}^*\right\|_{sp} \leqslant \widehat{s}_n - \widehat{\rho}^* \leqslant \widehat{S}_n - \widehat{\rho}^* \leqslant \beta^{n-1}\left\|\widehat{h}^*\right\|_{sp}, \quad \forall\ n \in \mathbb{N}.$$

En particular, $\widehat{\rho}_n(z)$ converge a $\widehat{\rho}^*$.

(b) Si $f \in \mathbf{F}$ es una política $\widehat{h}_n-greedy$, entonces

$$0 \leqslant \widehat{J}(f) - \widehat{\rho}^* \leqslant \|\widehat{\rho}_n\|_{sp} \leqslant \beta^{n-1}\left\|\widehat{h}^*\right\|_{sp}, \quad \forall\ n \in \mathbb{N}.$$

Esta proposición garantiza la convergencia geométrica del algoritmo IVA para $\widehat{\mathfrak{M}}$ quedando resuelto el problema (P1). Por los mismos argumentos de la observación 2.6 aplicados a $\widehat{\mathfrak{M}}$, el algoritmo IVA (5.3) y el (5.5), son equivalentes y por simplicidad computacional trabajaremos con el primero.

Este algoritmo IVA también es detenidos en alguna etapa $k \in \mathbb{N}$ mediante una regla de paro, obteniendose una política $f \in \mathbf{F}$, $\widehat{J}_k-greedy$ en $\widehat{\mathfrak{M}}$ y el valor óptimo $\widehat{\rho}^*$ es aproximado por medio de la función valor $\widehat{J}(f)$ con una cota del error de aproximación $\|\widehat{\rho}_n\|_{sp}$ por parar en alguna etapa.

Observación 5.2 *Las instrucciones del algoritmo* IVA *para el modelo perturbado $\widehat{\mathfrak{M}}$ son exactamente las mismas instrucciones que las del algoritmo* IV *costo promedio (2.2) para el modelo original $\mathfrak{M}$, unicamente hay que sustituir el operador T por el operador $\widehat{T}$ dado en (5.1).*

5.2.2 Modelo $\widetilde{\mathfrak{M}} = (\mathbf{X}, \mathbf{A}, \left\{\widetilde{C}_f, \widetilde{Q}_f,\ f \in \mathbf{F}\right\})$

En este modelo se considera que el espacio de estados, el espacio de controles, asi como el conjunto de controles admisible permanecen iguales a los del modelo original $\mathfrak{M}$, mientras que el costo en una etapa y la probabilidad de transición perturbada están definidos en (3.11) y (3.12) para la clase de políticas estacionarias $f \in \mathbf{F}$ como

$$\widetilde{C}_f := LC_f,$$

$$\widetilde{Q}_f(B \mid \cdot) := LQ_f(B \mid \cdot) = \int_{\mathbf{X}} Q_f(B \mid y)L(dy \mid \cdot),$$

respectivamente, para cada $B \in \mathcal{B}(\mathbf{X})$, $f \in \mathbf{F}$.

Se define el operadores de programación dinámica como

$$\widetilde{T}u(x) := \inf_{f\in\mathbf{F}} \left\{ \widetilde{C}_f(x) + \int_{\mathbf{X}} u(y)\widetilde{Q}_f(dy \mid x) \right\}, \quad x \in \mathbf{X},\, u \in \mathbf{M}_a(\mathbf{X}). \tag{5.11}$$

También para cada $f \in \mathbf{F}$, se define

$$\widetilde{T}_f u(x) := \widetilde{C}_f(x) + \int_{\mathbf{X}} u(y)\widetilde{Q}_f(dy \mid x), \quad x \in \mathbf{X},\, u \in \mathbf{M}_a(\mathbf{X}), \tag{5.12}$$

donde $\widetilde{Q}_f(dy \mid x) := \widetilde{Q}(dy \mid x, f(x))$. Utilizando (4.8), las definiciones anteriores quedan

$$\widetilde{T}u := \inf_{f\in\mathbf{F}} \left\{ \widetilde{C}_f + \widetilde{Q}_f u \right\} \quad \text{y} \quad \widetilde{T}_f u := \widetilde{C}_f + \widetilde{Q}_f u \quad \forall\, u \in \mathbf{M}_a(\mathbf{X})$$

Observe que estos operadores también son iguales a los de costo descontado con $\alpha = 1$. Entonces, por los mismos argumentos de la proposición 4.2 tomando $\alpha = 1$, se tiene la siguiente observación.

Observación 5.3 *Suponemos que en el modelo original $\mathfrak{M}$ se satisfacen las condiciones de la hipótesis 2.2 y L es un promediador. Entonces, $\widetilde{T} = LT$.*

A continuación siguiendo los mismo pasos realizados en la sub-sección anterior para el modelo $\widehat{\mathfrak{M}}$, se construyen las ecuaciones recursivas del algoritmos IVA para este modelo $\widetilde{\mathfrak{M}}$. Se definen

$$\widetilde{J}_0 := 0 \text{ y } \widetilde{J}_{n+1} = \widetilde{T}\widetilde{J}_n \quad n \in \mathbb{N}. \tag{5.13}$$

Se considera la ecuación de optimalidad

$$\widetilde{\rho} + \widetilde{h} = \widetilde{T}\widetilde{h}, \tag{5.14}$$

donde $\widetilde{\rho}$ es constante y $\widetilde{h} \in \mathbf{M}_a(\mathbf{X})$. Si existe una solución, digamos $(\widetilde{\rho}^*, \widetilde{h}^*)$, entonces existen políticas $\widetilde{f}^* \in \mathbf{F}$ (lo cual se demuestra más adelante), denominadas políticas $\widetilde{h}^*-greedy$, esto es,

$$\widetilde{\rho}^* + \widetilde{h}^* = \widetilde{T}_{\widetilde{f}^*}\widetilde{h}^*.$$

y a $(\widetilde{\rho}^*, \widetilde{h}^*, \widetilde{f}^*)$ se les llama terna canónica.

La solución de la ecuación de optimalidad se obtienen como límite de las sucesión

$$\widetilde{\rho}_{n+1} = \widetilde{J}_{n+1}(z) - \widetilde{J}_n(z), \quad n \in \mathbb{N}_0, \tag{5.15}$$

$$\widetilde{h}_{n+1} = \widetilde{J}_{n+1} - \widetilde{J}_{n+1}(z), \quad n \in \mathbb{N}_0, \tag{5.16}$$

con z un estado arbitrario pero fijo.

Estas sucesiones están relacionadas como sigue

$$\widetilde{\rho}_{n+1} + \widetilde{h}_{n+1}(\cdot) = \widetilde{T}\widetilde{h}_n(\cdot),, \quad n \in \mathbb{N}_0,$$

con $\widetilde{h}_0 \equiv 0$.

Note que, $\widetilde{\rho}_{n+1} = \widetilde{T}\widetilde{h}_n(z)$ para todo n$\in \mathbb{N}_0$, porque $\widetilde{h}_{n+1}(z) = 0$. Entonces,

$$\widetilde{h}_{n+1} = \widetilde{T}\widetilde{h}_n - \widetilde{T}\widetilde{h}_n(z), \quad n \in \mathbb{N}_0,$$

definiendo el operador $\widetilde{T}_z$ para cada $u \in \mathbf{M}_a(\mathbf{X})$ como

$$\widetilde{T}_z u := \widetilde{T}u - \widetilde{T}u(z), \tag{5.17}$$

entonces, se obtienen las otras ecuaciones recursivas del algoritmo IVA

$$\widetilde{h}_{n+1} = \widetilde{T}_z \widetilde{h}_n, \quad n \in \mathbb{N}_0, \tag{5.18}$$

que al considerar que $\widetilde{T} = LT$, se tiene explícitamente

$$\widetilde{h}_{n+1} := LT\widetilde{h}_n - LT\widetilde{h}_n(z), \quad n \in \mathbb{N}_0$$

que son las referidas en la introducción por (9).

En el siguiente lema se demuestra la condición de ergodicidad.

Lema 5.2 *Suponemos que en el modelo original $\mathfrak{M}$ se satisface la condición de ergodicidad (hipótesis 2.3). Entonces, en $\widetilde{\mathfrak{M}}$ también se satisface, es decir:*

$$\left\| \widetilde{Q}_f(\cdot \mid x) - \widetilde{Q}_g(\cdot \mid y) \right\|_{VT} \leqslant 2\beta, \quad \forall\ x,\ y \in \mathbf{X}, \quad f, g \in \mathbf{F},$$

β es la constante de la condición de ergodicidad en el modelo original $\mathfrak{M}$.

Demostración.

Como en $\mathfrak{M}$ se satisface la condición de ergodicidad (hipótesis 2.3), entonces para todas las políticas estacionaria f, $g \in \mathbf{F}$, se tiene que

$$\|Q_f(\cdot \mid x) - Q_g(\cdot \mid y)\|_{VT} \leqslant 2\beta \ \ \forall\ x,\ y \in \mathbf{X},$$

lo cual implica

$$|Q_f(B \mid x) - Q_g(B \mid y)| \leqslant \beta \ \ \forall\ x,\ y \in \mathbf{X},\ B \in \mathcal{B}(\mathbf{X}).$$

Entonces,

$$-\beta + Q_g(B \mid y) \leq Q_f(B \mid x) \leq Q_g(B \mid y) + \beta \ \ \forall\ x,\ y \in \mathbf{X},\ B \in \mathcal{B}(\mathbf{X}).$$

Tomando $y \in \mathbf{X}$ fija, al aplicar L a esta doble desigualdad y utlizar su propiedad de monotonia y linealidad, se tiene

$$-\beta + LQ_g(B \mid y) \leq LQ_f(B \mid x) \leq LQ_g(B \mid y) + \beta \ \ \forall\ x \in \mathbf{X},\ B \in \mathcal{B}(\mathbf{X}).$$

Con los mismos argumentos, pero ahora fijando $x \in \mathbf{X}$, resulta

$$-\beta + LQ_g(B \mid y) \leq LQ_f(B \mid x) \leq LQ_g(B \mid y) + \beta, \quad \forall\ y \in \mathbf{X},\ B \in \mathcal{B}(\mathbf{X}).$$

Entonces,

$$-\beta + \widetilde{Q}_g(B \mid y) \leq \widetilde{Q}_f(B \mid x) \leq \widetilde{Q}_g(B \mid y) + \beta, \ \ \forall\, x,\ y \in \mathbf{X},\ B \in \mathcal{B}(\mathbf{X}),$$

de donde se deduce que

$$\sup_{B \in \mathcal{B}(\mathbf{X})} \left| \widetilde{Q}_f(B \mid x) - \widetilde{Q}_g(B \mid y) \right| \leqslant \beta.$$

Como $f,\ g \in \mathbf{F}$ son arbitrarios, se llega al resultado deseado.■

Por el lema anterior en $\widetilde{\mathfrak{M}}$ también se cumple con las propiedades (2.18)-(2.21) para cada política estacionaria $f \in \mathbf{F}$, donde $\widetilde{\mu}_f$ es la medida de probabilidad invariante. En particular, se tiene que

$$\widetilde{J}(f) := \widetilde{J}_f(x) = \widetilde{\mu}_f(\widetilde{C}_f) \ \ \forall\, x \in \mathbf{X} \text{ y } f \in \mathbf{F}. \tag{5.19}$$

Los mismos argumentos dados en la proposición 5.1, justifican la siguiente proposición.

Proposición 5.4 *Suponga que en el modelo original $\mathfrak{M}$ se satisface las condiciones de las hipótesis 2.2 y 2.3. Entonces, el operador $\widetilde{T}_z$ es una contracción del espacio de Banach $(\mathbf{M}_a^0(\mathbf{X}), \|\cdot\|_{sp})$ en sí mismo, con módulo β igual al de T_z.*

Con argumentos análogos a los de la proposition 5.2, se obtiene el siguiente resultado para $\widetilde{\mathfrak{M}}$.

Proposición 5.5 *Suponemos que en el modelo original $\mathfrak{M}$ se satisfacen las condiciones de las hipótesis 2.2 y 2.3. Entonces, existe una terna canónica $(\widetilde{\rho}^*, \widetilde{h}^*, \widetilde{f}^*)$ para $\widetilde{\mathfrak{M}}$, con $\widetilde{h}^* \in \mathbf{M}_a(\mathbf{X})$ y $\widetilde{\rho}^* = \widetilde{h}^*(z)$ es el valor óptimo y $\widetilde{f}^*$ es la política óptima. Además,*

$$\left\| \widetilde{h}_n^* - \widetilde{h}^* \right\|_\infty \leqslant \left\| \widetilde{h}_n^* - \widetilde{h}^* \right\|_{sp} \leqslant \beta^n \left\| \widetilde{h}^* \right\|_{sp}, \quad n \in \mathbb{N}.$$

La siguiente proposición se refieren a la convergencia del algoritmo IVA y se resuelve el problema (P1) para $\widetilde{\mathfrak{M}}$. En vista de la proposición 5.5, las, pruebas es prácticamente la misma que la prueba del teorema 2.4, por lo tanto será omitida.

Proposición 5.6 *Suponemos que en el modelo original $\mathfrak{M}$ se satisfacen las condiciones de las hipótesis 2.2 y 2.3. Sea $(\widetilde{\rho}^*, \widetilde{h}^*, \widetilde{f}^*)$ una terna canónica para $\widetilde{\mathfrak{M}}$ (como en la proposición 5.5) y defina las sucesiones*

$$\widetilde{s}_n := \inf_{x \in \mathbf{X}} \widetilde{\rho}_n(x) \quad y \quad \widetilde{S}_n := \sup_{x \in \mathbf{X}} \widetilde{\rho}_n(x), \quad n \in \mathbb{N}.$$

Entonces, las siguientes propiedades se cumplen para todo $n \in \mathbb{N}$:

(a) La sucesión $\{\widetilde{s}_n\}$ es no-decreciente y $\left\{\widetilde{S}_n\right\}$ es no-creciente. Además,

$$-\beta^{n-1} \left\| \widetilde{h}^* \right\|_{sp} \leqslant \widetilde{s}_n - \widetilde{\rho}^* \leqslant \widetilde{S}_n - \widetilde{\rho}^* \leqslant \beta^{n-1} \left\| \widetilde{h}^* \right\|_{sp}, \ \ \forall\, n \in \mathbb{N}.$$

En particular, $\widetilde{\rho}_n(z)$ converge a $\widetilde{\rho}^*$.

(b) Si $f \in \mathbf{F}$ es una política $\widetilde{h}_n-greedy$, entonces

$$0 \leqslant \widetilde{J}(f) - \widetilde{\rho}^* \leqslant \|\widetilde{\rho}_n\|_{sp} \leqslant \beta^{n-1} \left\|\widetilde{h}^*\right\|_{sp}, \quad \forall\ n \in \mathbb{N}.$$

Ambas proposiciones garantizan la convergencia geométrica del algoritmo IVA para $\widetilde{\mathfrak{M}}$ quedando resuelto el problema (P1). Por los mismos argumentos de la observación 2.6 aplicados a $\widetilde{\mathfrak{M}}$, el algoritmos IVA (5.3) con (5.8) y el algoritmo (5.13) con (5.18), son equivalentes y por simplicidad computacional trabajaremos con el primero.

Este algoritmo IVA también es detenido en alguna etapa $k \in \mathbb{N}$ mediante una regla de paro, obteniendose una política $f \in \mathbf{F}$, $\widetilde{J}_k-greedy$ en $\widetilde{\mathfrak{M}}$. Además, el valor óptimo $\widetilde{\rho}^*$ es aproximado por medio de la función valor $\widetilde{J}(f)$, con una cota de error de aproximación $\|\widetilde{\rho}_n\|_{sp}$ por parar en alguna etapa.

Observación 5.4 *Las instrucciones del algoritmo* IVA *para el modelo perturbado* $\widetilde{\mathfrak{M}}$, *son exactamente las mismas instrucciones que las del algoritmo* IV *costo promedio (2.2) para el modelo original* $\mathfrak{M}$, *unicamente hay que sustituir el operador T por el operador $\widetilde{T}$ dado en (5.11).*

5.3 Cotas de aproximación del algoritmo IVA

En esta sección se resuelven los problemas (P2) y (P3) planteados desde la introducción, con la excepción de que las cotas para $\widetilde{h}^* - h^*$ y $\widehat{h}^* - h^*$ no son obtenidas. Se obtienen las cotas de aproximación de los algoritmos IVA para costo promedio para ambos modelos perturbados. Dichas cotas son expresadas en términos de la constante de aproximación $\delta_Q(\widehat{\mathbf{F}}_0)$ para $\widehat{\mathfrak{M}}$ dada en (3.19) y de las constantes de aproximación $\delta_C(\widetilde{\mathbf{F}}_0)$ y $\delta_Q(\widetilde{\mathbf{F}}_0)$ para $\widetilde{\mathfrak{M}}$ dadas en (3.20) y (3.21), respectivamente. Lo anterior es expresado en los siguientes teoremas.

Lema 5.3 *Suponemos que en el modelo original $\mathfrak{M}$ se satisfacen las condiciones de las hipótesis 2.3 y 2.1 y L es un promediador continuo. Sean $f \in \mathbf{F}$ una política estacionaria arbitraria y M la cota de la función de costo por etapa C. Entonces,*

$$\left|\widehat{J}(f) - J(f)\right| \leq \frac{M}{1-\alpha} \sup_{x \in \mathbf{X}} \left\|\widehat{Q}_f(\cdot \mid x) - Q_f(\cdot \mid x)\right\|_{VT}.$$

Demostración.

Se sigue de las propiedades (2.21), (5.10) y (5.19) que

$$\left|\widehat{J}(f) - J(f)\right| = |\widehat{\mu}_f(C_f) - \mu_f(C_f)| \leq M \|\widehat{\mu}_f - \mu_f\|_{VT} \tag{5.20}$$

Luego, en el teorema 2.3, al tomar $S(\cdot \mid \cdot) = Q_f(\cdot \mid \cdot)$ y $T(\cdot \mid x) = \widehat{Q}_f(\cdot \mid \cdot)$, se tiene que

$$\theta := \sup_{x \in \mathbf{X}} \left\|Q_f(\cdot \mid x) - \widehat{Q}_f(\cdot \mid x)\right\|_{VT}$$

$$\gamma := \frac{1}{2} \sup_{x,y \in \mathbf{X}} \|Q_f(\cdot \mid x) - Q_f(\cdot \mid y)\|_{VT} \leq \beta, \quad \text{(por la hipótesis 2.3).}$$

Entonces,

$$\|\mu_f - \widehat{\mu}_f\|_{VT} \leq \frac{1}{1-\beta} \sup_{x \in \mathbf{X}} \left\| Q_f(\cdot \mid x) - \widehat{Q}_f(\cdot \mid x) \right\|_{VT},$$

lo cual combinado con (5.20) implica el resultado deseado. ■

El resultado análogo para el modelo $\widetilde{\mathfrak{M}}$ es como sigue.

Lema 5.4 *Suponemos que en el modelo original $\mathfrak{M}$ se satisfacen las condiciones de las hipótesis 2.3 y 2.2 y L es un promediador. Sean $f \in \mathbf{F}$ una política estacionaria arbitraria y M la cota de la función de costo por etapa C. Entonces,*

$$\left| \widetilde{J}(f) - J(f) \right| \leq \left\| \widetilde{C}_f - C_f \right\|_\infty + \frac{M}{1-\alpha} \sup_{x \in \mathbf{X}} \left\| \widetilde{Q}_f(\cdot \mid x) - Q_f(\cdot \mid x) \right\|_{VT}.$$

Demostración.

De manera similar al lema anterior, note que

$$\begin{aligned} \left| \widetilde{J}(f) - J(f) \right| &= \left| \widetilde{\mu}_f(\widetilde{C}_f) - \mu_f(C_f) \right| \\ &\leq \left| \widetilde{\mu}_f(\widetilde{C}_f) - \mu_f(\widetilde{C}_f) \right| + \left| \mu_f(\widetilde{C}_f) - \mu_f(C_f) \right| \\ &\leq \left\| \widetilde{C}_f - C_f \right\|_\infty + M \left\| \widetilde{\mu}_f - \mu_f \right\|_{VT} \end{aligned}$$

Usando este resultado junto con el del teorema 2.3 al tomar $S = Q_f$ y $T(\cdot \mid x) = Q_f(\cdot \mid \cdot)$ se obiene el resultado deseado.■

Con los resultados anteriores procedemos a plantear los otros resultados principales de esta tesis.

Teorema 5.1 *Suponemos que en el modelo original $\mathfrak{M}$ se satisfacen las condiciones de las hipótesis 2.1 y 2.3, además L es un promediador continuo.* E*ntonces:*

(a) La cota que resuelve el problema (P2) para el modelo $\widehat{\mathfrak{M}}$, es la siguiente,

$$|\widehat{\rho}^* - \rho^*| \leqslant \frac{M}{1-\beta} \delta_Q(\widehat{\mathbf{F}}_0).$$

(b) La cota que resuelve el problema (P3) para el modelo $\widehat{\mathfrak{M}}$, si $f \in \mathbf{F}$ es una política $\widehat{h}_n-$greedy, es la siguiente,

$$0 \leqslant J(f) - \rho^* \leqslant \|\widehat{\rho}_n\|_{sp} + \frac{2M}{1-\beta} \delta_Q(\widehat{\mathbf{F}}_0). \tag{5.21}$$

*Al primer término del lado derecho le llamaremos **cota del error de paro** y al segundo **cota de error de aproximación**.*

Demostración.

(a) Esta parte se sigue usando el resultado del lema 5.3, puesto que

$$\begin{aligned}|\widehat{\rho}^* - \rho^*| &= \left| \inf_{f\in\widehat{\mathbf{F}}_0} \widehat{J}_f(x) - \inf_{f\in\widehat{\mathbf{F}}_0} J_f(x) \right| \\ &\leq \sup_{f\in\widehat{\mathbf{F}}_0} \left| \widehat{J}_f(x) - J_f(x) \right| \\ &\leq \sup_{f\in\widehat{\mathbf{F}}_0} \left[\frac{M}{1-\alpha} \sup_{x\in\mathbf{X}} \left\| \widehat{Q}_f(\cdot \mid x) - Q_f(\cdot \mid x) \right\|_{VT} \right] \\ &\leq \frac{M}{1-\beta} \delta_Q(\widehat{\mathbf{F}}_0).\end{aligned}$$

(b) Para la demostración de esta parte observemos que

$$0 \leq J(f) - \rho^* \leqslant \left| J(f) - \widehat{J}_f(x) \right| + \left| \widehat{J}_f(x) - \widehat{\rho}^* \right| + |\widehat{\rho}^* - \rho^*| \,.$$

Usando la parte (a), el lema 5.3 y la proposición 5.3 (b) se llega al resultado deseado. ■

El resultado corrrespondiente para el modelo $\widetilde{\mathfrak{M}}$ es como sigue.

Teorema 5.2 *Suponemos que en el modelo original $\mathfrak{M}$ se satisfacen las condiciones de la hipótesis 2.2 y las de la hipótesis 2.3. Entonces:*

(a) La cota que resuelve el problema (P2) para el modelo $\widetilde{\mathfrak{M}}$ es la siguiente,

$$|\widetilde{\rho}^* - \rho^*| \leqslant \delta_C(\widetilde{\mathbf{F}}_0) + \frac{M}{1-\beta} \delta_Q(\widetilde{\mathbf{F}}_0) \,|.$$

(b) La cota que resuelve el problema (P3) para el modelo $\widetilde{\mathfrak{M}}$, si $f \in \mathbf{F}$ es una política $\widetilde{h}_n-greedy$, es la siguiente,

$$0 \leqslant J(f) - \rho^* \leqslant \|\widetilde{\rho}_n\|_{sp} + 2\delta_C(\widetilde{\mathbf{F}}_0) + \frac{2M}{1-\beta} \delta_Q(\widetilde{\mathbf{F}}_0). \tag{5.22}$$

*Al primer término del lado derecho le llamaremos **cota del error de paro** y a la suma de los términos restantes **cota de error de aproximación**.*

Demostración. La demostración de este teorema sigue argumentos similares a los de la demostración del teorema anterior.

(a) Usando el resultado del lema 5.4, se tiene que

$$\begin{aligned}
|\widetilde{\rho}^* - \rho^*| &= \left| \inf_{f \in \widetilde{F}_0} \widetilde{J}_f(x) - \inf_{f \in \widetilde{F}_0} J_f(x) \right| \\
&\leq \sup_{f \in \widetilde{F}_0} \left| \widetilde{J}_f(x) - J_f(x) \right| \\
&\leq \sup_{f \in \widetilde{F}_0} \left[\left\| \widetilde{C}_f - C_f \right\|_\infty + \frac{M}{1-\alpha} \sup_{x \in \mathbf{X}} \left\| \widetilde{Q}_f(\cdot \mid x) - Q_f(\cdot \mid x) \right\|_{VT} \right] \\
&\leq \delta_C(\widetilde{\mathbf{F}}_0) + \frac{M}{1-\beta} \delta_Q(\widetilde{\mathbf{F}}_0).
\end{aligned}$$

(b) Para demostrar esta parte observemos que

$$0 \leq J(f) - \rho^* \leqslant \left| J(f) - \widetilde{J}_f(x) \right| + \left| \widetilde{J}_f(x) - \widetilde{\rho}^* \right| + |\widetilde{\rho}^* - \rho^* \,|.$$

Entonces, usando la parte (a), el lema 5.4 y la proposición 5.6 (b) se llega al resultado deseado.■

Capítulo 6

Aproximación sistemas de inventarios

6.1 Introducción

En este capítulo se ilustran los resultados del presente trabajo con un sistema de inventarios que bajo ciertas condiciones cumple con las hipótesis de los MCM desarrollados.

Considerando la existencia de una política óptima de umbral para el sistema de inventarios, se obtienen las constantes de aproximación y las cotas de aproximación del algoritmo IVA tanto para costo descontado como para costo promedio. Se finaliza este capítulo, con un reporte de los resultados numéricos de la implementación computacional del sistema de inventarios con demanda exponencial y para contraste de resultados se simula la cadena generada por la pólitica *greedy* obtenida.

6.2 Sistema de inventarios

Consideraremos un sistema de inventarios de un sólo producto con capacidad finita $\theta > 0$ tanto para capacidad de inventario como de producción, en el cual se pierde la demanda que no se satisface. Denotaremos por x_t y a_t al nivel de inventario y a la cantidad de producto que se ordena a producción al inicio de la etapa $t \in \mathbb{N}_0$, respectivamente, y por w_t a la demanda de dicho producto durante el mismo periodo. Suponemos que la cantidad de producto ordenada a_t es inmediatamente abastecida. Entonces, los niveles de inventario evolucionan de acuerdo al sistema recursivo

$$x_{t+1} = (x_t + a_t - w_t)^+, \quad t \in \mathbb{N}_0, \tag{6.1}$$

donde $x_0 = x \in \mathbf{X}$ es el nivel de inventario inicial dado y $r^+ = \max(0, r)$ para cada $r \in \mathbb{R}$.

También suponemos que el proceso de demandas $\{w_t\}$ está formado por variables aleatorias no-negativas, independientes e idénticamente distribuidas, con función de distribución común $G(\cdot)$ y que estas variables aleatorias además son independientes del inventario inicial $x_0 = x$.

Puesto que el sistema tiene capacidad θ, los procesos $\{x_t\}$ y $\{a_t\}$ toman valores en $\mathbf{X} = \mathbf{A} = [0, \theta]$. Además, para cada estado $x \in \mathbf{X}$ el conjunto de controles admisibles es $A(x) = [0, \theta - x]$. Note que el conjunto de pares admisibles queda $\mathbf{K} = \{(x, a) \in \mathbf{X} \times \mathbf{A} : x \in \mathbf{X},\ a \in [0, \theta - x]\}$.

La dinámica del sistema de inventario (6.1), puede expresarse equivalentemente por medio de la probabilidad de transición

$$Q(B \mid x, a) := E_{w_0} I_B(\ (x + a - w_0)^+), \quad B \in \mathcal{B}(\mathbf{X}),\ (x, a) \in \mathbf{K}, \tag{6.2}$$

donde E_{w_0} denota la esperanza con respecto a la función de distribución G de la variable aleatoriaria w_0. Observe que

$$Q(B \mid x, a) = I_B(0)\,[1 - G(x + a)] + \int_0^{x+a} I_B(x + a - w) G(dw). \tag{6.3}$$

Para completar la descripción del sistema de inventario solo hace falta especificar la función de costo por etapa $C : \mathbf{K} \to \mathbb{R}$. En general tendrá la siguiente forma:

$$C(x, a) = ca + R(x + a),\ (x, a) \in \mathbf{K} \tag{6.4}$$

donde c es una constante positiva y $R : \mathbf{X} \to \mathbb{R}$ es una función Borel medible no-negativa en $\mathbf{X}$. Usualmente la constante c se interpreta como el costo unitario de producción y la función R como un costo de penalización por el exceso o déficit de inventario en cada etapa.

En este trabajo requerimos que la función de costo $C : \mathbf{K} \to \mathbb{R}$ sea continua. Para esto basta que se cumpla con la siguiente condición.

Hipótesis 6.1 *En (6.4) la función costo de penalización $R : \mathbf{X} \to \mathbb{R}$ es continua.*

Un caso particular resulta al considerar una función $p : \mathbb{R}_+ \to \mathbb{R}_+$ que penaliza el déficit de inventario y una función $h : \mathbb{R}_+ \to \mathbb{R}_+$ que penaliza los inventarios positivos, en cuyo caso se tiene que

$$R(y) := h(y) + E_{w_0} p((w_0 - y)^+), \quad y \in \mathbb{R}_+.$$

Suponiendo que $E_{w_0} p((w_0 - y)^+)$ finita para toda $y \in \mathbf{X}$ y que las funciones p y h son continuas, se garantiza la continuidad de R.

Más concretamente, si las funciones son lineales, es decir, $p(y) := py$ y $h(y) := hy$, $y \in \mathbb{R}_+$, donde p y h son constantes positivas, tendremos que

$$R(y) := hy + pE_{w_0}(w_0 - y)^+, \quad y \in \mathbb{R}_+. \tag{6.5}$$

Aquí, solamente con suponer que la demanda esperada $\overline{w} := E_{w_0}(w_0)$ es finita, se garantiza la continuidad de R. En este caso las constantes h y p representan respectivamente el costo unitario por manejo de inventario y la penalización por cada unidad de déficit.

La función de costo en una etapa (6.4) con esta último costo de penalización R, queda

$$C(x, a) = ca + h(x + a) + pE_{w_0}(w_0 - x - a)^+,$$

para todo $(x, a) \in \mathbf{K}$. Siguiendo argumentos elementales, razón por la cual se omiten, se llega a

$$C(x,a) = p\overline{w} + ca + (h-p)(x+a) + p\int_0^{x+a} G(w)dw. \tag{6.6}$$

En las secciones siguientes estudiaremos los problemas de control óptimo en costo descontado y en costo promedio para el sistema de inventario (6.1) y función de costo por etapa (6.4), o alguno de estos casos particulares.

6.3 Existencia de políticas óptimas

Mostraremos que el sistema de inventarios de la sección anterior bajo condiciones muy generales satisface las condiciones de las hipótesis 2.1, 2.2 y 2.3. Para lo anterior se necesita de las siguientes definiciones.

Definición 6.1 *Sean S y S' espacios topológicos y $P(S')$ el conjunto potencia de S'.*

*(a) Un mapeo de valor conjustista de S a S' es un mapeo $\psi : S \to P(S')$. A este mapeo se le denomina **correspondencia** ó **multifunción**;*

*(b) Una multifunción es **semicontinua superiormente** si $\{s \in S: \psi(s) \cap F \neq \theta\}$ es un subconjunto cerrado en S para cada $F \subset S'$ cerrado.*

*(c) Una multifunción es **semicontinua inferiormente** si $\{s \in S: \psi(s) \cap F \neq \theta\}$ es un subconjunto abierto en S para cada $F \subset S'$ abierto.*

*(d) Una multifunción es **continua** si es semicontinua superiormente e inferiormente.*

Proposición 6.1 *Suponemos que el sistema de inventarios (6.1) con función función de costo por etapa (6.4) satisface las condiciones de la hipótesis 6.1. Entonces, el sistema de inventarios satisface las condiciones de la hipótesis 2.1.*

Demostración.

(a) Por la hipótesis 6.1 la función R es continua, por lo cual la función de costo por etapa C es continua en $\mathbf{K}$. Note que, C es acotada en $\mathbf{K}$, puesto que este conjunto es compacto.

(b) $\mathbf{A}(x) = [0, \theta - x]$ es un suconjunto de $\mathbf{A} = [0, \theta]$ no-vacío y compacto para cada $x \in \mathbf{X}$. Ahora verificaremos que el mapeo $x \to \mathbf{A}(x) = [0, \theta - x]$ es continuo. Primero, considere un subconjunto cerrado F de $\mathbf{A} = [0, \theta]$. Como $\mathbf{A}$ es compacto, también F es compacto; entonces, $a^1 := minF$ y $a^2 := maxF$ están bien definidos y pertenecen a F. Note que el conjunto

$$\{x \in \mathbf{X} : \mathbf{A}(x) \cap F \neq \theta\} = [0, \theta - a^1]$$

es un subconjunto cerrado en $\mathbf{X}$. Esto prueba que el mapeo es semicontinuo superiormente.
Ahora considere el intervalo abierto $(a^3, a^4) \subset \mathbf{A}$. Entonces,

$$\{x \in \mathbf{X} : \mathbf{A}(x) \cap (a^3, a^4) \neq \theta\} = [0, \theta - a^3)$$

es un subconjunto abierto en $\mathbf{X}$. Esto prueba que el mapeo es semicontinuo inferiormente.
Por lo tanto, el mapeo $x \to \mathbf{A}(x) = [0, \theta - x]$ es continuo.

(c) Ahora verificaremos la continuidad debil en $\mathbf{K}$ de la probabilidad de transición (6.2)

$$Q(B \mid x, a) = E_{w_0} I_B(\ (x + a - w_0)^+),$$

es decir, que el mapeo

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

es continuo para cada función $u \in \mathbf{C}_a(\mathbf{X})$. Esta propiedad es consecuencia directa de la igualdad

$$\int_{\mathbf{X}} v(y) Q(dy \mid x, a) = E_{w_0} v(\ (x + a - w_0)^+)\ \forall (x, a) \in \mathbf{K},\ v \in \mathbf{C}_a(\mathbf{X}), \quad (6.7)$$

y el teorema de convergencia acotada.■

Para garantizar que el sistema de inventarios satisface las condiciones de la hipótesis 2.2 supondremos que se cumple la siguiente condición.

Hipótesis 6.2 *La función de distribución G de las demandas tiene una densidad continua $\rho(\cdot)$ acotada por M' (vease hipótesis 6.4).*

Proposición 6.2 *Suponemos que el sistema de inventarios (6.1) con función función de costo por etapa (6.4) satisface las condiciones de las hipótesis 6.1 y 6.2. Entonces, el sistema de inventarios satisface las condiciones de la hipótesis 2.2.*

Demostración.

(a) Por la hipótesis 6.1 la función R es continua, por lo cual la función de costo por etapa C es continua en el compacto $\mathbf{K}$; entonces C es acotada en $\mathbf{K}$. En particular, $C(x, \cdot)$ es una función continua en $\mathbf{A}(x)$.

(b) $\mathbf{A}(x) = [0, \theta - x]$ es un suconjunto de $\mathbf{A} = [0, \theta]$ no-vacío y compacto para cada $x \in \mathbf{X}$.

(c) La continuidad fuerte de $Q(\cdot \mid x, \cdot)$ en $\mathbf{A}(x)$, se deduce de la igualdad

$$\int_{\mathbf{X}} v(y) Q(dy \mid x, a) = v(0)\,[1 - G(x + a)] + \int_0^{x+a} v(z) \rho(x + a - w) dw, \quad (6.8)$$

la cual se cumple para toda función $v \in \mathbf{C}_a(\mathbf{X})$ y el teorema de convergencia dominada.

Por lo tanto bajo estos supuestos adicionales, tenemos que el sistema de inventarios satisface la hipótesis 2.2. ■

Para que el sistema satisfaga la hipótesis 2.3 (condición de ergodicidad), supondremos que se satisface la siguiente condición.

Hipótesis 6.3 *La función de distribución G de las demandas cumple que $G(\theta) < 1$, donde θ es la capacidad del sistema.*

Proposición 6.3 *Suponemos que se satisface la hipótesis 6.3. Entonces, el sistema de inventarios satisface la hipótesis 2.3 (condición de ergodicidad).*

Demostración.

Para checar la condición de ergodicidad, recuerde que la probabilidad de transición dada por (6.8) es

$$Q(B \mid x, a) = I_B(0)\,[1 - G(x + a)] + \int_0^{x+a} I_B(x + a - w)G(dw).$$

Tomando $B = \{0\}$ se tiene que

$$\begin{aligned} Q(\{0\} \mid x, a) &= 1 - G(x + a) \\ &\geqslant 1 - G(\theta) \ \ \forall\ (x, a) \in \mathbf{K}. \end{aligned}$$

Entonces,

$$Q(0, \theta] \mid x, a) \leq G(\theta) \ \ \forall\ (x, a) \in \mathbf{K}$$

En general, se cumple que

$$Q(B \mid x, a) \leq G(\theta) \ \ \forall\ (x, a) \in \mathbf{K}, \text{ si } 0 \in B,$$

y

$$Q(B \mid x, a) \geqslant 1 - G(\theta)\ \forall\ (x, a) \in \mathbf{K}, \text{ si } B \subseteq (0, \theta].$$

Entonces,

$$|Q(B \mid x, a) - Q(B \mid x\prime, a\prime)| \leq G(\theta) \ \ \forall\ (x, a),\ (x\prime, a\prime) \in K,\ \ B \in B(X).$$

Por la caracterización de la norma de variación [10, B.2, p.125] se tiene que

$$\|Q(\cdot \mid x, a) - Q(\cdot \mid x\prime, a\prime)\| \leq 2G(\theta) \ \ \forall\ (x, a),\ (x\prime, a\prime) \in K,$$

lo cual implica que se cumple la condición de ergodicidad de la hipótesis 2.3 con

$$\beta = G(\theta) < 1.\blacksquare \tag{6.9}$$

Concluimos esta sección, enfatizando que si el sistema de inventarios (6.1) con función función de costo por etapa (6.4) satisface las condiciones de las hipótesis 6.1, 6.2 y 6.3, entonces cumple con los supuestos de las hipótesis 2.1, 2.2 y 2.3. Por lo cual, a este sistema de inventarios se le puede aplicar los resultados de este trabajo.

6.4 Políticas de umbral

Una política de umbral $S \geqslant 0.$ es una política de la forma

$$f_S(x) = \begin{cases} S - x & \text{si } 0 \leq x \leq S \\ 0 & \text{si } x > S. \end{cases} . \tag{6.10}$$

Para el el sistema de inventarios (6.1), bajo las hipótesis 6.1 y 6.2, existen políticas de umbral óptimo tanto para el índice en costo descontado como en costo promedio. Estos resultados pueden verificarse usando argumentos similares a los dados en [29] y se enuncian a continuación.

Proposición 6.4 *Considere el sistema de inventarios (6.1) con función de costo por etapa dada en (6.4) y (6.5). Suponga que la demanda esperada* $\overline{w} := E_{w_0}(w_0)$ *es finita y que se satisfacen las hipótesis 6.1 y 6.2.*

(a) Entonces, existe una política de umbral $f_{S_\alpha^*}$ *óptima en costo* $\alpha -$ *descontado. Además,* S_α^* *satisface la ecuación*

$$G(S_\alpha^*) = \frac{p - h - c}{p - \alpha c}, \tag{6.11}$$

si $p > c + h$. *En caso contrario* $S_\alpha^* = 0$.

(b) Si adicionalmente se satisface la hipótesis 6.3, entonces existe una política de umbral f_{S^*} *óptima en costo promedio. Además,* S^* *satisface la ecuación*

$$G(S^*) = \frac{p - h - c}{p - c}. \tag{6.12}$$

si $p > c + h$. *En caso contrario* $S^* = 0$.

6.5 Constantes de aproximación

Las cotas para las soluciones aproximadas que arroja el algoritmo de iteracion de valores se dieron en los teoremas 4.1 y 4.2 para el problema en costo descontado y en los teoremas 5.1 y 5.2 para el problema en costo promedio. Recuerde que dichas cotas están expresadas en térmios de las cantidades $\delta_Q(\widehat{\mathbb{F}}_0)$, $\delta_C(\widetilde{\mathbb{F}}_0)$ y $\delta_Q(\widetilde{\mathbb{F}}_0)$. En esta sección se proporcionan cotas para estas cantidades en términos de la malla cuando se usa interpoladores lineales para el problema de inventario descrito en las secciones previas.

Consideremos entonces una malla para el espacio de estados $X = [0, \theta]$ con nodos $m_1 = 0 < m_2 < \cdots < m_{k-1} < m_k = \theta$. Tomemos $\Delta m_{i+1} := m_{i+1} - m_i$ para $i = 0, \ldots, k-1$, y $\Delta := \sup_i \Delta m_{i+1}$. Para cada $v \in M_a(\mathbf{X})$ defina

$$Lv(x) := b_i(x)v(m_i) + (1 - b_i(x))v(m_{i+1}), \tag{6.13}$$

donde

$$b_i(x) := \frac{x - m_i}{\Delta m_{i+1}}, \quad \text{y} \quad \overline{b}_i(x) := 1 - b_i(x),$$

para $x \in [m_i, m_{i+1}], i = 0, \ldots, k-1$. Observe que el operador L es un promediador continuo.

6.5.1 Modelo $\widehat{\mathfrak{M}} = (\mathbf{X}, \mathbf{A}, \{\mathbf{A}(x);\ x \in \mathbf{X}\}, C, \widehat{Q})$

En este modelo únicamente se perturba la transición del sistema, por lo que no haremos hipótesis sobre la función de costo $C(\cdot, \cdot)$ y tomaremos $\widehat{\mathbb{F}}_0 = \mathbb{F}$. La ley de transición perturbada es

$$\widehat{Q}(B|x,a) = \int_{\mathbf{X}} L_B(y) Q(dy|x,a) \ \ \forall (x,a) \in \mathbb{K}.$$

Proposición 6.5 *Sea G la distribución de la demanda. Entonces,*

$$\delta_Q(\widehat{\mathbf{F}}_0) = 2G(\theta). \tag{6.14}$$

Demostración.

Definamos $B_1 := \mathbf{X} \backslash \{m_i\}_{i=1}^k$ y observe que $\widehat{Q}(B_1 \mid x,a) = 0$, $Q(B_1 \mid x,a) = G(x+a)$, para todo $(x,a) \in \mathbf{K}$. Entonces,

$$\begin{aligned} \left\| \widehat{Q}(\cdot \mid x,a) - Q(\cdot \mid x,a) \right\|_{VT} &= 2 \sup_{B \in \mathcal{B}(\mathbf{X})} \left| \widehat{Q}(\cdot \mid x,a) - Q(\cdot \mid x,a) \right| \\ &= 2 \left| \widehat{Q}(B_1 \mid x,a) - Q(B_1 \mid x,a) \right| \\ &= 2G(x+a) \ \ \forall\ (x,a) \in \mathbf{K}, \end{aligned}$$

de donde se sigue (6.14). ■

Se puede observar que $\delta_Q(\widehat{\mathbf{F}}_0)$ es computable, pero no depende de la malla, lo cual es una gran desventaja en términos de aproximación.

6.5.2 Modelo $\widetilde{\mathfrak{M}} = (\mathbf{X}, \mathbf{A}, \left\{\widetilde{C}_f, \widetilde{Q}_f,\ f \in \mathbf{F}\right\})$

Las cotas para los errores $\delta_C(\widetilde{F}_0)$ y $\delta_Q(\widetilde{F}_0)$ se obtienen bajo la siguiente hipótesis.

Hipótesis 6.4

(a) La demanda esperada $\overline{w} := Ew_0$ es finita;

(b) La densidad $\rho(\cdot)$ es acotada y Lipschitz continua con constante l, es decir,

$$|\rho(y) - \rho(x)| \leq l\,|y - x| \tag{6.15}$$

Proposición 6.6 *Sea L el interpolador lineal definido en (6.13). Suponga que se satisface la hipótesis 6.2 y que $p > c + h$. Entonces:*

(a)

$$\delta_C(\widetilde{\mathbf{F}}_0) \leq (4p - 2h + c)\Delta. \tag{6.16}$$

(b) Si además suponemos que se satisface la hipótesis 6.4. Entonces,

$$\delta_Q(\widetilde{\mathbf{F}}_0) \leq 2(\theta l + 2M')\Delta, \tag{6.17}$$

donde M' es una cota de la densidad $\rho(\cdot)$.

Para probar este resutado defina

$$U(y) := \int_0^y G(z)dz, \quad y \geq 0,$$

y observe que para cada política de umbral f_S se tiene que la función de costo en una etapa (6.6) queda

$$C_{f_S}(x) = \begin{cases} -cx + p\overline{w} + (c + h - p)S + pU(S) & \text{si } 0 \leq x \leq S \\ p\overline{w} + (h - p)x + pU(x) & \text{si } S < x \leq \theta. \end{cases} \tag{6.18}$$

Además, si el umbral $S \in [m_i, m_{i+1}]$, entonces la interpolación lineal de $C_{f_S}(\cdot)$ toma la forma

$$\widetilde{C}_{f_S}(x) = \begin{cases} -cx + p\overline{w} + (c + h - p)S + pU(S) & \text{si } 0 \leq x \leq m_i \\ b_i(x)C_{f_S}(m_i) + \overline{b_i}(x)C_{f_S}(m_{i+1}) & \text{si } m_i < x \leq m_{i+1} \\ p\overline{w} + (h - p)x + p\widetilde{U}(x) & \text{si } m_{i+1} < x \leq \theta. \end{cases} \tag{6.19}$$

donde $\widetilde{U}(\cdot)$ es la función de interpolación lineal de $U(\cdot)$.

Con el fin de dar expresiones compactas para $Q_{f_S}(\cdot|\cdot)$ y su perturbación con el interpolador lineal $\widetilde{Q}_{f_S}(\cdot|\cdot)$ se introducen las siguientes funciones:

$$pv(y) := \int_0^y v(y)\rho(y-r)dr$$

$$Pv(y) := v(0) + (1 - G(y))pv(y)$$

para cada $y \geq 0$. Entonces

$$Q_{f_S}v(x) = \begin{cases} Pv(S) & \text{si } 0 \leq x \leq S \\ Pv(x) & \text{si } S < x \leq \theta. \end{cases} \tag{6.20}$$

Además, si el umbral $S \in [m_i, m_{i+1}]$, entonces

$$\widetilde{Q}_{f_S}v(x) = \begin{cases} Pv(S) & \text{si } 0 \leq x \leq x_i \\ b_i(x)C_{f_S}(m_i) + \overline{b_i}(x)C_{f_S}(m_{i+1}) & \text{si } m_i < x \leq m_{i+1} \\ \widetilde{P}v(x) & \text{si } m_{i+1} < x \leq \theta. \end{cases} \tag{6.21}$$

Para la demostración de la proposición 6.6 usaremos el siguiente lema.

Lema 6.1 *$|Pv(y) - Pv(x)| \leq (2M' + l\theta)\,|y - x|$ para todo $x, y \in \mathbb{R}_+$, donde M' es una cota de $\rho(.)$ y l es la constante de Lipschitz en la hipótesis 6.4.*

Demostración. Observe que para cada $v : \mathbb{R}^+ \to \mathbb{R}$ acotada y para todo $0 \leq x < y$ se tiene

$$\begin{aligned} pv(y) - pv(x) &= \int_0^y v(r)\rho(y-r)dr - \int_0^x v(r)\rho(x-r)dr \\ &= \int_0^x v(r)\rho(y-r)dr + \int_x^y v(r)\rho(y-r)dr - \int_0^x v(r)\rho(x-r)dr \\ &= \int_0^x v(r)\left[\rho(y-r) - \rho(x-r)\right]dr + \int_x^y v(r)\rho(y-r)dr, \end{aligned}$$

En particular, si $||v||_\infty \leq 1$, resulta que

$$\begin{aligned} |pv(y) - pv(x)| &\leq \int_0^x |\rho(y-r) - \rho(x-r)|\,dr + \int_x^y \rho(y-r)dr \\ &\leq xl\,|y-x| + M'\,|y-x| \\ &\leq (\theta l + M')\,|y-x| \end{aligned} \tag{6.22}$$

Entonces,

$$\begin{aligned} |Pv(y) - Pv(x)| &= |v(0)(1-G(y)) + pv(y) - v(0)(1-G(x)) - pv(x)| \\ &\leq |G(y) - G(x)| + |pv(y) - pv(x)| \\ &\leq \left|\int_x^y \rho(r)dr\right| + (\theta l + M')\,|y-x| \\ &\leq M'\,|y-x| + (\theta l + M')\,|y-x|\,. \end{aligned}$$

lo cual implica que

$$|Pv(y) - Pv(x)| \leq (2M' + \theta l)\,|y-x|\,. \tag{6.23}$$

Demostración de la proposición 6.6. Considere una política f_S y suponga que $S \in [m_i, m_{i+1}]$.

(a) De (6.28) y (6.29) se obtiene con cálculos directos los siguientes resultados:

* $\left|\widetilde{C}_{f_S}(x) - C_{f_S}(x)\right| = 0 \ \forall \ x \in [0, m_i]$.
* $\left|\widetilde{C}_{f_S}(x) - C_{f_S}(x)\right| \leq (2c + 2p - h)\Delta \ \forall \ x \in [m_i, S]$.
* $\left|\widetilde{C}_{f_S}(x) - C_{f_S}(x)\right| \leq (4p - 2hc + c)\Delta \ \forall \ x \in [S, m_{i+1}]$.
* $\left|\widetilde{C}_{f_S}(x) - C_{f_S}(x)\right| \leq 2p\Delta. \ \forall \ x \in [m_{i+1}, \theta]$.

De estas desigulades se sigue que

$$\begin{aligned} ||C_f - \widetilde{C}_{f_S}||_\infty &\leq \max(2c + 2p - h, 4p - 2h + c, 2p)\Delta \\ &\leq (4p - 2h + c)\Delta. \end{aligned}$$

(b) La demostración de esta parte también se realiza por casos:

* Primero obseve que si $x \in [0, m_i]$, entonces

$$\left|\widetilde{Q}_{f_S} v(x) - Q_{f_S} v(x)\right| = |Pv(S) - Pv(S)| \ = 0.$$

* Por otra parte par $x \in [m_i, S]$, se tiene

$$\begin{aligned} \widetilde{Q}_{f_S} v(x) - Q_{f_S} v(x) &= b_i(x) Pv(S) + \overline{b_i}(x) Pv(m_{i+1}) - Pv(S) \\ &= -(1 - b_i(x)) Pv(S) + \overline{b_i}(x) Pv(m_{i+1}) \\ &= \overline{b_i}(x) \left[Pv(m_{i+1}) - Pv(S)\right]. \end{aligned}$$

Entonces,

$$\left|\widetilde{Q}_{f_S} v(x) - Q_{f_S} v(x)\right| \leq |Pv(m_{i+1}) - Pv(S) \ |.$$

Utilizando (6.23) se tiene

$$\begin{aligned} \left|\widetilde{Q}_{f_S} v(x) - Q_{f_S} v(x)\right| &\leq (2M' + \theta l) \left|m_{i+1} - S\right| \\ &\leq (2M' + \theta l)\Delta. \end{aligned}$$

* Si $x \in [S, m_{i+1}]$, se cumple que

$$\begin{aligned} \widetilde{Q}_{f_S} v(x) - Q_{f_S} v(x) &= b_i(x) Pv(S) + \overline{b_i}(x) Pv(m_{i+1}) - Pv(x) \\ &= -b_i(x) \left[Pv(x) - Pv(S)\right] \\ &\quad + \overline{b_i}(x) \left[Pv(m_{i+1}) - Pv(x)\right]. \end{aligned}$$

Entonces,

$$\left|\widetilde{Q}_{f_S}v(x) - Q_{f_S}v(x)\right| \leq |Pv(x) - Pv(S)| + |Pv(m_{i+1}) - Pv(x)|\,.$$

Utilizando (6.23) se tiene

$$\begin{aligned}\left|\widetilde{Q}_{f_S}v(x) - Q_{f_S}v(x)\right| &\leq (2M' + \theta l)\,|x - S| + (2M' + \theta l)\,|m_{i+1} - x| \\ &\leq (2M' + \theta l)\Delta + (2M' + \theta l)\Delta \\ &= 2(2M' + \theta l)\Delta.\end{aligned}$$

* Si $x \in [m_{i+1}, \theta]$, entonces

$$\begin{aligned}\left|\widetilde{Q}_{f_S}v(x) - Q_{f_S}v(x)\right| &= \left|\widetilde{P}v(x) - Pv(x)\right| \\ &= \left|v(0)(1 - \widetilde{G}(x)) + \widetilde{p}v(x) - v(0)(1 - G(x)) - pv(x)\right| \\ &\leq \left|\widetilde{G}(x) - G(x)\right| + |\widetilde{p}v(x) - pv(x)|\,.\end{aligned}$$

Considerando $x \in [m_j, m_{j+1}]$ con $j \geq i + 1$, se tiene que

$$\begin{aligned}\left|\widetilde{G}(x) - G(x)\right| &= \left|b_j(x)G(m_j) + \overline{b_j}(x)G(m_{j+1}) - G(x)\right| \\ &= \left|-b_j(x)\,[G(x) - G(m_j)] + \overline{b_j}(x)\,[G(m_{j+1}) - G(x)]\right| \\ &\leq \int_{m_j}^{x} \rho(r)dr + \int_{x}^{m_{j+1}} \rho(r)dr \\ &\leq M'(x - m_j) + M'(m_{j+1} - x) \\ &\leq 2M'\Delta.\end{aligned}$$

Además,

$$\begin{aligned}|\widetilde{p}v(x) - pv(x)| &= \left|b_j(x)pv(m_j) + \overline{b_j}(x)pv(m_{j+1}) - pv(x)\right| \\ &= \left|b_j(x)\,[pv(m_j) - pv(x)] + \overline{b_j}(x)\,[pv(m_{j+1}) - pv(x)]\right| \\ &\leq |pv(m_j) - pv(x)| + |pv(m_{j+1}) - pv(x)| \\ &\leq (\theta l + M')\,|m_j - x| + (\theta l + M')\,|m_{j+1} - x| \ \text{ por (6.22)} \\ &\leq 2(\theta l + M')\Delta.\end{aligned}$$

Entonces,

$$\begin{aligned}\left|\widetilde{Q}_{f_S}v(x) - Q_{f_S}v(x)\right| &\leq 2M'\Delta x + 2(\theta l + M')\Delta \\ &\leq 2(\theta l + 2M')\Delta.\end{aligned}$$

De los casos anteriores se concluye con la desigualda (6.17). ■

Las cotas de las constantes de aproximación a las que llegamos (6.16) y (6.17), pueden hacerse arbitrariamente pequeñas, basta con tomar mallas suficientementes finas.

Para casos especificos estas cotas pueden ser más ajustadas, por ejemplo si la demanda es exponencial, como se verá posteriormente.

6.6 Algoritmos IVA para el modelo $\widetilde{\mathfrak{M}}$

En la sección anterior se obtuvieron las constantes de aproximación para el modelo $\widetilde{\mathfrak{M}}$, las cuales dependen de la malla que genera a los promediadores, que para este sistema de inventarios en particular es un interpolador lineal. A continuación se presentan los pseudocodigos del algoritmo IVA tanto para costo descontado (4.9) como para costo promedio (5.18).

6.6.1 Costo descontado

Algoritmo 6.1 *Obtiene una política $\widetilde{V}_k$-greedy y necesita como entrada una función $\widetilde{V}_0 \in \mathbf{M}_a(\mathbf{X})$, una malla de puntos $\{m_i\}_{i=1}^t$ en $\mathbf{X}$ de tamaño finito t y el valor de la tolerancia de paro Tol.*

Paso 0.

$\widetilde{V}_0 \leftarrow$ *función dada*;

$Tol \leftarrow$ *valor dado*;

$k \leftarrow 0$;

Paso 1

Re*petir*

$k \leftarrow k+1$;

Calcular $T\widetilde{V}_{k-1}$ *para cada punto de la malla* $\{m_i\}_{i=1}^t$;

Obtener la aproximación $LT\widetilde{V}_{k-1}$ *usando resultados anteriores*;

$\widetilde{V}_k \leftarrow LT\widetilde{V}_{k-1}$;

$ToleParo \leftarrow \left\| \widetilde{V}_k - \widetilde{V}_{k-1} \right\|_\infty$;

Hasta que $ToleParo$ *sea menor que* Tol.

Paso 2

Obtener la política $\widetilde{V}_k-greedy$;

Terminar.

6.6.2 Costo promedio

Algoritmo 6.2 *Obiene una política $\widetilde{J}_n-greedy$. Necesita como entrada una función $\widetilde{J}_0 \equiv 0$, una malla de puntos $\{m_i\}_{i=1}^t$ en $\mathbf{X}$ de tamaño finito t y el valor de la tolerancia de paro Tol.*

Paso 0

Iniciar con $\widetilde{J}_0 \equiv 0$;

$Tol \leftarrow$ *valor dado*;

$n \leftarrow 0$;

Paso 1

Re*petir*

$n \leftarrow n+1$;

Calcular $T\widetilde{J}_{n-1}$ *para cada punto de la malla* $\{m_i\}_{i=1}^{t}$;

Obtener la aproximación $LT\widetilde{J}_{n-1}$ *usando resultados anteriores*;

$\widetilde{J}_n \leftarrow LT\widetilde{J}_{n-1}$;

$\widetilde{\rho}_n \leftarrow \widetilde{J}_n - \widetilde{J}_{n-1}$;

$\widetilde{s}_n \leftarrow \inf_{x\in\mathbf{X}} \widetilde{\rho}_n(x)$;

$\widetilde{S}_n \leftarrow \sup_{x\in\mathbf{X}} \widetilde{\rho}_n(x)$;

$\|\widetilde{\rho}_n\|_{sp} \leftarrow \widetilde{S}_n - \widetilde{s}_n$;

Hasta que $\|\widetilde{\rho}_n\|_{sp}$ *sea menor que* Tol.

Paso 2

Obtener la política $\widetilde{J}_n$*−greedy*;

Terminar.

Observación 6.1 *Como ya se hizo saber en las observaciones 5.4 y 6.1, las instrucciones de ambos algoritmo* IVA *son exactamente las mismas instrucciones que las de los respectivos algoritmos* IV, *2.1 y 2.2, solo se le han agregado la instrucción "**calcular los valores de la función iteración de valores en cada punto de la malla**" y la instrucción "**obtener la aproximación usando los resultados anteriores**".*

6.7 Sistema de inventarios demanda exponencial

Con el fin de implementar computacionalmente el sistema de inventarios con demanda exponencial se particularizan los resultados obtenidos anteriormente. A continuación se da un resumen de dichos resultados.

En (6.6) se da la función de costo en una etapa C, que al considerar demanda exponencial se reduce a

$$C(x,a) = ca + h(x+a) + (p/\lambda)e^{-\lambda(x+a)}. \tag{6.24}$$

Al considerar una política de umbral f_S (6.10), se tiene

$$C_{f_S}(x) = \begin{cases} -cx + (c+h)S + (p/\lambda)e^{-\lambda S} & \text{si } 0 \le x \le S \\ hx + (p/\lambda)e^{-\lambda x} & \text{si } S < x \le \theta, \end{cases}$$

de donde se deduce que C tiene por cota

$$M = (c+h)\theta + (p/\lambda)e^{-\lambda\theta}. \tag{6.25}$$

La probabilidad de transición Q en (6.8) es

$$Q(B \mid x, a) = I_B(0)e^{-\lambda(x+a)} + \int_0^{x+a} I_B(x+a-w)\lambda e^{-\lambda w} dw.$$

Las constantes de aproximación $\delta_Q(\widehat{\mathbf{F}}_0)$, $\delta_C(\widetilde{\mathbf{F}}_0)$ y $\delta_Q(\widetilde{\mathbf{F}}_0)$ obtenidas de (6.14), (6.16) y (6.17) respectivamente son

$$\delta_Q(\widehat{\mathbf{F}}_0) = 2G(\theta) = 2(1 - e^{-\lambda\theta}) \tag{6.26}$$

$$\delta_C(\widetilde{\mathbf{F}}_0) \leq (p + h + c)\Delta x. \tag{6.27}$$

$$\delta_Q(\widetilde{\mathbf{F}}_0) \leq 2(\theta\lambda^2 + 2\lambda)\Delta x, \tag{6.28}$$

ya que la densidad exponencial tiene cota $M' = \lambda$ y es Lipschitz continua con modulo $l = \lambda^2$.

Además, el umbral óptimo para costo descontado dado en (6.11) es

$$S^* = \frac{1}{\lambda} \ln \frac{p - \alpha c}{h + (1 - \alpha)c}. \tag{6.29}$$

y para costo promedio dado en (6.12) es

$$S^* = \frac{1}{\lambda} \ln \frac{p - c}{h}. \tag{6.30}$$

La cota β de la condición de ergodicidad dada en (6.9) con demanda exponencial se reduce a

$$\beta = 1 - e^{-\lambda\theta}. \tag{6.31}$$

6.8 Resultados numéricos costo descontado

A continuación se obtienen los resultados numéricos de la implementación computacional para el PCO en costo descontado para el sistema de inventarios con demanda exponencial. Se obtienen las cotas de error de interpolación lineal del algoritmo IVA para ambos modelos. Además se obtienen gráficas de algunas funciones IVA y las de sus correspondientes políticas, así como también la gráfica de la funcion $\widetilde{V}_k - greedy$ y la simulación de la cadena usando la pólitica *greedy* correspondiente.

El sistema se implementara con los siguientes valores paramétricos:

1. Demanda exponencial, λ =0.1
2. Costo unitario de produccion, c =1.5
3. Costo unitario por mantener inventario, h =0.5
4. Costo unitario por demanda no satifecha, p =3
5. Espacio de estados $[0, \theta]$ con θ =20.
6. La tolerancia de paro, Tol =0.01
7. Factor de descuento, α =0.6.

6.8.1 Salidas de la implementación computacional para el modelo $\widehat{\mathfrak{M}}$

Como se indicó en la sección 6.5.1 la constante de aproximación $\delta_Q(\widehat{\mathbf{F}}_0)$ no depende del tamaño de la malla, lo cual es una gran desventaja en términos de aproximación y su valor constante en este caso es

$$\delta_Q(\widehat{\mathbf{F}}_0) = 2(1 - e^{-\lambda\theta}) = 1.729329$$

por lo que, también es constante la cota del error de aproximación ($CotaErrorAprox$), que es el segundo término del lado derecho de (4.10), con un valor de

$$CotaErrorAprox = 1142.915.$$

Para obtener la cota de error total ($CotaErrorTotal$), al anterior resultado se le suma la cota del error de paro ($CotaErrorParo$), primer término del lado derecho de (4.10), cuyo valor si depende del tamaño de malla, como se puede ver en las tablas del siguiente modelo. Note que la cota de error total tiende en forma decreciente al valor constante $CotaErrorAprox$, es decir,

$$CotaErrorTotal = CotaErrorParo + 1142.915$$

Por lo que, estas cotas de aproximación del algoritmos IVA para el modelo $\widehat{\mathfrak{M}}$, son pésimas.

6.8.2 Salidas de la implementación computacional para el modelo $\widetilde{\mathfrak{M}}$

A diferencia del modelo anterior, en este caso $\delta_C(\widetilde{\mathbf{F}}_0)$ y $\delta_Q(\widetilde{\mathbf{F}}_0)$ si dependen del tamaño de la malla, como se puede observar en sus respectivas formulaciones (6.27) y (6.28) Para diferentes tamaños de malla, se obtuvieron los siguientes resultados de estas constantes de aproximación

	Malla 101	Malla 1001	Malla 5001	Malla 10001
$\delta_C(\widetilde{\mathbf{F}}_0)$	1	0.1	0.02	0.01
$\delta_Q(\widetilde{\mathbf{F}}_0)$	0.16	0.016	0.0032	0.0016

La función C dada por (6.24) tiene una cota superior M =44.06006 de acuerdo con (6.25), la política óptima tiene un umbral analítico S =6.466272 deducido de (6.29). Al utilizar (4.11) con distintos tamaños de malla se obtienen los resultados siguientes, donde hemos denotado $ToleParo := \left\|\widetilde{V}_k - \widetilde{V}_{k-1}\right\|_\infty$

	Malla 101	Malla 1001	Malla 5001	Malla 10001
$No.\ iteraciones$	17	17	17	17
$ToleParo$	0.007454653	0.007454679	0.007487855	0.007487864
$UmbralGreedy$	6.6	6.51	6.468	6.467
$CotaErrorAprox$	57.87207	5.787207	1.157441	0.5787207
$CotaErrorParo$	0.02236396	0.02236404	0.02246356	0.02246359
$CotaErrorTotal$	57.89443	5.809571	1.179905	0.60118439

$UmbralGreedy$ se deduce de la gráfica de la política $\widetilde{V}_k - greedy$.

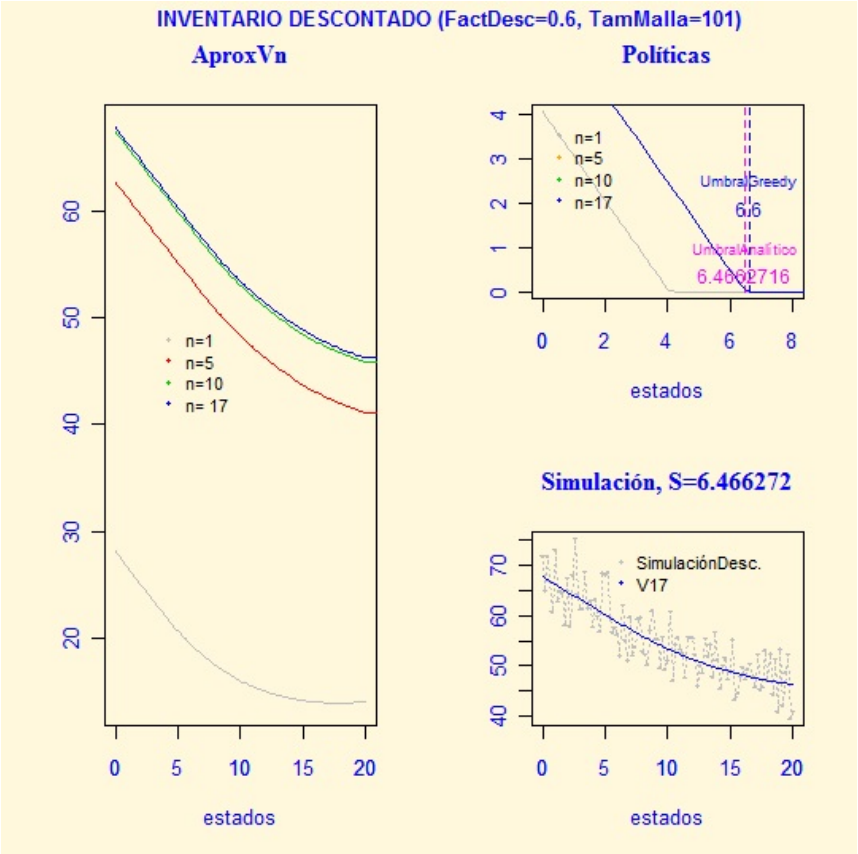
Los tiempos en segundos de las corridas tanto para IVA como la simulación para los distintos tamaños de malla son:

	Malla 101	Malla 1001	Malla 5001	Malla 10001
IVA	29.42	483.14	6321.33	75252.24
Simulación	52.28	567.11	3348.67	3468.53

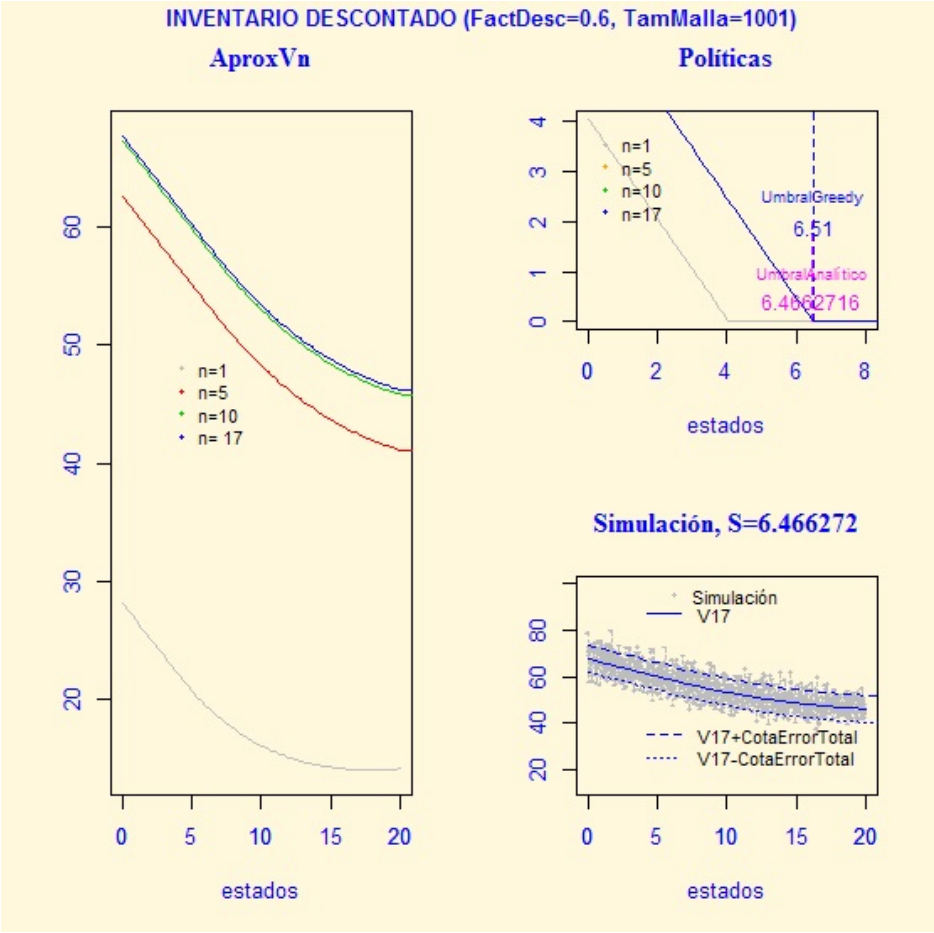
Gráficas para el modelo $\widetilde{\mathfrak{M}}$

En la primer gráfica se ilustran una sucesión de funciones IVA y en la segunda la de sus correspondientes políticas, hasta obtener la $\widetilde{V}_k - greedy$ que satisface la tolerancia Tol. En la tercer gráfica se ilustra la funcion $\widetilde{V}_k - greedy$ y la simulación de la cadena usando la pólitica $greedy$ asociada. Para obtener la gráfica de la simulación, en cada punto de la malla se obtiene el promedio de los valores de la la función valor esperado de 100 etapas, obtenidos de 50 simulaciones del sistema de inventarios considerando el umbral analítico S^* =6.466272. Las gráficas resultantes para distintos tamaños de malla son los siguientes:

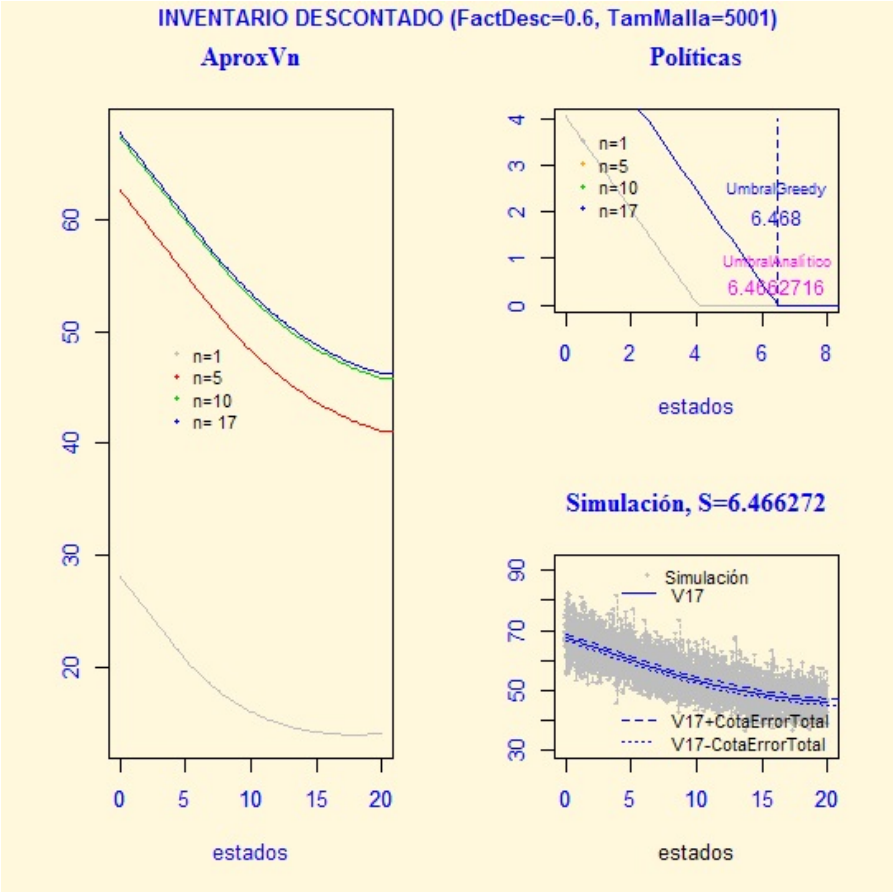
(a) Gráficas con malla de 101 puntos.



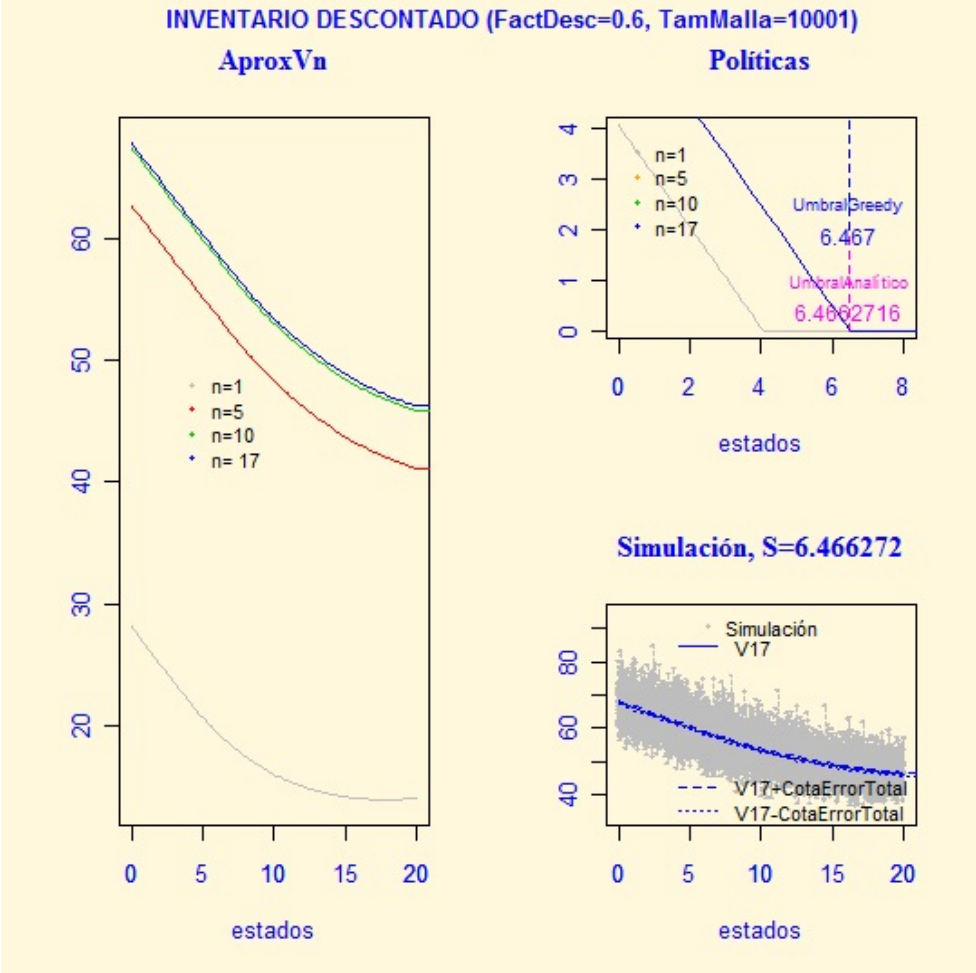
(b) Gráficas con malla de 1001 puntos. En la gráfica de la simulación, se añadió la banda $\widetilde{V}_{17} \pm CotaErrorTotal$ donde al utilizar (4.11) para este tamaño es $CotaErrorTotal$ =5.809571.



(c) Gráficas con malla de 5001 puntos. En la gráfica de la simulación, se añadió la banda $\widetilde{V}_{17} \pm CotaErrorTotal$ donde $CotaErrorTotal$ =1.179905.



(d) Gráficas con malla de 10001 puntos. En la gráfica de la simulación, se añadió la banda $\widetilde{V}_{17} \pm CotaErrorTotal$ donde $CotaErrorTotal$ =0.60118439.



Observación 6.2 *La $CotaErrorAprox$ es muy sensible a cambios del tamaño de malla, mientras que $CotaErrorParo$ esta determinada por Tol, la cual en este caso se mantiene fija en 0.01, por lo que $CotaErrorParo$ tampoco cambiara, entonces $CotaErrorTotal$ tiende a $CotaErrorParo$. Para obtener valores $CotaErrorTotal$ más pequeños se debe disminuir el valor de Tol.*

Observación 6.3 *Manteniendo fijo Tol, el No. iteraciones no cambia y el Umbral-Greedy mejoran poco al aumentar el tamaño de la malla.*

6.9 Resultados numéricos costo promedio

A continuación se desarrolla la implementación computacional del sistema de inventarios, pero ahora con índice en costo promedio. Se calculan las cotas de error de interpolación lineal del algoritmo IVA para ambos modelos.. Además se obtienen gráficas que ilustran la politica $f \in \mathbf{F}$, $\widetilde{h}_n - greedy$ y la función de costo promedio aproximado asociada a dicha política f. Se simula la cadena usando la pólitica *greedy* correspondiente.

Los valores paramétricos considerados para costo promedio son:

1. Demanda exponencial, λ =0.05
2. Costo unitario de produccion, c =1.5
3. Costo unitario por mantener inventario, h =0.5
4. Costo unitario por demanda no satifecha, p =3
5. Espacio de estados $[0, \theta]$ con θ =25.
6. La tolerancia de paro, Tol =0.0001

6.9.1 Salidas de la implementación computacional para el modelo $\widehat{\mathfrak{M}}$

Como se indicó en la sección 6.5.1, $\delta_Q(\widehat{\mathbf{F}}_0)$ no depende del tamaño de la malla, y su valor constante en este caso es

$$\delta_Q(\widehat{\mathbf{F}}_0) = 2(1 - e^{-\lambda\theta}) = 1.42699$$

por lo cual, también es constante ($CotaErrorAprox$), con un valor de

$$CotaErrorAprox = 669.3074$$

Por lo que, la cota de error total tiende en forma decreciente a $CotaErrorAprox$. Es decir,

$$CotaErrorTotal = CotaErrorParo + 669.3074$$

Por lo que, estas cotas de aproximación del algoritmos IVA para el modelo $\widehat{\mathfrak{M}}$, en costo promedio también son pésimas.

6.9.2 Salidas de la implementación computacional para el modelo $\widetilde{\mathfrak{M}}$

En este caso $\delta_C(\widetilde{\mathbf{F}}_0)$ y $\delta_Q(\widetilde{\mathbf{F}}_0)$ si dependen del tamaño de la malla. Para diferentes tamaños de malla de (6.27) y (6.28), se obtuvieron los siguientes resultados

	Malla 101	Malla 1001	Malla 5001	Malla 10001
$\delta_C(\widetilde{\mathbf{F}}_0)$	1.25	0.125	0.025	0.0125
$\delta_Q(\widetilde{\mathbf{F}}_0)$	0.08125	0.008125	0.001625	0.0008125

La función C de acuerdo con (6.25) tiene una cota superior M =67.19029. Se ha obtenido de (6.31) que β =0.7134952 y de (6.30) el $UmbralAnalítico$ =21.97225. Para distintos tamaños de malla se obtienen los resultados siguientes, donde hemos denotado $CotaErrorParo := \|\widetilde{\rho}_n\|_{sp}$.

	Malla 101	Malla 1001	Malla 5001	Malla 10001
No. iteraciones	7	7	7	7
$UmbralGreedy$	22	21.975	21.975	21.9725
$CotaErrorAprox$	40.60904	4.060904	0.8121807	0.4060904
$CotaErrorParo$	2.604605e-05	2.572659e-05	2.572063e-05	2.572063e-05
$CotaErrorTotal$	40.60906	4.060929	0.8122065	0.4061161

$UmbralGreedy$ se deduce de la gráfica de la política $\widetilde{J}_k - greedy$.

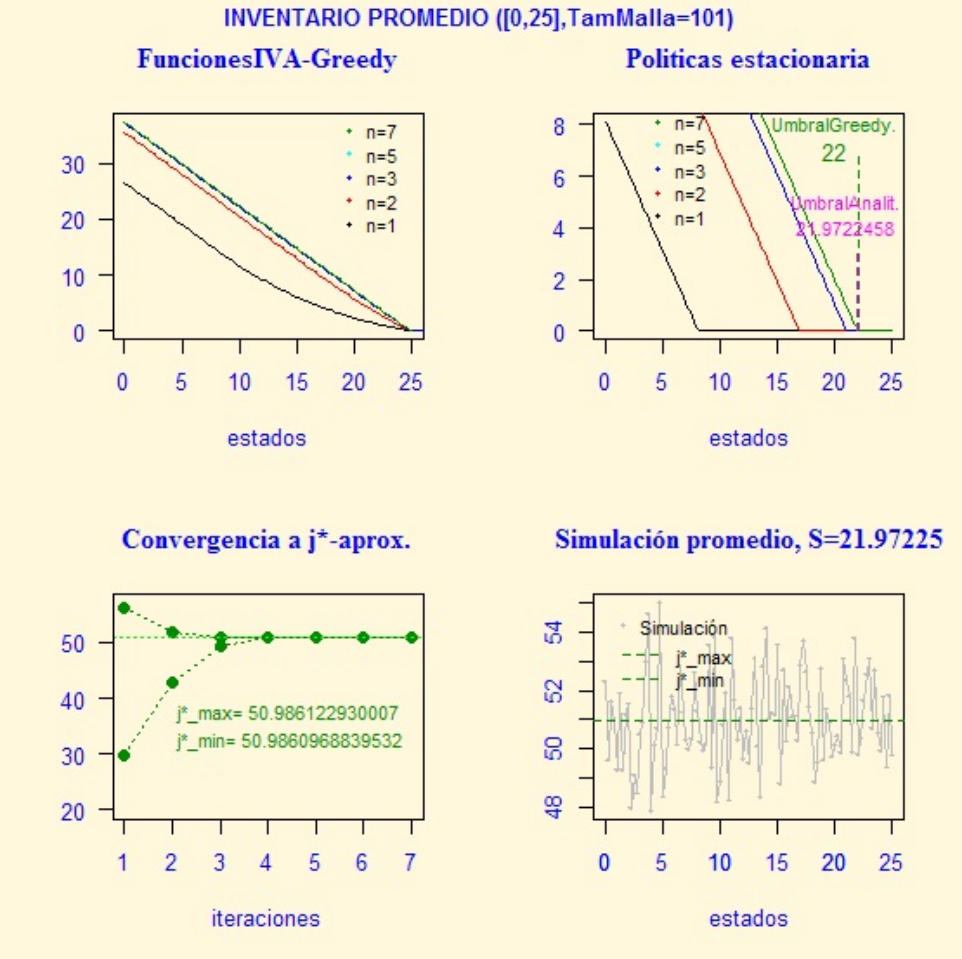
Los tiempos en segundos de las corridas tanto para IVA como para la simulación con distintos tamaños de malla son

	Malla 101	Malla 1001	Malla 5001	Malla 10001
IVA	9.27	129.94	1997.58	58619.43
Simulación	12.70	149.03	1186.81	745.31???

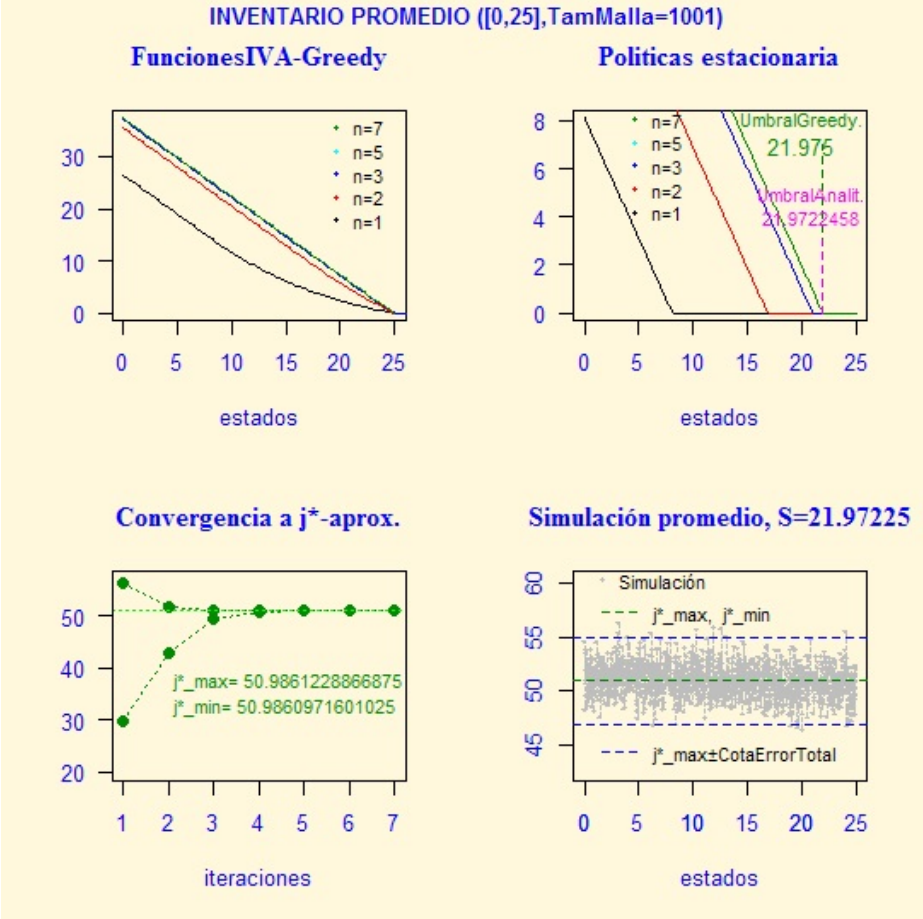
Gráficas para el modelo $\widetilde{\mathfrak{M}}$:

En las gráficas, se ilustran la sucesion de funciones IVA, las de sus correspondientes políticas y las sucesiones de $\widetilde{s}_n$ y $\widetilde{S}_n$ (por tipografia en la gráfica $\widetilde{s}_n = j^* - \min$ y $\widetilde{S}_n = j^* - \max$) hasta obtener la $\|\widetilde{\rho}_k\|_{sp}$ sea menor que Tol. En la otra gráfica se ilustra el resultado de 10 simulaciones del sistema, para cada punto de la malla y se ha obtenido el promedio de los valores de la función valor esperado en 100 etapas, considerando el umbral analítico $UmbralAnalítico$ =21.97225. Las gráficas resultantes para distintos tamaños de malla son los siguientes:

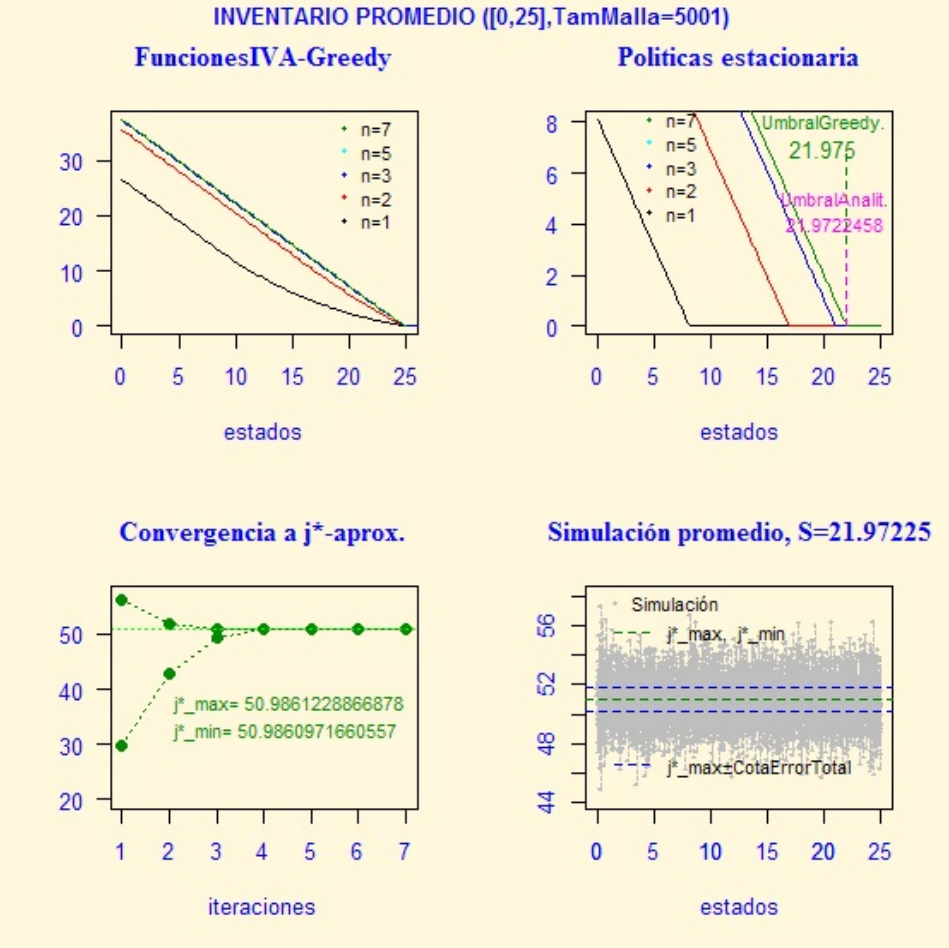
(a) Gráficas con malla de 101 puntos.



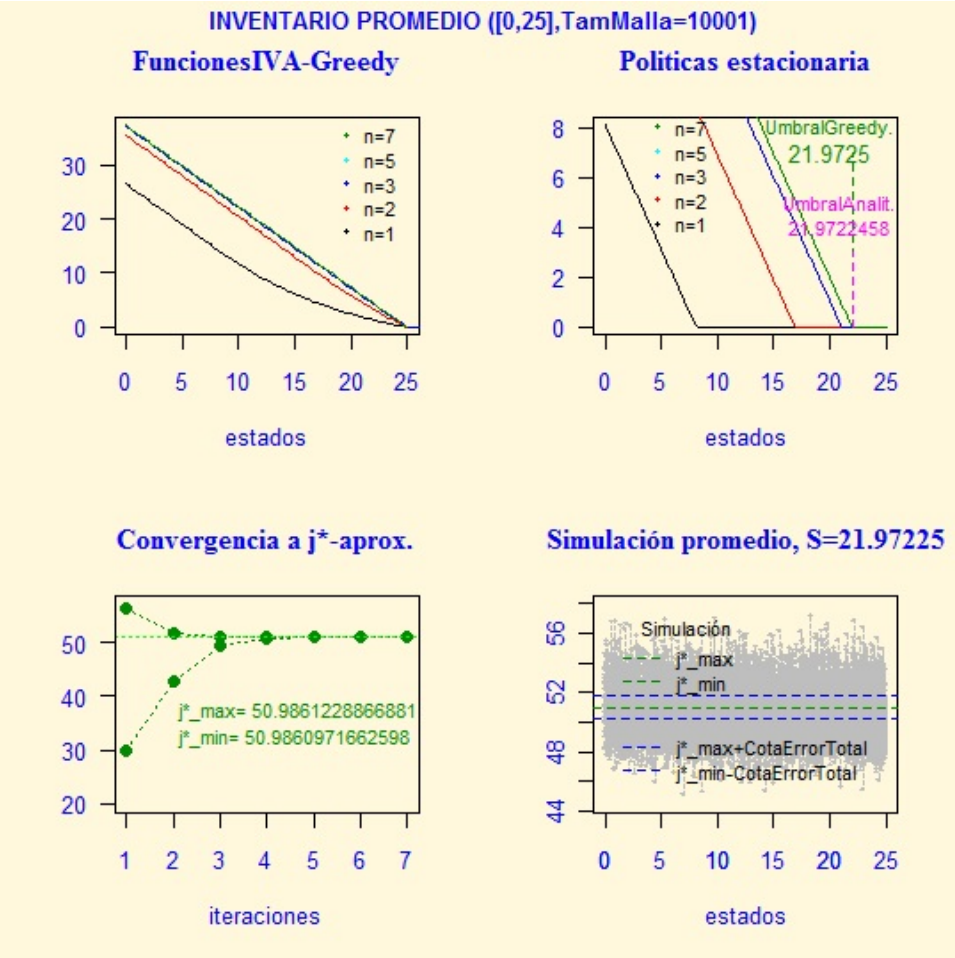
(b) Gráficas con malla de 1001 puntos. En la gráfica de la simulación se ha considerado la banda $\pm CotaErrorTotal = \pm 4.060929$



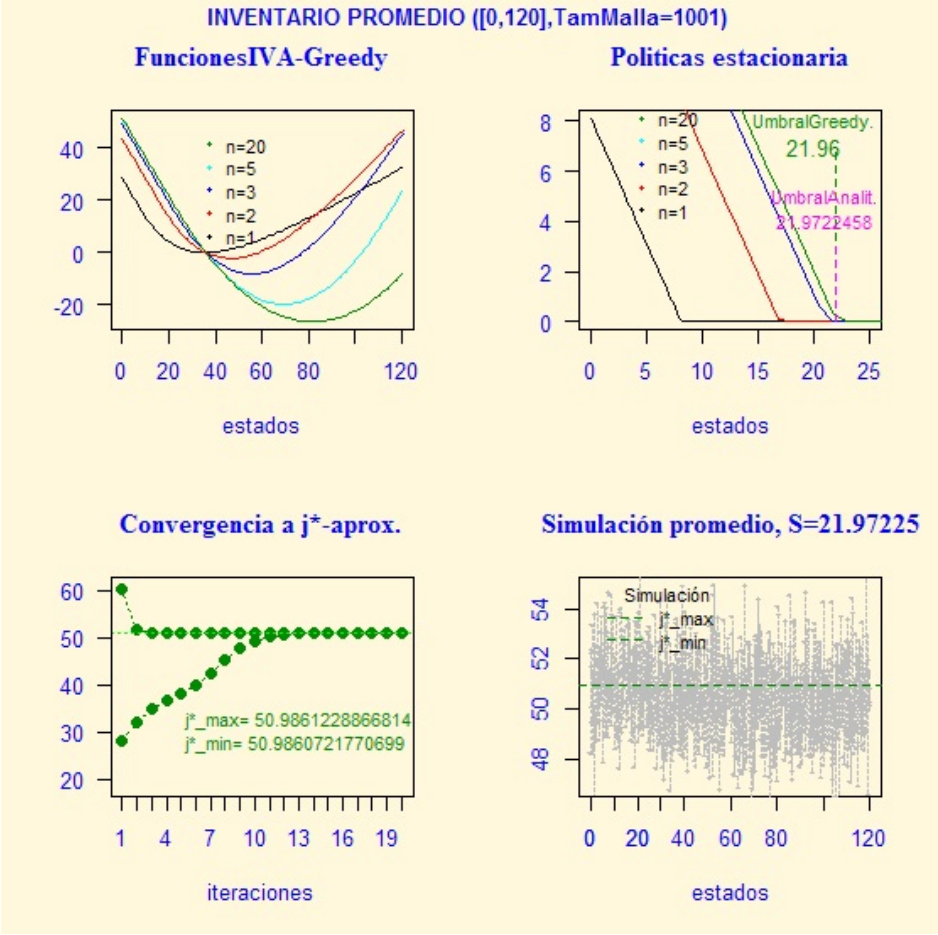
(c) Gráficas con malla de 5001 puntos. En la gráfica de la simulación se ha considerado la banda $\pm CotaErrorTotal = \pm 0.8122065$



(d) Gráficas con malla de 10001 puntos. En la gráfica de la simulación se ha considerado la banda $\pm CotaErrorTotal = \pm\ 0.4061161$



(e) En las siguientes gráficas se ha considerado un espacio de estados $[0, 120]$, con el fin de visualizar la evolución de las funciones $\widetilde{J}_k$ y sus respectivas políticas, por tal motivo, existe diferencia respecto a las tablas (en tablas $[0, 25]$) en el número de iteraciones, *UmbralGreedy* y sobre todo que las cotas son exesivamente grandes con espacio de estados $[0, 120]$.



Observación 6.4 *Al igual que en costo descontado la CotaErrorAprox es muy sensible a cambios del tamaño de malla, mientras que CotaErrorParo esta determinada fuertemente por Tol la cual en este caso se mantiene fija en 0.0001 por lo que CotaErrorParo tampoco cambiara.*

Capítulo 7

Conclusiones y perspectivas

En este trabajo se han desarrollado los fundamentos teóricos para obtener las cotas de desempeño del algorítmo IVA en procesos de control markoviano con espacios generales y costos acotados, tanto para el índice en costo descontado como para el índice en costo promedio.

En el desarrollo de la teoría, el resultado clave fue que los promediadores utilizados como operadores de aproximación definen una probabilidad de transición en el espacio de estados del sistema controlado. Esto permitió, por una parte, ver la función aproximante como el "*valor promedio*" de la función que se desea aproximar con respecto a la transición de probabilidad inducida por el operador, y por otra parte, ver el paso de aproximación en el algoritmo IVA como una perturbación del modelo original. Se resolvieron los problemas D1-D3 y P1-P3 planteados en la introducción tanto para índice en costo descontado como para índice en costo promedio.

La importancia de las cotas obtenidas radica en que son computables y se puede obtener la cota del error del algoritmo IVA con probabilidad 1, a diferencia de los algoritmos heuristicos cuyo error no es determinístico. Además, la implementación computacional del algoritmo IVA, permite acortar la brecha entre teoría y práctica lo anterior fué ejemplificado con resultados numéricos obtenidos de la implementación computacional de un sistema de inventarios, obteniendose excetentes resultados al contrastar con la solución analítica y con simulación.

Algunas posibles líneas de extensión de este trabajo son las siguientes:

(a) Procesos de control markoviano con costos no acotados.

(b) Procesos de control markoviano con espacios de estado y controles no compactos.

(c) Análizar e implementar los promediadores con otros métodos como; iteración de valores y programación lineal.

(d) Analizar la posibilidad de utilizar promediadores en procesos de control semimarkovianos.

(d) Analizar la posibilidad de utilizar promediadores en juegos de suma cero.

Bibliografía

[1] A. Almudevar (2008), *Approximate fixed point iteration with an application to infinite horizont Markov decision processes,* SIAM Journal on Control and Optimization 46, pp 5411-561.

[2] A. Araposthatis, B. S. Borkar, E. Fernández-Guacherand, M. K: Gosh, S. I. Marcus (1993), *Discrete-time contolled Markov processes with average cost criterion: a survey,* SIAM Journal on Control and Optimization 31, pp. 282-344.

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

[4] D. P. Bertsekas (1987), *Dynamic Programming: Deterministic and Stochastic Models*, Prentice-Hall, Englewood Cliffs, N. J.

[5] D. P Bertsekas and J. N. Tsitsiklis (1996), *Neuro-Dynamic Programming,* Athena Scientific, Belmont, MA.

[6] V. S. Borkar (1991), *Topics in Controlled Markov chains,* Pitman Research Notes in Math. No. 240, Longman, Harlow, U. K.O.

[7] C. C. Y. Dorea and A. G. C. Pereira (2006), *A note on a variations of Doeblin's condition for uniform ergodicity of Markov chains,* Acta Mathematica Hungarica 110, pp. 287-202.

[8] D. P. de Farias and B. Van Roy (2000), *On the existence of fixed points for approximate value iteration and temporal defference learning,* Journal of Optimization Theory and Applications 105, pp 589-608.

[9] G. J. Gordon (1995), *Stable function approximation in dynamic programming,* Proceedings of the 12th International Conference on Machine Learning, pp. 261-268.

[10] O. Hernández-Lerma (1989), *Adaptative Markov Control Processes,* Springer-Verlag, N.Y.

[11] O. Hernández-Lerma and J. B. Lasserre (1996), *Discrete-Time Markov Control Processes, Basic Optimaliy Criteria,* Springer-Verlag, N.Y.

[12] O. Hernández-Lerma and J. B. Lasserre (1999), *Furhter Topics on Discrete-Time Markov Control Processes,* Springer-Verlag, N.Y.

[13] O. Hernández-Lerma and J. B. Lasserre (2003), *Markov Chains and Invariant Probabilites,* Birkhauser Verlag, Basel, Switzerland.

[14] O. Hernández-Lerma, R. Montes-de-Oca, R. Cavazos-Cadena (1991), *Recurrence conditions for Markov decision processes with Borel Spaces: a survey,* Annals of Operations Research 29, pp. 29-46.

[15] J. M. Lee and J. H. Lee (2004), *Approximate dynamic programming strategies and their applicability for process control: a review and future directions,* International Journal of Control, Automation and Sistems 2(3), 263-278.

[16] S.P. Meyn and R. L. Tweedlie (1993), *Markov Chain and Stochastic Stability,* Springer-Verlag, London.

[17] R. Munos (2007), *Performance bounds in L_p-norm for approximate value iteration,* SIAM Journal on Control and Optimization 47, pp. 2303-2347.

[18] A.S. Nowak (1998), *A generalization of Ueno´s inequality for n-step transition probabilities.* Applicationes Mathematicae 25, pp. 295-299.

[19] M. L. Puterman (1994), *Markov Decision Processes: Discrete Stochastic Dynamic Programming,* Wiley, N. Y.

[20] W. P. Powell (2007), *Approximate Dynamic Programmin. Solving the Curse of Dimensinality,* John Wiley & Sons, Inc.

[21] J. Rust (1996), *Numerical dynamic programming in economics,* in Handbook of computational Economics, Vol.1, H Amman, D. Kendrick and J. Rust (eds), pp 619-728.

[22] M. S. Santos (1998), *Analysis of numerical dynamic programming algorithm applied to economic models,* Econometrica 66 pp. 409-426.

[23] L. I. Sennot (1999), *Stochastic Dynamic Programming and the Control of Queueing Systems,* John Wiley & Sons, Inc.

[24] J. Stachurski (2008), *Continuous state dynamic programming via nonexpansive approximation,* Computacional Economics 31, pp.141-160.

[25] J. Stachurski (2009), *Dynamic Economic: Theory and Computation*, The MIT Press, Cambridge, MA.

[26] R. S. Sutton and A. G. Barto (1998), *Reinforcement Learning: An Itroduction,* MIT Press, Cambridge, MA.

[27] O. Vega and J. López, *Performance Bounds for Discounted Approximate Value Iteration using Averagers*, Reporte de Investigación 13, Departamento de Matemáticas, Universidad de Sonora, México. Sometido para su publicación.

[28] O. Vega and J. López (2012), *A Perturbation Approach to Approximate Dynamic Programming for the Average Cost Criterion on Borel Spaces with Bounded Cost*, Reporte de Investigación 14, Departamento de Matemáticas, Universidad de Sonora, México. Sometido para su publicación.

[29] O. Vega-Amaya and R. Montes-de-Oca (1998). *Application of average dynamic programming to inventory systems,* Mathematical Methods of Operations Research 47, pp 451-471.

[30] D. J. White (1993) *A Survey of Applications of Markov Decision Processes,* J. Opl. Res. Soc. Vol. 44, No. 11, pp. 1073-1096.