Universidad Autónoma
Metropolitana-Iztapalapa

Casa abierta al tiempo

División de Ciencias Básicas e Ingeniería

Procesos de Decisión de Markov Descontados Perturbados

TESIS

QUE PARA OBTENER EL TíTULO DE:

Doctor en Ciencias Matemáticas

PRESENTA:

M. C. Rei Israel Ortega Gutiérrez

DIRECTOR DE TESIS:

Dr. José Raúl Montes de Oca Machorro

Jurado Calificador:

Presidente: Dr. José Raúl Montes de Oca Machorro

Secretario: Dr. Rolando Cavazos Cadena

Vocales: Dr. Julio César García Corte

Dr. Evgueni Ilich Gordienko

Dr. Víctor Hugo Vázquez Guevara

México, D.F. Septiembre 2017

Casa abierta al tiempo
UNIVERSIDAD AUTÓNOMA METROPOLITANA

ACTA DE DISERTACIÓN PÚBLICA

No. 00054
Matrícula: 2112800158

PROCESOS DE DECISIÓN DE MARKOV DESCONTADOS PERTURBADOS

En la Ciudad de México, se presentaron a las 12:00 horas del día 13 del mes de octubre del año 2017 en la Unidad Iztapalapa de la Universidad Autónoma Metropolitana, los suscritos miembros del jurado:

DR. JOSE RAUL MONTES DE OCA MACHORRO
DR. JULIO CESAR GARCIA CORTE
DR. EVGUENI ILICH GORDIENKO
DR. VICTOR HUGO VAZQUEZ GUEVARA
DR. ROLANDO CAVAZOS CADENA

Bajo la Presidencia del primero y con carácter de Secretario el último, se reunieron a la presentación de la Disertación Pública cuya denominación aparece al margen, para la obtención del grado de:

DOCTOR EN CIENCIAS (MATEMATICAS)

DE: REI ISRAEL ORTEGA GUTIERREZ

y de acuerdo con el artículo 78 fracción IV del Reglamento de Estudios Superiores de la Universidad Autónoma Metropolitana, los miembros del jurado resolvieron:

Aprobar

Acto continuo, el presidente del jurado comunicó al interesado el resultado de la evaluación y, en caso aprobatorio, le fue tomada la protesta.

REI ISRAEL ORTEGA GUTIERREZ
ALUMNO

REVISÓ

LIC. JULIO CESAR DE LARA ISASSI
DIRECTOR DE SISTEMAS ESCOLARES

DIRECTOR DE LA DIVISIÓN DE CBI	PRESIDENTE	VOCAL
DR. JOSE GILBERTO CORDOBA HERRERA	DR. JOSE RAUL MONTES DE OCA MACHORRO	DR. JULIO CESAR GARCIA CORTE

VOCAL	VOCAL	SECRETARIO
DR. EVGUENI ILICH GORDIENKO	DR. VICTOR HUGO VAZQUEZ GUEVARA	DR. ROLANDO CAVAZOS CADENA

Dedicado a
mi esposa e hijos

Agradecimientos

Doy gracias a Dios por darme la vida y permitirme llegar hasta este momento.

Agradezco a las personas que siempre han estado conmigo y me han apoyado en todo momento, mis padres Remedios Gutiérrez y Gerardo Ortega quienes han hecho un gran esfuerzo humano y económico para que yo salga adelante. También reconozco el apoyo y cariño de mis hermanos Paulina, Gerardo, Dalila, Areli y Rubén que siempre ha sido y seguirá siendo muy importante para mí.

Durante la realización de esta tesis recibí la valiosa ayuda de mi asesor, es un placer ahora expresar mi agradecimiento sincero al Dr. Raúl Montes de Oca por su tiempo, dedicación, amistad y sobre todo por todas las ideas y sugerencias que han constituido una aportación de valor incalculable.

Por su apoyo, motivación, comprensión, y sobre todo por compartir buenos y malos momentos, un agradecimiento a mi esposa Rosario Ávila y a mi mayor motivación, mis dos hijos, Israel y Manuel.

Al jurado calificador integrado por: Dr. Raúl Montes de Oca Machorro, Dr. Rolando Cavazos Cadena, Dr. Julio César García Corte, Dr. Evgueni Ilich Gordienko y Dr. Víctor Hugo Vázquez Guevara. Gracias a todos ustedes por sus comentarios, sugerencias y observaciones, los cuales han mejorado y enriquecido este trabajo.

También se agradece a CONACyT, por el apoyo económico otorgado para la realización de este doctorado.

INTRODUCCIÓN

En este trabajo se tratará con procesos de decisión de Markov (PDMs), a tiempo discreto, con criterio de rendimiento el costo total descontado (véase [7], [9], [25], [45]).Un Proceso de Decisión de Markov (PDM) tiene asociado un modelo conocido como Modelo de Control de Markov (MCM), cuyas componentes permiten caracterizar su desarrollo en el transcurso del tiempo. El objetivo de este tipo de procesos consiste en determinar una política que optimice el criterio de rendimiento. A dicha política se le llama *política óptima* y el *criterio de rendimiento* evaluado en la política óptima se le llama *función de valor óptimo*. Un método muy usado para resolver este tipo de problemas es la *programación dinámica* (véase [7], [9], [25], [45]). A través de la programación dinámica una forma muy usada para el cálculo de la política óptima, si ésta existe, es el método de iteración de valores (véase [25]).

Por otro lado, el principio variacional de Ekeland o "principio variacional" afirma que existen soluciones óptimas cercanas para algún problema de minimización (véase [21]). Es decir, este principio establece que para alguna función real, bajo ciertas condiciones, se le puede hacer una pequeña perturbación para hacer que alcance su mínimo (véase [12], [21] o [48]) e incluso cuando el conjunto sobre el que se está minimizando no sea compacto. Además, este minimizador resulta ser único para la función perturbada. El principio de Ekeland se basa en la completez del espacio métrico y también caracteriza la completez.

La regularización de Moreau-Yosida en una técnica de suavizado de funciones (véase [8], [12], [48]), las cuales a pesar de no ser diferenciables, con este tipo de regularización se convierten en funciones diferenciables, estrictamente convexas y finitas. Además el minimizador también resulta ser único usando la regularización de Moreau-Yosida. Es importante mencionar que para probar la diferenciabilidad en la regularización de Moreau-Yosida se prueba que es subdiferenciable y superdiferenciable con lo que se concluye que la regularización es diferenciable.

También, desde el punto de vista clásico (por ejemplo, el concepto de Hadamard sobre problemas bien planteados [1]) en un modelo matemático, la existencia y unicidad de soluciones siempre han sido las principales preocupaciones de la matemática aplicada. Sin embargo, en muchos ejemplos de optimización no siempre es posible garantizar tanto la existencia como la unicidad, inclusive si se tienen condiciones las cuales aseguren la existencia de soluciones, la unicidad de éstas no se obtiene de manera automática. Por ejemplo, en la programación lineal, incluso tenemos el caso extremo de que cuando hay dos vectores óptimos diferentes, todas sus combinaciones lineales convexas resultan también ser óptimas automáticamente. Un método muy común, para calcular los puntos que optimizan a dichos problemas, es el cálculo diferencial, pero en la literatura existen muchos problemas los cuales no presentan diferenciabilidad, aún cuando el problema sí tenga solución. Usando el método de iteración de valores en PDMs, surge la siguiente pregunta, suponiendo que existe una política óptima, ¿será que siempre o bajo qué condiciones las políticas de iteración de valores convergen a la política óptima? y si esto pasa, ¿la convergencia será puntual o uniforme? y ¿será posible determinar una cota de convergencia la cual nos ayude a hacer estimaciones sobre la política óptima única.

El objetivo principal de la tesis es para un PDM, con costo total descontado, al que llamaremos proceso original asociado a un MCM dado, se supondrá que este PDM tiene una política óptima, no necesariamente única y queremos encontrar un nuevo PDM al que llamaremos proceso perturbado, el cual tenga las mismas componentes que el MCM original, excepto en la función de costo. Esta función de costo se perturba adicionando una función adecuada, de tal manera que en el proceso perturbado se tengan ciertas propiedades estructurales de la política óptima y de la función de valor óptimo, tales como: la unicidad de la política óptima, la diferenciabilidad de la política óptima y de la función de valor óptimo y también que las políticas provenientes del método de iteración de valores converjan a la política óptima en el proceso proceso perturbado. Además, lo ideal es que las funciones de valor óptimo y las políticas óptimas sean las mismas para ambos modelos.

La no unicidad de la política óptima puede ser intrínseca al sistema o puede ser *estructuralmente no estable*, es decir, una versión ligeramente perturbada (dicho de otra manera, adicionando una función adecuada a la función de costo) del sistema original puede presentar unicidad de la política óptima. En esta tesis se analiza si la *propiedad de unicidad de la política óptima* *es genérica*, es decir, si para una familia de procesos de decisión de Markov descontados, siempre en el caso de que un PDM no presente unicidad, puede perturbarse de tal manera que el modelo perturbado y el modelo original tengan la misma política óptima y la misma función de valor óptimo y además el sistema perturbado tenga una política óptima única. Anteriormente se obtuvieron resultados en este sentido bajo restricciones de convexidad (véase [38]). También en [3] se obtienen resultados para la singularidad(unicidad) de políticas usando métodos de programación lineal. Sin embargo, en esta tesis, usando el principio variacional de Ekeland (véase [21], Sección 2.1, pág. 6 en [12], y la Proposición 1.43, p. 31 en [48]), es posible obtener nuevos resultados interesantes, sin las restricciones de convexidad (véase [40]), lo que hace que el artículo [38] sea obsoleto. Además, también se podrá encontrar una versión perturbada donde, además de unicidad, también se garantice la diferenciabilidad para poder calcular la política óptima y que tanto, la función de valor óptimo como la política óptima, sean diferenciables.

Cabe destacar que el resultado que se presenta en esta tesis, con respecto a la perturbación basada en el principio variacional de Ekeland, no es una consecuencia directa de este teorema. Sin embargo, se sigue de cerca el patrón de la prueba original de Ekeland, que, como él mismo afirma, es una adaptación de [11] (véase también el comentario 2.1.4, p. 9 en [12]). En particular, es necesario hacer la prueba paso a paso para establecer la medibilidad de la política óptima, lo cual no se sigue automáticamente del resultado clásico (véase [40]). Es importante mencionar que en este caso, ambos, el espacio de estados y de acciones son de Borel.

Con respecto a la regularización de Moreau-Yosida (véase [8], [12], [48]), anteriormente para PDMs se obtuvieron resultados en los cuales se implementan las técnicas de diferenciabilidad para obtener una política óptima (véase [19]), sin embargo estos resultados solo se pueden implementar cuando tenemos diferenciabilidad en la segunda variable, del lado derecho de la ecuación de programación dinámica (véase [7], [9], [25], [45]). Además, se garantiza que tanto la política óptima como la función de valor óptimo son diferenciables (véase [19]). Es importante mencionar que en [19] se pide que el costo sea de clase C^n, sin embargo en esta tesis se obtiene un nuevo resultado sin esta restricción y adicionalmente se garantiza la existencia de una política óptima única para cuando el espacio de estados y de acciones son subconjuntos de $\mathbb{R}$. En cuanto a la convergencia de las políticas provenientes del método de iteración de valores se tienen otros resultados en [16] y [17], que presentan condiciones bajo las cuales es posible garantizar que las políticas de iteración de valores convergen a la política óptima; cabe mencionar que la unicidad de la política óptima y la compacidad sobre los conjuntos de acciones admisibles es fundamental para tal convergencia. Sin embargo en esta tesis, a diferencia de [16] y [17], tanto la convergencia de las políticas de iteración de valores como las cotas de convergencia correspondientes se obtienen como una consecuencia de la perturbación basada en el método de Moreau-Yosida.

En este trabajo de tesis, también se presenta un capítulo de ejemplos, algunos de tipo teórico para el caso de la perturbación basada en el principio variacional de Ekeland, los que muestran que, con o sin restricciones de convexidad y sin tener unicidad en la política óptima se puede encontrar un nuevo modelo en donde sí se obtiene unicidad de la política óptima. En cuanto a la perturbación basada en la regularización de Moreau-Yosida se dan ejemplos teóricos los cuales muestran cómo obtener un modelo perturbado con propiedades adecuadas de diferenciabilidad. Es importante mencionar que se presenta un ejemplo el cual muestra que este tipo de perturbación también proporciona un nuevo método para la aproximación de la política óptima. Además, también se aborda otro ejemplo muy conocido en la teoría de PDMs, que es el modelo lineal cuadrático (véase [25]) para el cual se encuentran cotas de convergencia para las políticas de iteración de valores.

A continuación presentamos algunos antecedentes para PDMs de las metodologías usadas en esta tesis.

En lo que respecta a la perturbación basada en el principio variacional de Ekeland, anteriormente se encontró un modelo perturbado (véase [38]) donde se probó la unicidad en este nuevo modelo bajo hipótesis de convexidad. Cabe mencionar que este tipo de perturbación no esta basada en el principio variacional de Ekeland o la regularización de Moreau-Yosida, sin embargo los resultados obtenidos en [38] son un caso particular de ambas perturbaciones. Además, también en la literatura de PDMs se tenían resultados (véase [16]) donde se encontraban condiciones para obtener la unicidad de la política óptima, sin embargo éstas son muy restrictivas. Por otro lado, Tanaka (véase [54]) trabajó este tipo de perturbación aplicada a PDMs, sin embargo en su trabajo nunca se explotó la unicidad de la política óptima para el proceso perturbado. Es importante mencionar que este tipo de perturbación para PDMs no es una consecuencia inmediata del teorema de Ekeland, debido a que se debe probar la medibilidad de la política óptima para el proceso perturbado.

Ahora, cabe aclara que en la perturbación basada en el principio variacional de Ekeland a pesar de que se presente diferenciabilidad en el lado derecho de la ecuación de programación dinámica con este tipo de perturbación no necesariamente es diferenciable, entonces se usa otro tipo de perturbación, para garantizar diferenciabilidad, basada en la regularización de Moreau-Yosida. Cabe señalar que no se encontró literatura de antecedentes en su aplicación a PDMs. En el artículo [19] se trata la diferenciabilidad en PDMs pero se le pide al costo que sea diferenciable, sin embargo en esta tesis se puede aplicar para cualquier función de costo, incluyendo alguna que no sea diferenciable. También como consecuencia de la regularización de Moreau-Yosida se dan cotas de convergencia para las políticas provenientes del método de iteración de valores. Anteriormente en [17] se dieron condiciones para la convergencia de éstas a la política óptima.

La tesis está organizada de la siguiente manera: en primer lugar se presenta la notación de la tesis y a continuación se introduce un capítulo de preliminares donde se dan algunos resultados sobre procesos de decisión de Markov descontados, el principio variacional de Ekeland y la regularización de Moreau-Yosida. Luego se dan los antecedentes que se encontraron en la literatura sobre el trabajo y entonces se dan los resultados principales, como son la perturbación basada en el principio variacional de Ekeland y la regularización de Moreau-Yosida. Posteriormente se presentan ejemplos usando ambas perturbaciones y finalmente se presentan las conclusiones y la bibliografía.

ÍNDICE GENERAL

NOTACIÓN

$\arg\min f$: Conjunto de puntos donde la función f alcanza el mínimo

δ : Métrica para el espacio de estados X

$\langle \cdot, \cdot \rangle$: Producto interior

$\mathbb{B}(z, \varepsilon)$: Bola centrada en z y de radio ε

$\mathbb{F}$: Conjunto de políticas estacionarias

$\mathbb{H}_t$: Espacio de historias observadas del proceso hasta el tiempo t

$\mathbb{K}$: Conjunto de parejas estados-acciones admisibles

$\mathcal{B}(\mathbb{R}^n)$: σ-álgebra de Borel de $\mathbb{R}^n$

$\mathcal{B}(A)$: σ-álgebra de Borel de A

$\overline{\mathbb{R}}$: Números reales extendidos

$\|\cdot\|$: Norma euclidiana en $\mathbb{R}^n$

$\partial f(x)$: Subdiferencial de f en el punto x

Π : Conjunto de políticas

Π_{DM} : Conjunto de políticas deterministas Markovianas

Π_{DS} : Conjunto de políticas deterministas Markovianas estacionarias

Π_D : Conjunto de políticas deterministas

Π_{RM} : Conjunto de políticas Markovianas aleatorizadas

Π_{RS} : Conjunto de políticas Markovianas aleatorizadas estacionarias

φ : Métrica para el espacio de acciones A

$\widehat{c}$: Función de costo perturbada basada en la regularización de Moreau-Yosida.

$A(x)$: Conjunto de acciones admisibles para cada $x \in X$

A : Espacio de acciones

c : Función de costo en un paso

c^*: Función de costo perturbada basada en el principio variacional de Ekeland

$epi(f)$: Epigrafo de la función f

h_t : Historia observada hasta el tiempo t

I_C : Función indicadora del conjunto C

M: Proceso de decisión de Markov original

M_λ: Proceso de decisión de Markov perturbado basado en la regularización de Moreau-Yosida

M_ε: Proceso de decisión de Markov perturbado basado en el principio variacional de Ekeland

$Q(\cdot|\cdot)$: Ley de transición

X : Espacio de estados

X^* : Espacio dual de X

i.i.d : independientes e idénticamente distribuidas.

i.s.c. : Inferiomente semicontinua

Capítulo 1

PRELIMINARES

1.1 Procesos de decisión de Markov descontados

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

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

donde, X y A son espacios de Borel, llamados espacio de estados y espacio de acciones (o controles), respectivamente. $\{ A(x) | \, x \in X \}$ es una familia de subconjuntos $A(x)$ de A medibles y no vacíos, donde $A(x)$ denota al conjunto de acciones admisibles cuando el sistema se encuentra en el estado $x \in X$. El conjunto $\mathbb{K}$ de parejas de estados-acciones admisibles, está definido por

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

$Q(\cdot | \cdot)$ es llamada la *ley de transición*, es un kérnel estocástico definido en X dado $\mathbb{K}$, y $c : \mathbb{K} \to \mathbb{R}$ es una función medible, llamada la función de costo en un paso. En esta tesis asumiremos que (X, δ) y (A, φ) son ambos espacios de Borel, donde δ y φ son las métricas para X y A, respectivamente.

Para introducir el concepto de estrategia o política, considérese un MCM y defina $\mathbb{H}_t$, el *espacio de las historias* observadas del proceso hasta el tiempo t, como $\mathbb{H}_0 = X$, y $\mathbb{H}_t = \mathbb{K} \times \mathbb{H}_{t-1}$, para $t = 1, 2, \ldots$. Un elemento de $\mathbb{H}_t$ llamado *t-historia* es un vector de la forma

$$h_t = (x_0, a_0, x_1, a_1, \ldots, x_{t-1}, a_{t-1}, x_t),$$

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

1.1.1 Políticas

Una *política* es una sucesión $\pi = \{\pi_t\}$ de kérneles estocásticos, donde cada π_t está definido sobre A dado $\mathbb{H}_t$ y satisface que:

$$\pi_t(\, A(x_t)|\, h_t) = 1, \tag{1.1}$$

para toda $h_t \in \mathbb{H}_t$ y $t = 0, 1, 2, \ldots$. El conjunto de todas las políticas es denotado por Π .

Sea

$$\mathbb{F} := \{f : X \to A \mid f \text{ es medible y } f(x) \in A(x), \text{ para todo } x \in X\}\,.$$

A los elementos de $\mathbb{F}$ se les conoce como *funciones de decisión* o *selectores.*

Se dice que un kérnel estocástico π, definido sobre A dado $\mathbb{H}_t$, está concentrada en g, donde $g : X \to A$ satisface que $g(x) \in A(x)$ para todo $x \in X$ y g es una función medible en X, si $\pi(C|\, h_t) = I_C(g(x_t))$ para cada $C \in \mathcal{B}(A)$ y toda $h_t \in \mathbb{H}_t$. Donde, I_C denota la función indicadora sobre C y $\mathcal{B}(A)$ denota la σ-álgebra de Borel de A.

Una política $\pi = \{\pi_t\}$ es:

a) *Markoviana Aleatorizada.* Si existe una sucesión $\{\varphi_t\}$ de kérneles estocásticos, con φ_t definida sobre A dado X, tales que, $\pi_t(\cdot|\, h_t) = \varphi_t(\cdot|\, x_t)$, para toda $h_t \in \mathbb{H}_t$ y $t = 0, 1, 2, \ldots$. El conjunto de políticas Markovianas aleatorizadas será denotado por Π_{RM}.

b) *Markoviana Aleatorizada Estacionaria.* Si existe un kérnel estocástico φ sobre A dado X, tal que $\pi_t(\cdot|\, h_t) = \varphi(\cdot|\, x_t)$, para toda $h_t \in \mathbb{H}_t$ y $t = 0, 1, 2, \ldots$. El conjunto de políticas Markovianas aleatorizadas estacionarias será denotado por Π_{RS}.

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

d) *Determinista Markoviana.* Si existe una sucesión $\{f_t\} \subset \mathbb{F}$, tal que $\pi_t(\cdot|\, h_t)$ está concentrada en $f_t(x_t)$, para cada $h_t \in \mathbb{H}_t$ y $t = 0, 1, 2, \ldots$. El conjunto de políticas deterministas Markovianas será denotado por Π_{DM}.

e) *Estacionaria.* Si existe $f \in \mathbb{F}$ tal que $\pi_t(\cdot|\, h_t)$ está concentrada en $f(x_t)$, para cada $h_t \in \mathbb{H}_t$ y $t = 0, 1, 2, \ldots$. En este caso π es denotada por f. Π_S denotará el conjunto de políticas deterministas Markovianas estacionarias.

Observación 1.1.1 *Obsérvese que* $\Pi_{RS} \subset \Pi_{RM} \subset \Pi$ *y* $\Pi_S \subset \Pi_{DM} \subset \Pi_D \subset \Pi$.

Sea $(\Omega, \mathfrak{F})$ un espacio medible que consiste del espacio muestral canónico $\Omega := \overline{H}_\infty = (X \times A)^\infty$ y $\mathfrak{F}$ su correspondiente σ-álgebra producto (véase [2]). Los elementos de Ω son de la forma $w = (x_0, a_0, x_1, a_1, \ldots)$ con $x_t \in X$ y $a_t \in A$ para toda $t = 0, 1, \ldots$, denotando simultáneamente elementos en X y A y también las proyecciones x_t y a_t de Ω sobre X y A son llamados estado y acción, respectivamente. Obsérvese que $H_\infty := \mathbb{K}^\infty \subset \Omega$.

Sean $\pi \in \Pi$ una política arbitraria, υ una medida de probabilidad arbitraria en X y $x_0 = x \in X$. Entonces por el Teorema de Ionescu-Tulcea (véase [2], [25]), existe una única medida de probabilidad P_υ^π sobre $(\Omega, \mathfrak{F})$ tal que está concentrada en H_∞, es decir, $P_\upsilon^\pi(H_\infty) = 1$. Además, para cada $C \in \mathcal{B}(A)$, $B \in \mathcal{B}(X)$, $h_t \in \mathbb{H}_t$ y $t = 0, 1, 2, \ldots,$ se tiene que

$$P_\upsilon^\pi(x_0 \in B) = \upsilon(B), \tag{1.2}$$

$$P_\upsilon^\pi(a_t \in C|\, h_t) = \pi_t(C|\, h_t), \tag{1.3}$$

$$P_\upsilon^\pi(x_{t+1} \in B|\, h_t, a_t) = Q(B|\, x_t, a_t). \tag{1.4}$$

El proceso estocástico $((\Omega, \mathfrak{F}, P_\upsilon^\pi), \{x_t\})$ es llamado un *Proceso de Control de Markov a tiempo discreto* o *Proceso de Decisión de Markov (PDM).* La esperanza con respecto a P_υ^π es denotada por E_υ^π.

Observación 1.1.2 *En general, en lugar de dar* $x_0 = x \in X$*, se puede conocer una medida de probabilidad* ν *sobre* X*, referida como distribución inicial, la cual cumple que*

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

para cada $B \in \mathcal{B}(X)$*. Además, por (1.2), (1.3) y (1.4) se obtiene que para una distribución inicial* υ *y* $\pi = \{\varphi_t\} \in \Pi_{RM}$*,* $\{x_t\}$ *es un proceso de Markov no homogéneo con kérneles de transición* $\{Q(\cdot|\, x, \varphi_t)\}$*, esto es,*

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

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

Un *criterio de rendimiento* mide la calidad de las políticas aplicadas al proceso. En este trabajo se considera el criterio de *Costo Total Descontado*, el cual está definido para cada $x \in X$ y $\pi \in \Pi$ como

$$V(\pi, x) := E_x^\pi \left[\sum_{t=0}^{\infty} \alpha^t c(x_t, a_t) \right], \tag{1.5}$$

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

Definición 1.1.3 *Una política* $\pi^* \in \Pi$ *es óptima, si para cada* $x \in X$*,*

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

La función definida para cada $x \in X$ *como*

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

es llamada función de valor óptimo.

El *Problema de Control Óptimo* (PCO) consiste en determinar una política óptima si es que ésta existe.

1.1.2 Programación dinámica

En la literatura existente de PDMs se encuentra una herramienta esencial para resolver el problema de control óptimo, conocida como Programación Dinámica (PD) (véase [7], [9], [25], [45]). Bajo condiciones adecuadas sobre el MCM este procedimiento permite determinar la función de valor óptimo y/o a la política óptima.

Definición 1.1.4 *Diremos que una función medible* $\Theta : X \to \mathbb{R}$ *es solución de la Ecuación de Optimalidad (EO), si satisface que*

$$\Theta(x) = \inf_{a \in A(x)} \left\{ c(x,a) + \alpha \int \Theta(y) Q(dy|\, x, a) \right\}, \tag{1.6}$$

para cada $x \in X$*.*

Condición 1.1.5 *(a) c es inferiormente semicontinuo (i.s.c.), $c \geq 0$ e inferiormente compacto (i.e., para cada $x \in X$ y $r \in \mathbb{R}$, el conjunto $\{a \in A(x) \mid c(x,a) \leq r\}$ es compacto).*

(b) La ley de transición Q es fuertemente continua, i.e.,

$$\theta(x,a) = \int_X u(y) Q(dy|x,a),$$

es continua para $(x,a) \in \mathbb{K}$ y acotada por abajo en $\mathbb{K}$, para cada función medible y acotada $u : X \to \mathbb{K}$.

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

Lema 1.1.6 *([25], Teorema 4.2.3) Supóngase que la Condición 1.1.5 se satisface. Entonces*

(a) La función de valor óptimo υ^* es una solución de

$$\upsilon^*(x) = \min_{a \in A(x)} \left[c(x,a) + \alpha \int_X \upsilon^*(y) Q(\, dy|\, x,a) \right], \tag{1.7}$$

para cada $x \in X$ y, si μ es otra solución de la ecuación, entonces $\mu(\cdot) \geq \upsilon^*(\cdot)$.

(b) Existe un selector $f^* \in \mathbb{F}$ tal que en (1.7) se alcanza el mínimo, i.e., para cada $x \in X$,

$$\upsilon^*(x) = c(x, f^*(x)) + \alpha \int_X \upsilon^*(y) Q(\, dy|\, x, f^*(x)) \tag{1.8}$$

y f^* es óptima.

(c) Para todo $x \in X$, $V_n(x) \uparrow \upsilon^*(x)$, con V_n definido como

$$V_n(x) = \min_{a \in A(x)} \left\{ c(x,a) + \alpha \int_X V_{n-1}(y) Q(dy|x,a) \right\},$$

$x \in X$, $n = 1, 2, \ldots$ y $V_0 \equiv 0$. Además, para cada $n \in \mathbb{N}$ existe $f_n \in \mathbb{F}$ tal que

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

para todo $x \in X$.

Notación 1.1.7 *Para cada $(x,a) \in \mathbb{K}$, G y G_n denotarán lo siguiente*

$$G(x,a) := c(x,a) + \alpha \int_X \upsilon^*(y) Q(dy|x,a),$$

y

$$G_n(x,a) := c(x,a) + \alpha \int_X V_n(y) Q(dy|x,a).$$

1.2 Principio variacional de Ekeland

En esta sección se enunciará y se probará el principio variacional de Ekeland. Además, también se dará otra versión del principio de Ekeland, donde la unicidad del minimizador se encuentra de manera explícita (véase [12], [21] ó [48]).

Teorema 1.2.1 (Principio variacional de Ekeland (véase [21])) *Sea (X,δ) un espacio métrico completo y sea $f : X \to \mathbb{R} \cup \{+\infty\}$ una función inferiormente semicontinua y acotada por abajo. Supóngase que para cada $\varepsilon > 0$ existe $z \in X$ tal que*

$$f(z) < \inf_{x \in X} f(x) + \varepsilon.$$

Entonces, existe $y \in X$ tal que:

(i) $\delta(z,y) \leq 1$,

(ii) $f(y) + \varepsilon\delta(z,y) \leq f(z)$ *y*

(iii) $f(x) + \varepsilon\delta(x,y) > f(y)$ *para todo* $x \in X \setminus \{y\}$.

Demostración. Se definirá una sucesión $\{z_i\}$ de manera inductiva, comenzando con $z_0 := z$. Supóngase que se tiene definido z_i. Defínase un conjunto S_i de la siguiente manera

$$S_i := \{x \in X \mid f(x) + \varepsilon\delta(x,z_i) \leq f(z_i)\}$$

y considérese dos posibles casos:

(a) Si $\inf_{S_i} f = f(z_i)$. Entonces se define $z_{i+1} = z_i$.

(b) Si $\inf\limits_{S_i} f < f(z_i)$, se elige $z_{i+1} \in S_i$ tal que

$$f(z_{i+1}) < \inf_{S_i} f + \frac{1}{2}\left[f(z_i) - \inf_{S_i} f\right] = \frac{1}{2}\left[f(z_i) + \inf_{S_i} f\right] < f(z_i). \quad (1.9)$$

Se demostrará que $\{z_i\}$ es una sucesión de Cauchy. En efecto, si (a) se cumple entonces z_i es la misma para i suficientemente grande, por lo tanto es de Cauchy. De otra forma, de la definición de S_i y de z_{i+1}, se obtiene que

$$\varepsilon\delta(z_i, z_{i+1}) \leq f(z_i) - f(z_{i+1}). \quad (1.10)$$

Sumando (1.10) desde i a $j-1 > i$ se tiene que

$$\varepsilon\delta(z_i, z_j) \leq f(z_i) - f(z_j). \quad (1.11)$$

Obsérvese que la sucesión $\{f(z_i)\}$ es decreciente y acotada por abajo por el $\inf\limits_X f$, por lo tanto convergente. Se concluye de (1.11) que $\{z_i\}$ es de Cauchy. Sea $y := \lim\limits_{i\to\infty} z_i$. Se probará que y satisface las conclusiones del teorema. Tomando $i = 0$ en (1.11), se tiene que

$$\varepsilon\delta(z, z_j) + f(z_j) \leq f(z). \quad (1.12)$$

Tomando el límite cuando $j \to \infty$ se llega a (ii). Como $f(z) - f(y) \leq f(z) - \inf_X f < \varepsilon$, (i) se sigue de (ii). Sólo resta probar que y satisface (iii). Sea i fijo en (1.11) y tómese el límite cuando $j \to \infty$ se concluye que $y \in S_i$. Es decir,

$$y \in \bigcap_{i=1}^{\infty} S_i.$$

Por otro lado, si $x \in \bigcap\limits_{i=1}^{\infty} S_i$ entonces, para todo $i = 1, 2, \ldots,$

$$\varepsilon\delta(x, z_{i+1}) \leq f(z_{i+1}) - f(x) \leq f(z_{i+1}) - \inf_{S_i} f. \quad (1.13)$$

Se sigue de (1.9) que $f(z_{i+1}) - \inf\limits_{S_i} f \leq f(z_i) - f(z_{i+1})$, y por lo tanto

$$\lim_{i\to\infty}\left[f(z_{i+1}) - \inf_{S_i} f\right] = 0.$$

Tomando límites en (1.13) cuando $i \to \infty$ se tiene que $\varepsilon\delta(x, y) = 0$. De esto se obtiene que

$$\bigcap_{i=1}^{\infty} S_i = \{y\}. \tag{1.14}$$

Nótese que por construcción la sucesión de conjuntos $\{S_i\}$ son anidados, i.e., para todo i, $S_{i+1} \subset S_i$. En efecto, para algún $x \in S_{i+1}$, tal que $f(x) + \varepsilon\delta(x, z_{i+1}) \leq f(z_{i+1})$ y $z_{i+1} \in S_i$ se concluye que

$$\begin{aligned} f(x) + \varepsilon\delta(x, z_i) &\leq f(x) + \varepsilon\delta(x, z_{i+1}) + \varepsilon\delta(z_i, z_{i+1}) \\ &\leq f(z_{i+1}) + \varepsilon\delta(z_i, z_{i+1}) \leq f(z_i), \end{aligned} \tag{1.15}$$

lo cual implica que $x \in S_i$. Ahora, para algún $x \neq y$, se obtiene de (1.14) que cuando i es suficientemente grande $x \notin S_i$. Por lo tanto, $f(x) + \varepsilon\delta(x, z_i) \geq f(z_i)$. Tomando límite cuando $i \to \infty$ se llega a la desigualdad (iii). ■

Nota 1.2.2 *El principio variacional de Ekeland o "principio variacional" es un teorema que fue propuesto por Ivar Ekeland el cual afirma que existen soluciones óptimas cercanas para algún problema de minimización (véase [21]). Es decir, este principio establece que para alguna función real inferiormente semicontinua y acotada por abajo (en la cual no necesariamente exista un mínimo), se le puede hacer una pequeña perturbación para hacer que alcance su mínimo (véase [12], [21] o [48]). El principio variacional de Ekeland puede ser usado cuando los conjuntos de nivel de un problema de minimización no sean compactos, lo cual significa que los teoremas de análisis clásico no pueden ser aplicados. El principio de Ekeland se basa en la completez del espacio métrico y además la caracteriza. Más aún, el principio variacional provee una herramienta poderosa en análisis moderno. Sus aplicaciones cubren numerosas áreas incluyendo optimización, geometría en espacios de Banach, análisis no suave, economía, teoría de control y teoría de juegos, por nombrar algunos (véase [12] y [21]).*

Es importante mencionar que el caso (iii) del Teorema 1.2.1, establece la unicidad del minimizador para la función perturbada. El siguiente teorema es una forma equivalente del Teorema 1.2.1, en el cual se da de manera explícita la unicidad del minimizador, la demostración puede consultarse en la Proposición 1.43 de [48].

Teorema 1.2.3 *Sea $f : \mathbb{R}^n \to \mathbb{R}$ una función i.s.c. con $\inf f < +\infty$. Supóngase que para cada $\varepsilon > 0$ y $z \in X$ se satisface que*

$$f(z) < \inf_{x \in X} f(x) + \varepsilon.$$

Entonces, existe $y \in X$ tal que

(a) $y \in \mathbb{B}(z, \varepsilon)$ *con* $f(y) \leq f(z)$; *y*

(b) $\arg\min_{x \in X} \{f(x) + \varepsilon\delta(x, y)\} = \{y\}$

donde, $\mathbb{B}(z, \varepsilon)$ representa la bola centrada en z y de radio ε, y arg mín *es el conjunto de puntos donde la función alcanza el mínimo.*

1.3 Regularización basada en envolventes de Moreau-Yosida

Un método muy común para calcular el punto mínimo o los puntos mínimos de una función cuando éstos existen es mediante las técnicas del cálculo diferencial. Sin embargo, se han encontrado muchos problemas de optimización en los cuales la función a minimizar resulta ser no diferenciable en el(los) punto(s) mínimo(s). Una forma de poder resolver los problemas en los cuales la función a optimizar no es diferenciable, es usando el concepto de subdiferencial (véase [12] y [48]).

La subdiferencial (véase Definición 1.3.5 abajo) de una función en un punto es un conjunto, el cual se considera el sustituto de la noción de derivada, en donde la función objetivo no necesariamente es diferenciable. La subdiferencial es un concepto muy importante, debido a que muchos resultados del cálculo se pueden extender, lo cual hace que sea una herramienta útil cuando se hace algún tipo de análisis sobre funciones no diferenciables.

Una forma de poder resolver los problemas de optimización no diferenciable, en los cuales se conoce de manera explícita el conjunto de puntos donde la función no es diferenciable (por ejemplo, las funciones con valores absolutos), es usar una técnica especial de suavizado, por ejemplo, la regularización de Moreau-Yosida (véase [48]), la cual se desarrollará en esta sección.

En esta sección se usará la siguiente notación: $\langle \cdot, \cdot \rangle$ denota el producto interior y $\| \cdot \|$ la norma euclidiana en $\mathbb{R}^n$. También, $\overline{\mathbb{R}}$ representa los números reales extendidos, i.e., $\overline{\mathbb{R}} = [-\infty, +\infty]$. Además, $\mathbb{R}^n$ lo consideraremos como un espacio medible con $\mathcal{B}(\mathbb{R}^n)$ la σ-álgebra de Borel de $\mathbb{R}^n$.

1.3.1 Subdiferenciabilidad y Convexidad

Sea $g : \mathbb{R}^n \to \mathbb{R}$ una función definida en al menos una vecindad del 0. Se dice que $g(h)$ es $o(h)$ y se escribe $g(h) = o(h)$ para indicar que

$$\lim_{h\to 0} \frac{g(h)}{\| h \|} = 0.$$

Definición 1.3.1 *Sea $f : \mathbb{R}^n \to \mathbb{R}$ una función. Se dice que una función f es diferenciable en $x \in \mathbb{R}^n$, si existe $y \in \mathbb{R}^n$ tal que*

$$f(x+h) - f(x) = \langle y, h\rangle + o(h),$$

para toda $h \in \mathbb{R}^n$.

Definición 1.3.2 *Sea $f : \mathbb{R}^n \to \mathbb{R}$ una función.*

(a) Un conjunto $C \subseteq \mathbb{R}^n$ es convexo si para todo $x, y \in C$ y para todo $\beta \in [0,1]$ se tiene que

$$\beta x + (1-\beta)y \in C.$$

(b) Una función $f : \mathbb{R}^n \to \mathbb{R}$ es convexa, si para todo $x, y \in \mathbb{R}^n$ y $\beta \in [0,1]$,

$$f(\beta x + (1-\beta)y) \leq \beta f(x) + (1-\beta)f(y).$$

En caso de que

$$f(\beta x + (1-\beta)y) < \beta f(x) + (1-\beta)f(y),$$

se dice que la función es estrictamente convexa.

(c) Sea $f : \mathbb{R}^n \to \mathbb{R}$ una función convexa. Se define el dominio de f como el conjunto de puntos donde f es finita, es decir, $dom(f) = \{x \in \mathbb{R}^n \mid f(x) < +\infty\}$. f es propia si $dom(f) \neq \emptyset$.

(d) Para $f : \mathbb{R}^n \to \mathbb{R}$ una función, se define el epigrafo como

$$epi(f) = \{(x,t) \in \mathbb{R}^n \times \mathbb{R} \mid x \in dom(f), t \geq f(x)\}.$$

La demostración del siguiente Lema puede consultarse en [48] (Teorema 2.9).

Lema 1.3.3 *Sea $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ una función convexa. Suponga que se quiere minimizar a f sobre $\mathbb{R}^n$, entonces cada solución localmente óptima es globalmente óptima y el conjunto* $\arg\min f$ *es convexo. Más aún, si f es estrictamente convexa y propia, entonces* $\arg\min f$*, si es no vacío, es de un sólo punto.*

La demostración del siguiente Lema puede consultarse en [47], Teorema 1.6.

Lema 1.3.4 *Sea $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ una función. Entonces, las siguientes condiciones son equivalentes*

(a) f es inferiormente semicontinua (i.s.c.) en todo $\mathbb{R}^n$.

(b) $\{x \in \mathbb{R}^n \mid f(x) \leq \gamma\}$ es cerrado, para cada $\gamma \in \mathbb{R}$.

(c) El epigrafo de f es un conjunto cerrado en $\mathbb{R}^{n+1}$.

Sea X^* el espacio dual de X (véase [49]). También $\langle \cdot, \cdot \rangle$ denota el producto interior en X o en $A(x)$ para todo $x \in X$.

Ahora, se introducirá el concepto de subdiferencial de una función convexa. Es importante mencionar que este concepto generaliza el concepto de derivada clásica (véase Ejemplo 1.3.6 abajo).

Definición 1.3.5 *La subdiferencial de una función convexa $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ en el punto $x \in \mathbb{R}^n$ está definida como*

$$\partial f(x) = \{s \in X^* \mid f(y) \geq f(x) + \langle s, y - x \rangle \ \textit{para todo } y \in \mathbb{R}^n\}.$$

En este caso en particular X^* representa el dual de $\mathbb{R}^n$. Una función se dice que es subdiferenciable, si $\partial f(x) \neq \emptyset$.

Ejemplo 1.3.6 *Sea $f : \mathbb{R} \to \mathbb{R}$ una función definida por $f(x) = \mid x \mid$. Entonces*

$$\partial f(x) = \begin{cases} \{-1\} & si \quad x < 0 \\ \\ \{1\} & si \quad x > 0 \\ \\ [-1, 1] & si \quad x = 0. \end{cases}$$

Observe que esta función no es diferenciable en 0, sin embargo sí es subdiferenciable en 0.

El siguiente lema afirma que para toda función convexa existe la subdiferencial en $x \in \mathbb{R}^n$. Para la demostración véase [8].

Lema 1.3.7 *Sea $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ una función convexa y $x \in int\ dom(f)$. Entonces, $\partial f(x) \neq \emptyset$.*

Definición 1.3.8 *Una función $f : X \to \mathbb{R}$ es* ***Fréchet diferenciable*** *en x si existe alguna $g^* \in X^*$ tal que*

$$\lim_{h \to 0} \frac{f(x+h) - f(x) - \langle g^*, h\rangle}{|h|} = 0,$$

g^* es llamada la derivada de Fréchet de f en x y puede ser escrita como $f'(x)$ o $\dfrac{df}{dx}$.

Definición 1.3.9 *La* ***subdiferencial de Fréchet*** *de $f : X \to \mathbb{R}$ es*

$$\partial_F f(x) = \left\{ g^* \in X^* : \liminf_{h \to 0} \frac{f(x+h) - f(x) - \langle g^*, h\rangle}{|h|} \geq 0 \right\},$$

y f es Fréchet subdiferenciable si $\partial_F f(x)$ es no vacío. Similarmente, la ***superdiferencial de Fréchet*** *de f es*

$$\partial^F f(x) = \left\{ g^* \in X^* : \limsup_{h \to 0} \frac{f(x+h) - f(x) - \langle g^*, h\rangle}{|h|} \leq 0 \right\},$$

y es Fréchet superdiferenciable si $\partial_F f(x)$ es no vacío.

Observación 1.3.10 *Una función f es Fréchet diferenciable en x_0 si y sólo si f es a la vez Fréchet subdiferenciable y superdiferenciable en x_0 en este caso $\{df(x_0)\} = \partial_F f(x) = \partial^F f(x)$.*

Debe observarse que la subdiferencial presentada en la definición precedente generaliza la subdiferencial del análisis convexo clásico (véase [8], [12], [48]). Recuerde que si f es una función convexa, la subdiferencial clásica de f en un punto x se define por

$$\partial f(x) = \{p \in X^* \mid \langle p, y - x\rangle \leq f(y) - f(x)\ \forall y \in X\}.$$

1.3.2 Regularización de Moreau-Yosida

Definición 1.3.11 *Sea $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ una función propia (véase Definición 1.3.2 (c)) y convexa. Se define la regularización de Moreau-Yosida de f para $\lambda > 0$ y para cada $x \in \mathbb{R}^n$ como*

$$F(x) = \min_{y\in\mathbb{R}^n} \left\{ f(y) + \frac{\lambda}{2} \parallel y - x \parallel^2 \right\}, \tag{1.16}$$

y

$$p(x) = \arg\min_{y\in\mathbb{R}^n} \left\{ f(y) + \frac{\lambda}{2} \parallel y - x \parallel^2 \right\}. \tag{1.17}$$

Donde el $\arg\min$ *está definido como en el Teorema 1.2.3. Observe que el operador proximal $p(x)$ está bien definido, ya que, la función a minimizar es estrictamente convexa (véase 1.3.14 abajo) y por lo tanto el minimizador es único (véase [8], [48]). Debido a la unicidad se usará la notación $p(x)$ como un punto. Es importante mencionar que tanto los siguientes lemas como su demostración aparecen en [8].*

Para ver más detalles del siguiente ejemplo véase [6].

Ejemplo 1.3.12 *Sea $f : \mathbb{R} \to \mathbb{R}$ una función definida por $f(x) = |x|$, para cada $x \in X$. Entonces*

$$\begin{aligned} F(x) &= \min_{y\in\mathbb{R}} \left\{ f(y) + \frac{\lambda}{2} (y-x)^2 \right\} \\ &= \min_{y\in\mathbb{R}} \left\{ |y| + \frac{\lambda}{2} (y-x)^2 \right\} \\ &= \min_{y\in\mathbb{R}} \left\{ |y| + \frac{\lambda}{2} \left(y^2 - 2xy + x^2 \right) \right\}. \end{aligned}$$

Es fácil ver que si $x \geq 0$, entonces $y \geq 0$ ó si $x < 0$, entonces $y < 0$. Entonces

$$\begin{aligned} F(x) &= \min_{\eta\geq 0} \min_{|y|=\eta} \left\{ \eta + \frac{\lambda}{2} \left(\eta^2 - 2|x|\eta + x^2 \right) \right\} \\ &= \frac{\lambda}{2} x^2 + \frac{\lambda}{2} \min_{\eta\geq 0} \min_{|y|=\eta} \left\{ \eta^2 + 2 \left(\frac{1}{\lambda} - |x| \right) \eta \right\}. \end{aligned} \tag{1.18}$$

Ahora, observe que si $\frac{1}{\lambda} - |x| \geq 0$, es decir, $|x| \leq \frac{1}{\lambda}$, entonces $\eta = 0$ y $F(x) = \frac{\lambda}{2} x^2$. Por otro lado, $\frac{1}{\lambda} - |x| < 0$, es decir, $|x| > \frac{1}{\lambda}$, entonces derivando *con respecto a* η *e igualando a cero se obtiene que*

$$2\eta + 2(\frac{1}{\lambda} - |x|) = 0.$$

Despejando η *se obtiene que* $\eta = |x| - \frac{1}{\lambda}$, *sustituyendo en la ecuación (1.18) se llega a que*

$$\begin{aligned} F(x) &= \frac{\lambda}{2}x^2 + \frac{\lambda}{2}\min_{\eta\geq 0}\min_{|y|=\eta}\left\{\eta^2 + 2\left(\frac{1}{\lambda} - |x|\right)\eta\right\} \\ &= \frac{\lambda}{2}x^2 + \frac{\lambda}{2}\left\{\left(|x| - \frac{1}{\lambda}\right)^2 + 2\left(\frac{1}{\lambda} - |x|\right)\left(|x| - \frac{1}{\lambda}\right)\right\} \\ &= |x| - \frac{1}{2\lambda}. \end{aligned}$$

Por lo tanto,

$$F(x) = \begin{cases} \frac{\lambda}{2}x^2 & si \quad |x| \leq 1/\lambda \\ |x| - \frac{1}{2\lambda} & si \quad |x| > 1/\lambda \end{cases}$$

y

$$p(x) = \begin{cases} 0 & si \quad |x| \leq 1/\lambda \\ x - \frac{1}{\lambda} & si \quad x > 1/\lambda \\ \frac{1}{\lambda} + x & si \quad x < -1/\lambda, \end{cases}$$

para cada $x \in X$.

Proposición 1.3.13 *Si* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *es una función propia, entonces* F, *dada en la Definición 1.3.11 es una función finita valuada.*

Demostración. Si f es una función propia, entonces existe $x \in \mathbb{R}^n$ tal que $f(x) < \infty$. De esta manera, F es finita valuada, ya que,

$$F(x') \leq f(x) + \frac{\lambda}{2} \parallel x - x' \parallel^2 < \infty,$$

para algún $x' \in \mathbb{R}^n$. ■

Proposición 1.3.14 *Si $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ es una función convexa, entonces F, dada en la Definición 1.3.11 es una función convexa. Además, la función definida por $g_\lambda(x,\omega) = f(\omega) + \dfrac{\lambda}{2} \parallel \omega - x \parallel^2$ es estrictamente convexa en ω.*

Demostración. Sean $x, x' \in \mathbb{R}^n$ y $\beta \in (0,1)$, entonces

$$
\begin{aligned}
& \beta F(x) + (1-\beta)F(x') \\
= \; & \beta f(p(x)) + (1-\beta) f(p(x')) \\
+ \; & \frac{\lambda}{2}\left(\beta \parallel p(x) - x \parallel^2 + (1-\beta) \parallel p(x') - x' \parallel^2\right) \\
\geq \; & f(\beta p(x) + (1-\beta)p(x')) \\
+ \; & \frac{\lambda}{2} \parallel \beta p(x) + (1-\beta)p(x') - (\beta x + (1-\beta)x') \parallel^2 \\
\geq \; & F(\beta x + (1-\beta)x')
\end{aligned}
$$

Por otro lado, nótese que la función norma $\parallel \cdot \parallel$ es estrictamente convexa (recuérdese que se está hablando de la norma euclidiana), entonces g_λ es estrictamente convexa, ya que es la suma de una función convexa con una función estrictamente convexa. ■

Corolario 1.3.15 *Si $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ es una función propia y convexa, entonces $p(x)$ es único para cada $x \in X$.*

Demostración. Es una consecuencia directa de la Proposición 1.3.14 y del Lema 1.3.3. ■

Lema 1.3.16 *Si $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ es una función i.s.c., propia y convexa, entonces $p(x)$ es continuo en $x \in \mathbb{R}^n$.*

Demostración. Véase Teorema 2.26 en [48]. ■

Proposición 1.3.17 *Si $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ es una función i.s.c., propia y convexa, entonces F, dada en la Definición 1.3.11 es una función diferenciable con gradiente, dado para $x \in \mathbb{R}^n$, por*

$$\nabla F(x) = \lambda \left(x - p(x)\right),$$

en donde, F y λ son como en la Definición 1.3.11 arriba.
Más aún,

$$\parallel \nabla F(x) - \nabla F(x') \parallel \leq \lambda \left\langle \nabla F(x) - \nabla F(x'), x - x' \right\rangle$$

y

$$\| \nabla F(x) - \nabla F(x') \| \leq \lambda \| x - x' \|,$$

para todo $x, x' \in \mathbb{R}^n$.

Demostración. Fijemos una dirección arbitraria $d \in \mathbb{R}^n$, $\| d \| = 1$, y sea $t > 0$. Por definición

$$\begin{aligned}
\frac{F(x+td) - F(x)}{t} &= \frac{\min_y \left[f(y) + \frac{\lambda}{2} \| y - x - td \|^2 \right]}{t} \\
&- \frac{\min_w \left[f(w) + \frac{\lambda}{2} \| w - x \| \right]^2}{t} \\
&\geq \frac{\left[f(p(x+td)) + \frac{\lambda}{2} \| p(x+td) - x - td \|^2 \right]}{t} \\
&- \frac{\left[f(p(x+td)) + \frac{\lambda}{2} \| p(x+td) - x \|^2 \right]}{t} \\
&= \frac{\lambda}{2} \frac{\| p(x+td) - x - td \|^2 - \| p(x+td) - x \|^2}{t} \\
&= \frac{\lambda}{2} \frac{\| p(x+td) - p(x) + p(x) - x - td \|^2}{t} \\
&- \frac{\lambda}{2} \frac{\| p(x+td) - p(x) + p(x) - x \|^2}{t} \\
&= \frac{\lambda}{2} \frac{\| p(x) - x - td \|^2 - \| p(x) - x \|^2}{t} \\
&- \lambda \langle p(x+td) - p(x), d \rangle \qquad (1.19)
\end{aligned}$$

en donde se usa que $F(x) \leq f(p(x+td)) + \frac{\lambda}{2} \| p(x+td) - x \|^2$. (Nota: Recuerde que para $\alpha, \beta \in \mathbb{R}^n$ se tiene que $\| \alpha + \beta \|^2 = \| \alpha \|^2 + 2 \langle \alpha, \beta \rangle + \| \beta \|^2$. En este caso, para probar la igualdad (1.19) sólo basta sustituir tomando $\alpha = p(x+td) - p(x)$, $\beta = p(x) - x - td$ y $\gamma = p(x) - x$.)

Ahora, tomando el límite cuando $t \longrightarrow 0$, $p(x+td) \longrightarrow p(x)$ (ya que, $p(x)$ es continuo véase Lema 1.3.16) lo cual implica que

$$\lim_{t \longrightarrow 0} \frac{F(x+td) - F(x)}{t} \geq \langle \lambda(x - p(x)), d \rangle .$$

Por otro lado, como $F(x+td) \leq f(p(x)) + \frac{\lambda}{2} \parallel p(x) - x - td \parallel^2$,

$$\begin{aligned}
\frac{F(x+td) - F(x)}{t} &= \frac{\text{mín}_y \left[f(y) + \frac{\lambda}{2} \parallel y - x - td \parallel^2 \right]}{t} \\
&- \frac{\text{mín}_w \left[f(w) + \frac{\lambda}{2} \parallel w - x \parallel \right]^2}{t} \\
&\leq \frac{\left[f(p(x)) + \frac{\lambda}{2} \parallel p(x) - x - td \parallel^2 \right]}{t} \\
&- \frac{\left[f(p(x)) + \frac{\lambda}{2} \parallel p(x) - x \parallel^2 \right]}{t} \\
&= \frac{\lambda}{2} \frac{\parallel p(x) - x - td \parallel^2 - \parallel p(x) - x \parallel^2}{t}
\end{aligned}$$

Otra vez, tomando el límite cuando $t \longrightarrow 0$,

$$\lim_{t \longrightarrow 0} \frac{F(x+td) - F(x)}{t} \leq \langle \lambda(x - p(x)), d \rangle .$$

Antes de continuar y poder probar las desigualdades restantes, obsérvese lo siguiente

$$\begin{aligned}
\parallel \nabla F(x) - \nabla F(x') \parallel^2 &= \lambda \left\langle \nabla F(x) - \nabla F(x'), x - p(x) - x' + p(x') \right\rangle \\
&= \lambda \left\langle \nabla F(x) - \nabla F(x'), x - x' \right\rangle \\
&+ \lambda \left\langle \nabla F(x) - \nabla F(x'), p(x') - p(x) \right\rangle .
\end{aligned}$$

Ahora, se probará que los términos anteriores son negativos. Para ello, se necesitará lo siguiente (conocido como la monotonicidad de $\partial f(x)$). Sea $s \in \partial f(x)$ y $s' \in \partial f(x')$, entonces $\langle s - s', x - x' \rangle \geq 0$. Para probar esto, considérese las desigualdades de los subgradientes asociados con s y s' aplicados a x y x' respectivamente,

$$f(x') \geq f(x) + \langle s, x - x' \rangle$$

y

$$f(x) \geq f(x') + \langle s, x - x' \rangle$$

Sumando estas desigualdades se obtiene que $\langle s - s', x - x' \rangle \geq 0$. Ahora, obsérvese que por la optimalidad de $p(x)$ y $p(x')$,

$$0 \in \partial \left(f(p(x)) + \frac{\lambda}{2} \parallel p(x) - x \parallel^2 \right)$$

y

$$0 \in \partial \left(f(p(x')) + \frac{\lambda}{2} \parallel p(x') - x' \parallel^2 \right).$$

Así, $\nabla F(x) = \lambda(x - p(x)) \in \partial f(p(x))$ y $\nabla F(x') = \lambda(x' - p(x')) \in \partial f(p(x'))$. Usando la monotonicidad de ∂f,

$$\langle \nabla F(x) - \nabla F(x'), p(x') - p(x) \rangle \leq 0. \tag{1.20}$$

La prueba de la última desigualdad se obtiene de la desigualdad (1.20),

$$\begin{aligned} \parallel \nabla F(x) - \nabla F(x') \parallel^2 &\leq \lambda \langle \nabla F(x) - \nabla F(x'), x - x' \rangle \\ &\leq \lambda \parallel \nabla F(x) - \nabla F(x') \parallel \parallel x - x' \parallel . \end{aligned}$$

Por lo tanto,

$$\parallel \nabla F(x) - \nabla F(x') \parallel \leq \lambda \parallel x - x' \parallel .$$

■

Capítulo 2

PROBLEMAS ABIERTOS Y ANTECEDENTES

2.1 Problemas abiertos

En esta sección se presentan los problemas que se tratarán y se resolverán es los siguientes capítulos de este trabajo. Los problemas que se abordan están motivados principalmente en perturbar el modelo original de un PDM de tal manera que el modelo perturbado presente características que nos ayuden a obtener unicidad en la política óptima o diferenciabilidad tanto en la política óptima como en la función de valor óptimo. Además, de encontrar cotas de convergencia para las políticas provenientes del método de iteración de valores.

Considérese un proceso de decisión de Markov descontado fijo, asociado a un modelo de control de Markov $(X, A, \{A(x)|\, x \in X\}, Q, c)$, el cual le llamaremos proceso original y será denotado por M. Se supondrá que el PDM original tiene una política óptima única. Ahora, dado el PDM original, se construirá un nuevo PDM con las mismas componentes del MCM asociado al proceso original, excepto en la función de costo, la cual se remplazará por otra función de costo c^* o $\widehat{c}$ que difiere del costo por una función apropiada. A la nueva función de costo le llamaremos costo perturbado. Al PDM con el MCM $(X, A, \{A(x)|\, x \in X\}, Q, c^*$ o $\widehat{c})$ le llamaremos PDM perturbado. A continuación se presentarán los problemas que se resuelven en este trabajo:

1. Encontrar formas de perturbar la función de costo c, de tal manera que la política óptima y la función de valor óptimo del PDM original y el

PDM perturbado coincidan.

2. Garantizar que en el modelo perturbado la política óptima f^* siempre sea única.

3. Dado un PDM no necesariamente diferenciable, encontrar un nuevo PDM perturbado de tal manera que se pueda garantizar la diferenciabilidad de la política óptima y de la función de valor óptimo del PDM perturbado.

4. Encontrar una forma de perturbar el PDM original, en la cual no necesariamente se tenga convexidad estricta, para poder garantizar convexidad estricta en el modelo perturbado.

5. Encontrar cotas de convergencia para las políticas provenientes del método de iteración de valores.

2.2 Antecedentes

En esta sección se dará un resumen de los antecedentes que existen en la literatura relacionados con problemas planteados en la sección anterior.

2.2.1 Condiciones para la unicidad de las políticas óptimas de un proceso de decisión de Markov

En el trabajo de Daniel Cruz-Suárez, Raúl Montes-de-Oca y Francisco Salem-Silva (véase [16]) se presentan condiciones, las cuales garantizan la unicidad de las políticas óptimas. Dichas condiciones que se presentan son sobre el espacio de estados X, el espacio de acciones A, el conjunto de acciones admisibles $A(x)$, $x \in X$, la ley de transición Q y la función de costo $c \geq 0$.
Sea X un espacio de Borel y supóngase también que X es completo y parcialmente ordenado. Por simplicidad denotaremos el orden parcial en X por "$\prec$". Más aún, se dice que una función $g : X \to \mathbb{R}$ es **creciente** si para cada $x, y \in X$, $x \prec y$ implica que $g(x) \leq g(y)$, donde $\leq$ es el orden usual en $\mathbb{R}$.

Definición 2.2.1 *Sea X un espacio de Borel completo y parcialmente ordenado. Suponga que P y P' son dos medidas de probabilidad en $(X, \mathcal{B}(X))$. Se dice que P' **domina a P estocásticamente** si $\int g dP \leq \int g dP'$ para toda*

$g : X \to \mathbb{R}$ medible, acotada y creciente. Si P' domina a P estocásticamente lo denotaremos por $P \overset{st}{\leq} P'$.

Condición 2.2.2 *(a) X es un subconjunto convexo de $\mathbb{R}^n$, para algún entero positivo n.*

(b) A es un subconjunto convexo de $\mathbb{R}^m$, para algún entero positivo m.

(c) $(1-\beta)a + \beta a' \in A((1-\beta)x + \beta x')$, para todo $x, x' \in X$, $a \in A(x)$, $a' \in A(x')$ y $\beta \in [0,1]$. Además, se supone también lo siguiente: si $x, y \in X$, $x < y$, entonces $A(x) \subset A(y)$, y $A(x)$ es convexo para todo $x \in X$.

(d) Q es inducida por la ecuación en diferencias $x_{t+1} = F(x_t, a_t, \xi_t)$, con $t = 0, 1, \ldots$, en donde $F : X \times A \times S \to X$ es una función medible y $\{\xi_t\}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas (i.i.d) con valores en $S \subseteq \mathbb{R}$ y con densidad común Δ. En resumen, se supone que $F(\cdot, \cdot, s)$ es una función convexa en $\mathbb{K}$, para cada $s \in S$; y si $x, y \in X$, $x < y$, entonces $F(x, a, s) \leq F(y, a, s)$ para cada $a \in A(x)$ y $s \in S$.

(e) c es convexo en $\mathbb{K}$, y si $x, y \in X$, $x < y$, entonces $c(x, a) \leq c(y, a)$ para cada $a \in A(x)$.

Condición 2.2.3 *(a) Lo mismo que en la Condición 2.2.2 (a) y (b).*

(b) $(1-\beta)a + a' \in A((1-\beta)x + x')$ para todo $x \in X$, $a \in A(x)$, $a' \in A(x')$ y $\beta \in [0,1]$. Además, también se supone que $A(x)$ es convexo para cada $x \in X$.

(c) Q es dada por la relación $x_{t+1} = \gamma x_t + \delta a_t + \xi_t$, $t = 0, 1, \ldots$, donde $\{\xi_t\}$ son variables aleatorias i.i.d con valores en $S \subseteq \mathbb{R}$ y con densidad Δ, γ y δ son números reales.

(d) c es estrictamente convexo en $\mathbb{K}$.

Condición 2.2.4 *(a) X es un espacio completo y parcialmente ordenado.*

(b) Suponemos lo siguiente: si $x, y \in X$, con $x < y$, entonces $A(y) \subset A(x)$.

(c) *Existe* $\overline{f} \in \mathbb{F}$, *tal que*

$$Q(\cdot \mid x, \overline{f}(x)) \overset{st}{\leq} Q(\cdot \mid x, a),$$

para todo $x \in X$, $a \in A(x)$. *Más aún, se supone que*

$$c(x, \overline{f}(x)) < c(x, a), \tag{2.1}$$

para todo $x \in X$, *y* $a \in A(x)$, $a \neq \overline{f}(x)$ *(note que, en este caso,* $\overline{f}$ *es la única política estacionaria que satisface (2.1)).*

(d) *Si* $x, y \in X$ *con* $x < y$, *entonces*

$$Q(\cdot \mid x, a) \overset{st}{\leq} Q(\cdot \mid y, a),$$

para todo $a \in A(y)$.

(e) *Si* $x, y \in X$, $x < y$, *entonces* $c(x, a) \leq c(y, a)$, *para todo* $a \in A(y)$.

Teorema 2.2.5 *Supóngase que la Condición 1.1.5 se cumple. Entonces existe una única política estacionaria bajo cada Condición 2.2.2, 2.2.3 o 2.2.4.*

2.2.2 Convergencia uniforme de las políticas de iteración de valores para procesos de decisión de Markov descontados

En el trabajo de Daniel Cruz-Suárez y Raúl Montes-de-Oca (véase [17]) se asume la unicidad de la política óptima, ésta será denotada por f^* y bajo esta condición se dan otras condiciones adicionales para demostrar la convergencia uniforme de las políticas provenientes del método de iteración de valores. Donde una condición fundamental para demostrar tanto la convergencia puntual como la convergencia uniforme es la compacidad sobre el conjunto de acciones admisibles.

Condición 2.2.6 (a) *La multifunción* $x \to A(x)$ *es semicontinua superiormente y cerrada valuada.*

(b) $c(\cdot, \cdot)$ *es una función continua en* $\mathbb{K}$.

(c) *Las integrales*

$$\int V_n(y)Q(dy \mid x,a),\; n=1,2,\ldots,$$

y

$$\int V^*(y)Q(dy \mid \cdot,\cdot),$$

son finitas y continuas en $\mathbb{K}$.

Condición 2.2.7 (a) *El espacio de acciones* A *es un conjunto compacto.*

(b) *La multifunción* $x \to A(x)$ *es compacta valuada.*

(c) *La función de costo en una etapa es estrictamente acotada, es decir, existe una sucesión no decreciente de conjuntos compactos* $X_n \uparrow X$, $n \to \infty$ *y* $A_n \uparrow A$, $n \to \infty$ *tal que* $\Lambda_n := X_n \times A_n$ *es un subconjunto de* $\mathbb{K}$, *y*

$$\lim_{n\to\infty} \inf_{(x,a)\in\Lambda_n^c} c(x,a) = +\infty,$$

donde Λ_n^c *denota el complemento de* Λ_n.

Notación 2.2.8 1. *Para* $x \in X$, *denote*

$$\widehat{A}(x) := A_{\upsilon^*(x)}(x) = \{a \in A(x) \mid c(x,a) \le \upsilon^*(x)\},$$

donde υ^* *es la función de valor óptimo como en la Definición 1.1.3.*

2. *Denote*

$$\widehat{\mathbb{K}}_\varsigma := \left\{(x,a) \in \mathbb{K} \mid x \in \varsigma, a \in \widehat{A}(x)\right\},$$

donde $\varsigma \subset X$ *es un conjunto compacto no vacío.*

Lema 2.2.9 *Asuma que las Condiciones (2.2.6(a)), (2.2.6(b)) y (2.2.6(c)) se cumplen. Entonces, cada una de las Condiciones (2.2.7(a)), (2.2.7(b)) y (2.2.7(c)) implican que* $\widehat{\mathbb{K}_\varsigma}$ *es un conjunto compacto para cada conjunto compacto no vacío* $\varsigma \subset X$.

Lema 2.2.10 *Asuma que la Condición (2.2.6) y una de las Condiciones (2.2.7(a)), (2.2.7(b)) y (2.2.7(c)) se cumple. Entonces, la política óptima estacionaria* f^* *es una función continua.*

Lema 2.2.11 *Asuma que la Condición (2.2.6) se cumple. Entonces, G_n converge uniformemente a G en cada subconjunto compacto de $\mathbb{K}$.*

Lema 2.2.12 *Bajo la Condición (2.2.6) y bajo alguna de las Condiciones (2.2.7(a)), (2.2.7(b)) y (2.2.7(c)) se cumplen. Entonces, para cada compacto no vacío $\varsigma \subset X$, se cumple lo siguiente: para cada $\varepsilon > 0$, existe $\delta > 0$ tal que, para todo $(x, a) \in \widehat{\mathbb{K}_\varsigma}$, si*

$$\mid G(x, f^*(x)) - G(x, a) \mid < \delta$$

entonces

$$\varphi(f^*(x), a) < \varepsilon.$$

Teorema 2.2.13 *Supóngase que la Condición (2.2.6) y una de las Condiciones (2.2.7(a)), (2.2.7(b)) y (2.2.7(c)) se cumplen. Entonces $\{f_n\}$ converge uniformemente sobre conjuntos compactos a f^*.*

2.2.3 Teorema de la envolvente y algunas aplicaciones a procesos de decisión de Markov descontados

En el artículo realizado por Hugo Cruz-Suárez y Raúl Montes-de-Oca (véase [19]) se presenta el teorema de la envolvente en espacios Euclideanos, con el objetivo de analizar un problema de optimización y obtener las correspondientes función de valor óptimo y política óptima, usando técnicas de diferenciación. En esta sección se definirá nuevamente un MCM, debido a que en el trabajo [19] se usa un criterio de rendimiento similar al criterio de costo total descontado descrito en el capítulo de Preliminares. En este trabajo se presenta el teorema de la envolvente para PDMs de dos maneras:

- Usando condiciones de concavidad y diferenciabilidad
- Sólo usando condiciones de diferenciabilidad.

Notaciones Preliminares:

Sean W, Y y Z espacios Euclideanos. Para algún subconjunto $B \subset W$, un punto $x \in B$ es llamado *punto interior* de B si existe un conjunto abierto U tal que $x \in U \subset B$. El *interior* de B es el conjunto de todos los puntos interiores de B y es denotado por $int(B)$.

Dado $\eta : W \to Z$ una función. La función inversa de η (si tal función existe) es denotada por η^{-1}. Además, si η es diferenciable, su derivada es denotada por η'.

Suponga que $\widehat{W}$ es un subconjunto de $W \times Y$.

Sea $\theta : \widehat{W} \to Z$ una función. Supóngase que $\theta = \theta(w, y)$ es dos veces diferenciable. Las derivadas parciales con respecto a w y y son denotadas por θ_w y θ_y, respectivamente. La notación para las segundas derivadas parciales de θ con respecto a w y y son θ_{ww} y θ_{yy}, respectivamente.

También, para $\phi = 1, 2$, defínase

$$C^{\phi}(\widehat{W}; Z) := \Big\{\theta : \widehat{W} \to Z \mid \text{la derivada parcial de orden } \phi \text{ de } \theta \\ \text{existe y es continua}\Big\}.$$

Ahora, sea $(X, A, \{A(x) \mid x \in X\}, Q, r)$ un modelo de control de Markov fijo, donde

(a) X y A son (no vacíos) ambos espacios de Borel de $\mathbb{R}^p$ y $\mathbb{R}^m$ ($p, m \geq 1$ enteros), respectivamente. X y A son llamados el espacio de estados y de acciones, respectivamente.

(b) $A(x)$ representa el conjunto de acciones admisibles en el estado x.

(c) Q es ley de transición.

(d) $r : \mathbb{K} \to \mathbb{R}$ es una función medible, llamada función de recompensa en un paso.

Ahora, para cada política π y estado inicial $x \in X$, considérese

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

$\Upsilon(\pi, x)$ es llamado el criterio de la recompensa total esperado descontado, donde $\alpha \in (0, 1)$ es el factor de descuento.

Recuerde que para PDMs, $\mathbb{K} := \{(x, a) : x \in X, a \in A(x)\}$. Sea $\Psi : \mathbb{K} \to \mathbb{R}$ una función medible. Supóngase que existe una función $\widehat{\Phi} : X \to \mathbb{R}$ tal que $\Psi(x, a) \leq \widehat{\Phi}(x)$, para todo $x \in X$ y $a \in A(x)$. Defina $\psi : X \to \mathbb{R}$ por

$$\psi(x) = \sup_{a \in A(x)} \Psi(x, a),$$

para todo $x \in X$.

Condición 2.2.14 *(a)* $\Psi \in C^2(int(\mathbb{K}), \mathbb{R})$. *Más aún,* $\Psi_{aa}(x, \cdot)$ *es definida negativa, para cada* $x \in X$.

(b) Existe una función $h : X \to A$ *tal que* $h(x) \in int(A(x))$ *y* $\psi(x) = \Psi(x, h(x))$, *para cada* $x \in X$.

Teorema 2.2.15 *Asuma que la Condición 2.2.14 se cumple. Entonces* $h \in C^1(int(X); A)$, $\psi \in C^2(int(X); \mathbb{R})$, *entonces se obtiene la siguiente fórmula de la envolvente*

$$\psi'(x) = \Psi_x(x, h(x)),$$

para todo $x \in int(X)$.

El *problema de control óptimo* consiste en determinar una política π^*, tal que:

$$\Upsilon(\pi^*, x) = \sup_{\pi \in \Pi} \Upsilon(\pi, x),$$

$x \in X$, y π^* es llamada *política óptima*. La función Φ definida por

$$\Phi(x) := \sup_{\pi \in \Pi} \Upsilon(\pi, x),$$

para todo $x \in X$, es llamada *función de valor óptimo*.

Supóngase que Q es inducida por la ecuación en diferencias

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

con $t = 0, 1, \ldots$, en donde $F : X \times A \times S \to X$ es una función medible y $\{\xi_t\}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas (i.i.d) con valores en $S \subseteq \mathbb{R}$ y con densidad común Δ y función de distribución de probabilidad μ. Obsérvese que en este caso r es una función de recompensa en un paso.

Sea $G^\infty : \mathbb{K} \to \mathbb{R}$ la función definida por

$$G^\infty(x, a) := r(x, a) + \alpha \int \Phi(y) Q(dy \mid x, a).$$

Condición 2.2.16 *(a)* $r \in C^2(int(\mathbb{K}); \mathbb{R})$.

(b) $F(\cdot, \cdot, s) \in C^2(int(\mathbb{K}); X)$ *para cada* $s \in S$. *Además tiene inversa en la tercera variable* $R : \mathbb{K} \times X \to S$ *tal que* $R(\cdot, \cdot, s) \in C^2(int(\mathbb{K} \times X); S)$ *y* $\mid detR_{(3)}(\cdot, \cdot, s) \mid \in C^2(int(\mathbb{K}); \mathbb{R})$, *para todo* s, *en donde en este caso* $R_{(3)}$ *denota la derivada de* R *en la tercera variable, y el determinante de* $R_{(3)}$ *es denotado por* $detR_{(3)}$.

(c) S es un conjunto abierto y $\Delta \in C^2(int(S); \mathbb{R})$.

(d) El intercambio entre derivadas e integrales es válido.

Lema 2.2.17 *Bajo la Condición 2.2.16, $G^\infty \in C^2(int(\mathbb{K}); \mathbb{R})$.*

Condición 2.2.18 *(a) $A(x)$ es convexo, para cada $x \in X$.*

(b) $G^\infty(x, \cdot)$ es definida negativa, para cada $x \in X$.

(c) $f(x) \in int(A(x))$, para cada $x \in X$.

Teorema 2.2.19 *Supóngase que las Condiciones 2.2.16 y 2.2.18 se cumplen. Entonces $f \in C^1(int(X); A)$ y $\Phi \in C^2(int(X); \mathbb{R})$.*

Lema 2.2.20 *Si se cumplen las Condiciones 2.2.16 y 2.2.18. Entonces, existe una única política estacionaria f tal que $f(x) \in int(A(x))$, para cada $x \in X$.*

Lema 2.2.21 *$G^\infty \in C^2(int(\mathbb{K}); \mathbb{R})$ y $G^\infty_{aa}(x, \cdot) > 0$, para cada $x \in X$.*

Lema 2.2.22 *$\Phi \in C^2(int(X; \mathbb{R})$ y $f \in C^1(int(X; A)$.*

Resultados para el Caso no Cóncavo

Condición 2.2.23 *(a) Existe una función $h : X \to A$ tal que $h(x) \in int(A(x))$ y $\psi(x) = \Psi(x, h(x))$, para cada $x \in X$.*

(b) $\Psi \in C^2(int(\mathbb{K}); \mathbb{R})$ y el determinante de $\Psi_{aa}(x, h(x))$ es diferente de cero, para cada $x \in X$.

Condición 2.2.24 *(a) f es única y $f(x) \in int(A(x))$, para cada $x \in X$.*

(b) $G^\infty \in C^2(int(\mathbb{K}); \mathbb{R})$ y el determinante de $G^\infty_{aa}(x, f(x))$ es diferente de cero, para cada $x \in X$.

Teorema 2.2.25 *Asuma que las Condiciones 2.2.23 se cumple. Entonces $h \in C^1(int(X); A)$, $\psi \in C^2(int(X); \mathbb{R})$, y la siguiente fórmula de la envolvente se cumple*

$$\psi'(x) = \Psi_x(x, h(x)),$$

para todo $x \in int(X)$.

Teorema 2.2.26 *Asuma que la Condición 2.2.16 y 2.2.24 se cumplen. Entonces $f \in C^1(int(X); A)$, y $\Phi \in C^2(int(X); \mathbb{R})$.*

2.2.4 No unicidad versus unicidad de las políticas óptimas en procesos de decisión de Markov convexos

Raúl Montes-de-Oca, Enrique Lemus-Rodríguez y Francisco Salem-Silva en el trabajo [16] proponen un método de perturbación, que se puede probar en el modelo resultante, bajo hipótesis de convexidad sobre el espacio de estados X, el conjunto de acciones A, la ley de transición Q y la función de costo c, que tiene una política óptima única. Es importante mencionar que en el trabajo [16], tanto el espacio de estados X como el conjunto de acciones A son ambos subconjuntos de $\mathbb{R}$.

Para $\eta > 0$, considere el siguiente PDM denotado por M_η con el modelo de control $(X, A, \{A(x) : x \in X\}, Q, \bar{c})$, donde $\bar{c}(x,a) = c(x,a) + \eta a^2$, $(x,a) \in \mathbb{K}$, donde c es la función de costo para M.

Condición 2.2.27 *1. X y A son convexos;*

2. $(1-\beta)a + a' \in A((1-\beta)x + x')$ para todo $x, x' \in X$, $a \in A(x)$, $a' \in A(x')$ y $\beta \in [0,1]$. Además, también se supone que: si $x, y \in X$, $x < y$, entonces $A(y) \subseteq A(x)$, y $A(x)$ son convexos para cada $x \in X$;

3. Q es inducida por la ecuación en diferencias $x_{t+1} = F(x_t, a_t, \xi_t)$, con $t = 0, 1, \ldots$, donde $F : X \times A \times S \to X$ es una función medible y $\{\xi_t\}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas (i.i.d) con valores en $S \subseteq \mathbb{R}$ y con densidad común Δ. En resumen, se supone que $F(\cdot,\cdot,s)$ es una función convexa en $\mathbb{K}$, para cada $s \in S$; y si $x, y \in X$, $x < y$, entonces $F(x,a,s) \leq F(y,a,s)$ para cada $a \in A(x)$ y $s \in S$;

4. c es convexo en $\mathbb{K}$, y si $x, y \in X$, $x < y$, entonces $c(x,a) \leq c(y,a)$ para cada $a \in A(y)$.

Condición 2.2.28 *1. Lo mismo que en la Condición 2.2.27 (1);*

2. $(1-\beta)a + a' \in A((1-\beta)x + x')$ para todo $x, x' \in X$, $a \in A(x)$, $a' \in A(x')$ y $\beta \in [0,1]$. Además, también se supone que $A(x)$ son convexos para cada $x \in X$;

3. Q es dada por la relación $x_{t+1} = \gamma x_t + \delta a_t + \xi_t$, $t = 0, 1, \ldots$, donde $\{\xi_t\}$ son variables aleatorias i.i.d con valores en $S \subseteq \mathbb{R}$ y con densidad Δ, γ y δ son números reales.

4. c es convexo en $\mathbb{K}$.

Condición 2.2.29 *Existe una política* ϕ *tal que*

$$E_x^\phi\left[\sum_{t=0}^{\infty}\alpha^t\overline{c}(x_t,a_t)\right]<\infty,$$

para cada $x\in X$.

Condición 2.2.30 *Existe una función medible* $Z:X\rightarrow\mathbb{R}$*, la cual podría depender de* α*, tal que* $\overline{c}(x,a)-c(x,a)=\eta a^2\leq\eta Z(x)$*, y* $\int Z(y)Q(dy\mid x,a)\leq Z(x)$ *para cada* $x\in X$ *y* $a\in B(x)$.

Para M_η, L^*, l^* y l_n, $n=1,2,\ldots$, denotarán la función de valor óptimo, la política óptima y las políticas provenientes de iteración de valores, respectivamente. Más aún, sea ϖ_n, $n=1,2,\ldots$, las correspondientes funciones de iteración de valores para M_η.

Teorema 2.2.31 *Supóngase que se cumplen las Condiciones 1.1.5 y 2.2.29, y que para* M*, también una de las dos Condiciones 2.2.27 ó 2.2.28 se cumplen. Sea* η *un número positivo. Entonces,*

(a) Si A *es compacto, entonces* $\mid L^*(x)-v^*(x)\mid<\eta K^2/(1-\alpha)$, $x\in X$*, donde* K *es el diámetro de un conjunto compacto* D *tal que* $0\in D$ *y* $A\subseteq D$.

(b) Bajo la Condición 2.2.30, se tiene que $\mid L^*(x)-v^*(x)\mid<\eta Z(x)/(1-\alpha)$, $x\in X$.

Corolario 2.2.32 *Supóngase que se cumplen las Condiciones 1.1.5 y 2.2.29 y que para* M*, una de las dos Condiciones 2.2.27 ó 2.2.28 se cumplen (por tanto,* M *no necesariamente tiene una política óptima única). Sea* η *un número positivo. Si* A *es compacto o la condición 2.2.30 se cumple, entonces existe un PDM* M_η *con una única política óptima* l^**, tal que las desigualdades del Teorema 2.2.31 (a) o (b) se cumplen, respectivamente.*

2.2.5 Una política ε-óptima para procesos de control estocástico a tiempo discreto

En el trabajo que se desarrolla en [54] se da una versión del principio variacional de Ekeland para procesos de decisión de Markov descontados a tiempo discreto, donde el modelo perturbado se hace en base a una política ε-óptima (véase [9], [54], [45]), sin embargo en este trabajo no se prueba la medibilidad de la política óptima para el modelo perturbado y no se trabaja la unicidad de ella.

Definición 2.2.33 *Para $\varepsilon > 0$, una política $\pi_\varepsilon \in \Pi$ es llamada ε-óptima en un estado inicial x si*

$$V(\pi_\varepsilon, x) \leq \upsilon^*(x) + \varepsilon.$$

Si π_ε es ε-óptima para cada estado inicial x, ésta es llamada ε-óptima.

En este caso se tratarán los tres siguientes casos para la función de costo.

(D) Si $0 < \alpha < 1$ y para algún $N \in \mathbb{R}$, $| c(x,a) | \leq N$ para cada $(x,a) \in \mathbb{K}$.

(P) Si $\alpha = 1$ y $0 \leq c(x,a)$, para cada $(x,a) \in \mathbb{K}$.

(N) Si $\alpha = 1$ y $c(x,a) \leq 0$, para cada $(x,a) \in \mathbb{K}$.

Lema 2.2.34 *Para un estado inicial $x \in X$ y para alguna política $\pi \in \Pi$, existe una política Markoviana $\pi_M \in \Pi_M$ tal que*

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

En el siguiente lema se demuestra la existencia de políticas ε-óptimas. La demostración se puede consultar en Bertsekas and Shreve [10], Proposición 9.19.

Lema 2.2.35 *Para cada $\varepsilon > 0$, existe una política ε-óptima markoviana no aleatorizada y si $0 < \alpha < 1$ ésta puede ser tomada estacionaria.*

Sea $N(X)$ el conjunto de todas las funciones real valuadas en X, las cuales son inferiormente semicontinuas. Definimos los operadores T y T_f en $N(X)$ de la siguiente manera: para cada $\upsilon \in N(X)$, $x \in X$ y $a \in A(x)$,

$$T\upsilon(x) = c(x,a) + \alpha \int_X \upsilon(y) Q(dy \mid x, a),$$

$$T_f v(x) = c(x, f(x)) + \alpha \int_X v(y) Q(dy \mid x, f(x)).$$

El siguiente Teorema es una consecuencia del Teorema 1.2.1. Tanaka (véase [54]) dio una demostración de este teorema.

Teorema 2.2.36 *Sea $\varepsilon > 0$. Dada la política ε-óptima f^ε del Lema 2.2.35 la cual cumple que*

$$T_{f^\varepsilon} u(x) \leq Tu(x) + \varepsilon.$$

Entonces existe $f^* \in \mathbb{F}$ tal que

(i) $\varphi(f^\varepsilon(x), f^*(x)) \leq \sqrt{\varepsilon}$,

(ii) $T_{f^*} u(x) + \sqrt{\varepsilon}\varphi(f^\varepsilon(x), f^*(x)) \leq T_{f^\varepsilon} u(x)$ *y*

(iii) $T_{f^*} u(x) < T_f u(x) + \sqrt{\varepsilon}\varphi(f(x), f^*(x))$ *para toda* $f \in \mathbb{F} \backslash \{f^*\}$.

A continuación se presentará un Teorema que ha sido aplicado por Tanaka (véase [54]) a Procesos de Decisión de Markov. En ese artículo se mencionan los siguientes resultados.

Teorema 2.2.37 *Para alguna política ε-óptima $\pi_\varepsilon \in \Pi$, existe una política no aleatorizada markoviana $\pi^* = (f_0^*, f_1^*, \ldots) \in \Pi$ tal que para todo estado inicial $x_o = x \in X$,*

$$v^*(x) + \varepsilon \sum_{k=0}^{\infty} \alpha^{k+1} E\left[\varphi(f_k^*(x_k), f_k^\varepsilon(x_k)) \mid x_0 = x\right] \leq V(\pi_\varepsilon, x)$$

y

$$\sum_{k=0}^{\infty} \alpha^{k+1} E\left[\varphi(f_k^*(x_k), f_k^\varepsilon(x_k)) \mid x_0 = x\right] \leq 1.$$

Tanaka (véase [54]) da una forma de cómo perturbar el Modelo de Control de Markov, para después enunciar el siguiente teorema:

Teorema 2.2.38 *La política $\pi^* = (f_0^*, f_1^*, \ldots) \in \Pi$ dada en el teorema anterior para una política $\pi_\varepsilon \in \Pi^{HR}$ (conjunto de políticas no aleatorizadas Markovianas) es óptima para el problema de minimización perturbado.*

Capítulo 3

PERTURBACIÓN BASADA EN EL PRINCIPIO VARIACIONAL DE EKELAND

Desde el punto de vista clásico (por ejemplo, el concepto de Hadamard sobre problemas bien planteados [1]) en un modelo matemático la existencia y unicidad de soluciones siempre han sido una de las principales preocupaciones de la matemática aplicada. Sin embargo, en muchos ejemplos de optimización no siempre es posible garantizar tanto la existencia como la unicidad, inclusive si se tienen condiciones que aseguren la existencia de soluciones, la unicidad de éstas no se obtiene de manera automática. Por ejemplo, en la programación lineal, incluso tenemos el caso extremo de que cuando hay dos vectores óptimos diferentes, todas sus combinaciones lineales convexas resultan también ser óptimas automáticamente.

La no unicidad puede ser intrínseca al sistema o puede ser *estructuralmente no estable*, es decir, una versión ligeramente perturbada del sistema original puede presentar unicidad. En este capítulo se indagará si la *propiedad es genérica*, es decir, si para alguna familia de procesos de decisión de Markov descontados (PDM), siempre en el caso de que un sistema presente no unicidad de la política óptima puede perturbarse de tal manera que la política óptima y función de valor óptimo del modelo perturbado y del modelo original coincidan y además el modelo perturbado tenga una política óptima única. Anteriormente se obtuvieron resultados en este sentido (ver [38]) bajo restricciones de convexidad. Usando el principio variacional de Ekeland (véase [21], Sección 2.1, pág. 6 en [12], y la Proposición 1.43, p. 31

en [48]), es posible obtener nuevos resultados interesantes sin las restricciones de convexidad, lo que hace que el artículo [38] sea obsoleto.

Vale la pena mencionar que el presente resultado (véase el Teorema 3.3.2 abajo) se cumple bajo condiciones bastante generales y poco restrictivas (ver Condición 1.1.5 atrás), que satisfacen una gran familia de PDMs.

Cabe destacar que el resultado principal (véase Teorema 3.3.2 abajo) de este capítulo, no es una consecuencia directa del principio variacional de Ekeland. Sin embargo, la demostración sigue de cerca el patrón de la prueba original de Ekeland, que, como él mismo afirma, es una adaptación de [11] (véase también el comentario 2.1.4, p. 9 en [12]). En particular, es necesario hacer la prueba paso a paso para establecer la medibilidad de la política óptima, lo cual no sigue automáticamente del resultado clásico.

Es importante mencionar que Tanaka (en [54]) presenta una versión del teorema de Ekeland para PDMs, relacionada con las políticas ε-optimas, pero no discute la unicidad.

Este capítulo está organizado de la siguiente manera: en primer lugar se presenta un resultado en relación con la existencia de un único maximal, entonces se da el resultado principal y su prueba. Por último, se incluyen algunas observaciones y comentarios.

3.1 Modelo perturbado usando el principio variacional de Ekeland

Sea $(X, A, \{A(x) \mid x \in X\}, Q, c)$ un MCM asociado a un PDM descontado fijo, el cual será denotado por M y lo llamaremos proceso original. En lo que resta de este capítulo, se supondrá que el proceso original M satisface la Condición 1.1.5 (a), (b) y (c). La Condición 1.1.5 (a), (b) y (c) para el proceso M no será mencionada en cada lema o teorema en este capítulo, pero se supone que se cumple. Más aún, dado un modelo M sea f^* la correspondiente política óptima cuya existencia está asegurada en el Lema 1.1.6, la cual no necesariamente es única.

Dado ε un número entero positivo el cual se supondrá que es fijo en el resto del capítulo. Ahora sea el siguiente PDM, denotado por M_ε (del proceso original M): $(X, A, \{A(x) \mid x \in X\}, Q, c^*)$, donde $c^*(x,a) = c(x,a) + \varepsilon\varphi(a, f^*(x))$, $x \in X$, $a \in A(x)$ y f^* como en el Lema 1.1.6, donde c es la función de costo original de M. Nótese que ambos PDMs, M y M_ε, son iguales excepto por la función de costo; más aún, el conjunto de $\mathbb{F}$ políticas estacionarias en el mismo para ambos modelos (en particular, el conjunto Π de todas las políticas admisibles es también el mismo para ambos modelos). M_ε le llamaremos proceso perturbado basado en el principio variacional de Ekeland.

Para M_ε, sea $W(\pi, x)$ el costo total descontado cuando se aplica la política π, dado el estado inicial x, y w^* denota su correspondiente función de valor óptimo.

Observación 3.1.1 *Observe que ahora el criterio de rendimiento para el proceso perturbado está definido como*

$$\begin{aligned} w^*(\pi, x) &= E_x^\pi\left[\sum_{t=0}^{\infty}\alpha^t c^*(x_t, a_t)\right] \\ &= E_x^\pi\left[\sum_{t=0}^{\infty}\alpha^t\left[c(x_t, a_t) + \varepsilon\varphi\left(a_t, f^*(x_t)\right)\right]\right] \\ &= E_x^\pi\left[\sum_{t=0}^{\infty}\alpha^t c(x_t, a_t)\right] + \varepsilon E_x^\pi\left[\sum_{t=0}^{\infty}\alpha^t\varphi\left(a_t, f^*(x_t)\right)\right] \\ &= v(\pi, x) + \varepsilon E_x^\pi\left[\sum_{t=0}^{\infty}\alpha^t\varphi\left(a_t, f^*(x_t)\right)\right], \end{aligned}$$

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

3.2 Existencia de una política óptima maximal para el modelo perturbado

Defínase para cada $x \in X$, una relación de orden en el conjunto de acciones admisibles $A(x)$, de la siguiente manera, para a_1, $a_2 \in A(x)$,

$$a_1 \prec a_2 \text{ si y sólo si } G(x, a_2) - G(x, a_1) + \varepsilon\varphi(a_1, a_2) \leq 0.$$

Esta relación claramente establece una relación de orden parcial, i.e., es una relación reflexiva, antisimétrica y transitiva.

Para $x \in X$, un control $d \in A(x)$ se le llama *maximal* para $A(x)$ si $a \prec d$ para todo $a \in A(x)$.

Lema 3.2.1 *Para cada $\eta \in \mathbb{F}$, existe una única política estacionaria $g^* \in \mathbb{F}$ tal que, para cada $x \in X$, $\eta(x) \prec g^*(x)$ y $g^*(x)$ es maximal para $A(x)$.*

Demostración. Sea $\eta \in \mathbb{F}$. Se definirá una sucesión de políticas estacionarias $\{f_k\}_{k=1,2,...}$ de la siguiente manera.
Tómese $f_1 \equiv \eta$. Defínase, para cada $x \in X$,

$$S_1(x) = \{a \in A(x) \,|f_1(x) \prec a\}, \tag{3.1}$$

en donde se supone que el conjunto $\{(x, a) \mid x \in X \text{ y } a \in S_1(x)\}$ es medible, es decir, la multifunción $x \twoheadrightarrow S_1(x)$ es medible.

Como $S_1(x) \neq \emptyset$, para cada $x \in X$ (observe que $f_1(x) \in S_1(x)$), f_2 se elegirá como se demuestra abajo. Por definición de $S_1(x)$, $x \in X$, se sigue que para cada $g \in S_1(x)$:

$$\begin{aligned} 0 \leq \varepsilon\varphi(g, f_1(x)) &\leq G(x, f_1(x)) - G(x, g) &(3.2)\\ &\leq G(x, f_1(x)) - \inf_{a \in S_1(x)} G(x, a). &(3.3) \end{aligned}$$

Considérese que $x \in X$ es arbitrario. Nótese que las Condiciones 1.1.5 (a) y 1.1.5 (b) implican que $G(x, \cdot)$ es i.s.c. Por tanto, como la función distancia es continua, implica que el conjunto $S_1(x)$ es cerrado. Más aún, es fácil probar que $S_1(x) \subseteq \{a \in A(x) \mid c(x, a) \leqslant G(x, f_1(x))\}$. De esta manera, $S_1(x)$ es un conjunto compacto (véase Condición 1.1.5 (a)).

Observe que como $G(x, \cdot)$ es i.s.c. en $S_1(x) \subseteq A(x)$ y $S_1(x)$ es compacto, para cada $x \in X$, por Proposición D.5, p. 182 en [25], se sigue que existe $f_2 \in \mathbb{F}$ tal que, para todo $x \in X$, $f_2(x) \in S_1(x)$, y

$$\inf_{a \in S_1(x)} G(x, a) = G(x, f_2(x)).$$

Ahora, tómese $\frac{1}{2}\varepsilon\varphi(f_1(x), f_2(x)) \geqslant 0$, para cada $x \in X$. Entonces, la siguiente desigualdad evidentemente se cumple para cada $x \in X$,

$$G(x, f_2(x)) \leq \inf_{a \in S_1(x)} G(x, a) + \frac{1}{2}\varepsilon\varphi(f_1(x), f_2(x)). \tag{3.4}$$

Por (3.3) y (3.4) se obtiene que, para cada $x \in X$,

$$\begin{aligned} G(x, f_2(x)) &\leq \inf_{a \in S_1(x)} G(x, a) + \frac{1}{2}\left[G(x, f_1(x)) - \inf_{a \in S_1(x)} G(x, a)\right]\\ &= \frac{1}{2}\left[G(x, f_1(x)) + \inf_{a \in S_1(x)} G(x, a)\right], \end{aligned}$$

por tanto, resulta que, para cada $x \in X$,

$$2G(x, f_2(x)) - G(x, f_1(x)) \ \ \leq \ \inf_{a \in S_1(x)} G(x, a). \tag{3.5}$$

Así, tomando para cada $x \in X$,

$$S_2(x) = \{a \in A(x) \, | f_2(x) \prec a\}. \tag{3.6}$$

Otra vez, considerando que $x \in X$ es arbitrario. Ahora, sea $f_k \in \mathbb{F}$ dado un entero positivo $k > 2$. Supóngase que para cada $x \in X$, $f_k(x) \in S_{k-1}(x)$ y $f_k(x)$ minimiza $G(x, a)$ sobre todos los $a \in S_{k-1}(x)$. Defínase $S_k(x)$ como:

$$S_k(x) = \{a \in A(x) \, | f_k(x) \prec a\}, \tag{3.7}$$

en donde se supone que el conjunto $\{(x, a) \mid x \in X \text{ y } a \in S_k(x)\}$ es medible, es decir, la multifunción $x \twoheadrightarrow S_k(x)$ es medible.

Entonces $f_{k+1} \in \mathbb{F}$, la cual satisface que $f_{k+1}(x) \in S_k(x)$ para $x \in X$ es elegida de manera similar como $f_2 \in \mathbb{F}$ se obtuvo de f_1. Entonces, las siguientes desigualdades también se cumplen para cada $x \in X$:

$$2G(x, f_{k+1}(x)) - G(x, f_k(x)) \ \ \leq \ \inf_{a \in S_k(x)} G(x, a) \tag{3.8}$$

y

$$S_{k+1}(x) = \{a \in A(x) \, | f_{k+1}(x) \prec a\}. \tag{3.9}$$

Con este procedimiento, la sucesión $\{f_k\}_{k=1,2,\ldots}$ es construida. Nótese que, para cada k y $x \in X$, $S_k(x)$ es compacto.

Ahora, una vez más considérese que $x \in X$ es arbitrario. Obsérvese que, para cada k, $f_{k+1}(x) \in S_k(x)$, satisface

$$G(x, f_{k+1}(x)) \leq \inf_{a \in S_k(x)} G(x, a) + \frac{1}{2}\left[G(x, f_k(x)) - \inf_{a \in S_k(x)} G(x, a)\right]. \tag{3.10}$$

En particular, la desigualdad anterior es consecuencia de (3.8). Ahora, por construcción, para cada $k \geq 1$, $S_{k+1}(x) \subset S_k(x)$ (recuerde que $\prec$ es transitiva), entonces

$$\inf_{a \in S_k(x)} G(x, a) \leq \inf_{a \in S_{k+1}(x)} G(x, a);$$

esto implica que

$$G(x, f_{k+1}(x)) - \inf_{a \in S_{k+1}(x)} G(x,a) \leq G(x, f_{k+1}(x)) - \inf_{a \in S_k(x)} G(x,a), \quad (3.11)$$

para cada $x \in X$.

De las desigualdades (3.10) y (3.11) (recuerde que $f_{k+1}(x) \in S_{k+1}(x)$, para cada $x \in X$), resulta que, para cada $x \in X$,

$$\begin{aligned} \left| G(x, f_{k+1}(x)) - \inf_{a \in S_{k+1}(x)} G(x,a) \right| &\leq \frac{1}{2} \left| G(x, f_k(x)) - \inf_{a \in S_k(x)} G(x,a) \right| \\ &\leq \frac{1}{2^k} \left| G(x, f_1(x)) - \inf_{a \in S_1(x)} G(x,a) \right|. \end{aligned}$$

De la desigualdad anterior, se obtiene que, para cada $x \in X$ y $l \in S_{k+1}(x)$ (obsérvese que $f_{k+1}(x) \prec l$):

$$\begin{aligned} |G(x, f_{k+1}(x)) - G(x,l)| &= G(x, f_{k+1}(x)) - G(x,l) \\ &\leq G(x, f_{k+1}(x)) - \inf_{a \in S_{k+1}(x)} G(x,a) \\ &\leq \frac{1}{2^k} \left| G(x, f_1(x)) - \inf_{a \in S_1(x)} G(x,a) \right|. \end{aligned}$$

De lo anterior y de la definición de $S_{k+1}(x)$, implica que, para cada $x \in X$,

$$\varepsilon\varphi(f_{k+1}(x), l) \leq \frac{1}{2^k} \left| G(x, f_1(x)) - \inf_{a \in S_1(x)} G(x,a) \right|.$$

De esta forma, usando la desigualdad triangular, se sigue que

$$\varepsilon\varphi(\beta, l) \leq \frac{2}{2^k} \left| G(x, f_1(x)) - \inf_{a \in S_1(x)} G(x,a) \right|,$$

para todo $x \in X$, $\beta, l \in S_{k+1}(x)$, and $k \geq 1$.

Por lo tanto, para cada $x \in X$, $\mathrm{Diam}(S_k(x)) \to 0$ (donde $\mathrm{Diam}(D)$ es el supremo de la distancias entre alguna pareja de elementos de D, para D un subconjunto de un espacio métrico), cuando $k \to \infty$; como, para cada $x \in X$, $S_1(x)$ es completo, los conjuntos $S_k(x)$ tienen para cada $k \geq 1$ un único punto en común, denotado por $g^*(x)$, que es,

$$\{g^*(x)\} = \bigcap_{k=1}^{\infty} S_k(x).$$

Ahora, por definición de $S_k(x) \subseteq A(x)$, se obtiene que $f_k(x) \prec g^*(x)$ para todo k y $x \in X$. En particular, para $k = 1$, se sigue que $f_1(x) \prec g^*(x)$. Supóngase que $g^*(x)$ no es maximal para $A(x)$. Entonces existe $z(x) \in A(x)$ tal que $g^*(x) \prec z(x)$. Por transitividad, $f_k(x) \prec z(x)$ para algún k. Consecuentemente, $z(x) \in \{g^*(x)\}$, i.e., $z(x) = g^*(x)$. Por lo tanto, $g^*(x)$ es el único elemento maximal para $A(x)$.

Como x es arbitrario, es posible definir una función $g^* : X \to A$ tal que $g^*(x) \in A(x)$ es maximal y única, para cada $x \in X$.

Ahora se establecerá que g^* es una política estacionaria. Para ello sea $x \in X$ fijo. Como $f_k(x) \prec f_{k+1}(x)$, entonces:

$$0 \le \varepsilon\varphi(f_k(x), f_{k+1}(x)) \le G(x, f_k(x)) - G(x, f_{k+1}(x)), \tag{3.12}$$

para $k = 1, 2, \ldots$. Sumando, para $m > k$:

$$\begin{aligned} \varepsilon\varphi(f_k(x), f_m(x)) &\le \varepsilon \sum_{j=k}^{m-1} \varphi(f_j(x), f_{j+1}(x)) && (3.13) \\ &\le \sum_{j=k}^{m-1} \{G(x, f_j(x)) - G(x, f_{j+1}(x))\} \\ &= G(x, f_k(x)) - G(x, f_m(x)). && (3.14) \end{aligned}$$

Por (3.12) se obtiene que $\{G(x, f_k(x))\}_{k=1,2,\ldots}$ es decreciente, y acotada por abajo, y entonces $\{G(x, f_k(x))\}_{k=1,2,\ldots}$ converge. Esto establece que el término de la parte derecha de la ecuación (3.14) converge a 0, cuando $k, m \to \infty$. Consecuentemente $\{f_k(x)\}_{k=1,2,\ldots}$ es una sucesión de Cauchy en $S_1(x) \subset A(x)$. El conjunto $S_1(x)$ es completo, y por lo tanto la sucesión de controles $\{f_k(x)\}_{k=1,2,\ldots}$ converge a $g^*(x)$ en $A(x)$. Por tanto, como x es arbitrario, resulta que $\{f_k(x)\}_{k=1,2,\ldots}$ converge a $g^*(x)$ en $A(x)$, para cada $x \in X$. Como $f_k \in \mathbb{F}$, que es, cada f_k es medible, resulta que g^* es también medible, i.e., $g^* \in \mathbb{F}$. ■

3.3 Unicidad de la política óptima en el proceso perturbado

Lema 3.3.1 *Para una política óptima f^*, dada por Lema 1.1.6, para el PDM M, existe $g^* \in \mathbb{F}$ (dada por el Lema 3.2.1 tomando $\eta \equiv f^*$), tal que, las siguientes propiedades se satisfacen:*

(i) Si $\varphi(f^*(x), g^*(x)) = 0$*, entonces* $f^*(x) = g^*(x)$ *y* $G(x, g^*(x)) = G(x, f^*(x))$*, para todo* $x \in X$*;*

(ii)

$$G(x, g^*(x)) < G(x, f(x)) + \varepsilon\varphi(f(x), g^*(x)), \tag{3.15}$$

para cada $f \in \mathbb{F}$ *con* $f(x) \neq g^*(x)$*, para cada* $x \in X$*. Más aún,* g^* *es la* **única** *política óptima para el proceso perturbado* M_ε*.*

Demostración. Tomando $\eta \equiv f^* \equiv f_1$ en el Lema 3.2.1, y sea g^* definida como en el Lema 3.2.1.

(i) Tomando $k = 1$ en (3.14), se obtiene para cada $m = 2, \ldots,$ y $x \in X$, la siguiente desigualdad

$$\varepsilon\varphi(f^*(x), f_m(x)) \leq G(x, f^*(x)) - G(x, f_m(x)). \tag{3.16}$$

Como $G(x, \cdot)$ es i.s.c. en $A(x)$ para cada $x \in X$, de (3.16), y como $f_m(x) \to g^*(x)$ para cada $x \in X$, se sigue que

$$\begin{aligned} G(x, g^*(x)) &\leq \liminf_{m\to\infty} G(x, f_m(x)) \\ &\leq \liminf_{m\to\infty} \{G(x, f^*(x)) - \varepsilon\varphi(f^*(x), f_m(x))\} \\ &= G(x, f^*(x)) - \varepsilon\varphi(f^*(x), g^*(x)), \end{aligned}$$

para cada $x \in X$. De esta manera,

$$G(x, g^*(x)) + \varepsilon\varphi(f^*(x), g^*(x)) \leq G(x, f^*(x))$$

y, como f^* es óptima para el proceso original M,

$$\begin{aligned} \varepsilon\varphi(f^*(x), g^*(x)) &\leq G(x, f^*(x)) - G(x, g^*(x)) \\ &\leq G(x, f^*(x)) - \inf_{a\in A(x)} G(x, a) = 0, \end{aligned}$$

para cada $x \in X$. Por tanto, $f^*(\cdot) = g^*(\cdot)$.

(ii) De (i) resulta que para cada $x \in X$, $W(g^*, x) = V(f^*, x) = v^*(x)$. Es fácil verificar que para cada $\pi \in \Pi$ y $x \in X$, $W(\pi, x) \geq V(\pi, x) \geq v^*(x)$ (obsérvese que $c^*(x, a) \geq c(x, a)$, para todo $x \in X$ y $a \in A(x)$).

Combinando estas igualdades y desigualdades, se sigue que, para cada $\pi \in \Pi$ y $x \in X$, $W(\pi, x) \geq W(g^*, x)$, por tanto $w^*(\cdot) = v^*(\cdot) = W(g^*, \cdot)$ y g^* es óptima para M_ε. Ahora, tomando $f \in \mathbb{F}$ y $x \in X$, tal que $f(x) \neq g^*(x)$, usando que $\varphi(g^*(x), f(x)) > 0$ y la maximalidad de g^*, resulta que

$$\begin{aligned} G(x, g^*(x)) &\leq G(x, f(x)) - \varepsilon\varphi(g^*(x), f(x)) \\ &\leq G(x, f(x)) \\ &< G(x, f(x)) + \varepsilon\varphi(g^*(x), f(x)), \end{aligned}$$

por tanto, la desigualdad (3.15) se sigue.

Finalmente, sea h una política estacionaria óptima para M_ε y $h \neq g^*$. Tomando $x \in X$, tal que $h(x) \neq g^*(x)$. De (3.15) y usando que $w^*(\cdot) = v^*(\cdot) = W(h, \cdot)$ y $f^*(\cdot) = g^*(\cdot)$, se obtiene que

$$\begin{aligned} w^*(x) &= G(x, g^*(x)) \\ &< G(x, h(x)) + \varepsilon\varphi(g^*(x), h(x)) \\ &= c(x, h(x)) + \varepsilon\varphi(f^*(x), h(x)) + \alpha \int_X v^*(y) Q(dy|x, h(x)) \\ &= c(x, h(x)) + \varepsilon\varphi(f^*(x), h(x)) + \alpha \int_X W(h, y) Q(dy|x, h(x)). \end{aligned}$$

De esta manera,

$$\begin{aligned} w^*(x) &< c^*(x, h(x)) + \alpha \int_X W(h, y) Q(dy|x, h(x)) \\ &= W(h, x), \end{aligned}$$

en donde la desigualdad anterior se obtiene de (4.2.15), p. 51 en [25]. Por lo tanto, $w^*(x) < W(h, x)$, i.e., h no es óptima, lo cual es una contradicción a la optimalidad de h. Esto finaliza la prueba de la parte (ii) del Lema 3.3.1.

■

Ahora, todo lo anterior en conjunto establece el resultado principal de este capítulo, que es claramente una consecuencia inmediata de los dos lemas previos:

Teorema 3.3.2 *Para un PDM M, sea f^* una política óptima (nótese que M no necesariamente tiene una política óptima única). Entonces existe un PDM M_ε con una única política óptima que coincide con f^*.*

El resultado principal que se obtuvo en este capítulo es la unicidad de la política óptima del modelo perturbado.

Las siguientes observaciones de este capítulo son importantes:

1. Con los resultados obtenidos en este capítulo se generalizó [38], debido a que se extendieron los resultados para funciones de costo no necesariamente convexas.

2. No hay una manera de obtener el resultado principal de este capítulo como un corolario del principio variacional de Ekeland, sin embargo, la prueba clásica (véase [21]) puede ser adaptada.

Además, el teorema principal de este capítulo, puede ser interpretado como una propiedad de densidad específica en el conjunto de modelos de decisión de Markov con una política óptima única. Debido a que la construcción del proceso perturbado usa explícitamente una política óptima del modelo original, este enfoque no ha sido aún establecido, si bajo alguna topología natural, el conjunto de todos los modelos de decisión de Markov con una política óptima es abierto o de primera categoría y por tanto es genérica.

Capítulo 4

PERTURBACIÓN BASADA EN LA REGULARIZACIÓN DE MOREAU-YOSIDA

En términos generales, la optimización trata de obtener máximos o mínimos de una función sobre un conjunto de restricciones y usar éstos para poder calcular el valor máximo o mínimo de la función objetivo. Un método muy desarrollado para encontrar los máximos o mínimos son los criterios de la primera y segunda derivada (véase [35], [53]). Sin embargo, no siempre es posible hacer uso de este tipo de herramienta debido a que existen muchos ejemplos de optimización en los cuales no es posible encontrar la derivada de la función objetivo, inclusive si se tienen condiciones las que garanticen la existencia de soluciones óptimas. Por ejemplo, la función valor absoluto tiene un mínimo en 0, sin embargo no es diferenciable en este punto. Otra de las grandes preocupaciones en optimización es también garantizar la unicidad del optimizador.

Ahora, dado un PDM asociado a un MCM, en este capítulo se tratará de encontrar un nuevo modelo perturbado para la familia de procesos de decisión de Markov descontados cuando el espacio de estados y conjunto de acciones son subconjuntos de $\mathbb{R}$, siempre que el sistema no presente diferenciabilidad de la política óptima y de la función de valor correspondiente, así como la no unicidad de la política óptima, de tal manera que el modelo perturbado presente diferenciabilidad en la función de valor óptimo y la política óptima y ésta resulte ser única. Además, también se está buscando que, tanto las funciones de valor óptimo como las políticas óptimas sean las mismas. Como una consecuencia de la diferenciabilidad en el modelo perturbado se quiere encontrar una tasa de convergencia para las políticas provenientes del método de iteración de valores.

Anteriormente se obtuvieron resultados para el criterio de recompensa total descontado en los cuales se implementan las técnicas de diferenciabilidad para obtener una política óptima (véase [19]), sin embargo estos resultados sólo se pueden implementar cuando tenemos diferenciabilidad en el lado derecho de la ecuación de programación dinámica. En esta tesis se obtiene un nuevo resultado sin la restricción de diferenciabilidad en la función de costo y adicionalmente se garantiza la existencia de una política óptima única. En cuanto a la convergencia se tienen otros resultados ([16] y [17]) los cuales presentan condiciones bajo las cuales es posible garantizar que las políticas provenientes del método de iteración de valores convergen a la política óptima; cabe mencionar que la unicidad de la política óptima es fundamental en estos trabajos y no se dan cotas de convergencia.

En este capítulo se usan las técnicas de regularización de Moreau-Yosida (véase Capítulo de Preliminares) la cual consiste en perturbar el modelo original de tal manera que el modelo perturbado sea diferenciable. Cabe mencionar que este tipo de regularización hasta el momento no se ha aplicado a PDMs. Usando este tipo de regularización no sólo se obtiene la unicidad y la diferenciabilidad de la política óptima y de la función de valor óptimo del modelo perturbado, sino que además se obtiene que la función de valor para el proceso perturbado es finito valuada y estrictamente convexa, sin tener condiciones de convexidad estricta. Aún más, como consecuencia de este tipo de perturbación se proporciona de manera explícita una cota de convergencia para las políticas de iteración de valores.

Este capítulo está organizado de la siguiente manera: en primer lugar se presenta la manera de perturbar el PDM basado en la regularización de Moreau-Yosida. En segundo lugar se prueba la existencia y unicidad de la política óptima del modelo perturbado; posteriormente se prueba la convexidad del modelo perturbado; en cuarto lugar se demuestra la diferenciabilidad del modelo perturbado y por último se dan las cotas de convergencia de las políticas de iteración de valores.

4.1 Regularización de Moreau-Yosida

Sean $X \subseteq \mathbb{R}$, $A \subseteq \mathbb{R}$ conjuntos de Borel no vacíos y suponga que $A(x) \subset A$, para toda $x \in X$. Denotemos por $\mathbb{K} := \{(x,a) : x \in X, a \in A(x)\}$. Defínase $G : \mathbb{K} \to \mathbb{R}^+$ una función medible y

$$\psi(x) = \inf_{a \in A(x)} G(x,a),$$

$x \in X$.

Definición 4.1.1 *Defínase para cada $(x,a) \in \mathbb{K}$ y $\lambda > 0$ fijo*

$$H(x,a) := \inf_{b \in A(x)} \left\{ G(x,b) + \frac{\lambda}{2}(b-a)^2 \right\},$$

$$P(x,a) := \arg\inf_{b \in A(x)} \left\{ G(x,b) + \frac{\lambda}{2}(b-a)^2 \right\},$$

a H y P se le llama la regularización de Moreau-Yosida y el operador proximal, respectivamente.

Ejemplo 4.1.2 *Sean $X = A = A(x) = \mathbb{R}$, para cada $x \in X$. Defina $G(x,a) = |x| + |a|$ para cada $(x,a) \in \mathbb{K}$. Entonces*

$$\begin{aligned} H(x,a) &= \min_{b \in A(x)} \left\{ G(x,b) + \frac{\lambda}{2}(b-a)^2 \right\} \\ &= \min_{b \in A(x)} \left\{ |x| + |b| + \frac{\lambda}{2}(b-a)^2 \right\} \\ &= |x| + \min_{b \in A(x)} \left\{ |b| + \frac{\lambda}{2}(b-a)^2 \right\}. \end{aligned}$$

Análogamente, como en el Ejemplo 1.3.12,

$$H(x,a) = \begin{cases} |x| + \frac{\lambda}{2}a^2 & si \quad |a| \leq 1/\lambda \\ |x| + |a| - \frac{1}{2\lambda} & si \quad |a| > 1/\lambda \end{cases}$$

y

$$P(x,a) = \begin{cases} 0 & si \quad |a| \leq 1/\lambda \\ a - \frac{1}{\lambda} & si \quad a > 1/\lambda \\ \frac{1}{\lambda} + a & si \quad a < -1/\lambda, \end{cases}$$

para cada $x \in X$ y $a \in A(x)$.

Condición 4.1.3 *Para cada $x \in X$, se cumple que $G(x, \cdot)$ es una función convexa.*

Observación 4.1.4 *a) Nótese que el operador proximal $P(x, a)$ está bien definido, para cada $(x, a) \in \mathbb{K}$, ya que la función a minimizar es estrictamente convexa bajo la Condición 4.1.3 (véase Teorema 4.1.7 abajo).*

b) Nótese que para todo $(x, a) \in \mathbb{K}$, siempre se cumple que $H(x, a) \leq G(x, a)$. En efecto, sea $(x, a) \in \mathbb{K}$ entonces

$$\begin{aligned} H(x,a) &= \inf_{b \in A(x)} \left\{ G(x,b) + \frac{\lambda}{2} (b-a)^2 \right\} \\ &\leq G(x,a) + \frac{\lambda}{2} (a-a)^2 \\ &= G(x,a). \end{aligned}$$

Lema 4.1.5 *Sea $a \in A(x)$, fijo. Defínase $L(b) = (b-a)^2$, para cada $b \in A(x)$. Entonces L es una función estrictamente convexa en $A(x)$.*

Demostración. Sea $a \in A(x)$, fijo. Entonces $L'(b) = 2(b-a)$, más aún $L''(b) = 2 > 0$. Por lo tanto, L es estrictamente convexa. ■

Notación 4.1.6 *Para cada $(x, a) \in \mathbb{K}$ fijo, se usará la siguiente notación*

$$\overline{H}(x,a,b) := G(x,b) + \frac{\lambda}{2} (b-a)^2 ,$$

para todo $b \in A(x)$.

Lema 4.1.7 *Supóngase que la Condición 4.1.3 se cumple. Entonces $\overline{H}(x, a, \cdot)$ es una función estrictamente convexa, para cada $(x, a) \in \mathbb{K}$. Además, $P(x, a)$ tiene un sólo elemento, para cada $(x, a) \in \mathbb{K}$.*

Demostración. Sea $(x, a) \in \mathbb{K}$ fijo. Usando la Condición 4.1.3 y el Lema 4.1.5, $\overline{H}(x, a, \cdot)$ es suma de una función estrictamente convexa y una función convexa. De esto se obtiene que $\overline{H}(x, a, \cdot)$ es estrictamente convexa. ■

Lema 4.1.8 *Supóngase que para todo $x \in X$, existe $a \in A(x)$ tal que $G(x, a) < \infty$. Entonces, para toda $(x, a') \in \mathbb{K}$ se cumple que $H(x, a') < \infty$.*

Demostración. Sea $x \in X$, fijo. Por hipótesis, existe $a \in A(x)$ tal que $G(x,a) < \infty$. Entonces,

$$\begin{aligned} H(x,a') &\leq G(x,a) + \frac{\lambda}{2}(a-a')^2 \\ &< \infty, \end{aligned}$$

para toda $a' \in A(x)$. ■

Lema 4.1.9 *Suponga que la Condición 4.1.3 se cumple. Si $f^* : X \to A$, $f^*(x) \in A(x)$, para cada $x \in X$ es la única función tal que $\psi(x) = G(x, f^*(x))$, para cada $x \in X$, entonces $P(x, f^*(x)) = \{f^*(x)\}$, para cada $x \in X$, es decir, f^* es el único ínfimo para $H(x, f^*(x))$. Además, $G(x, f^*(x)) = H(x, f^*(x))$ para cada $x \in X$.*

Demostración. Usando el Lema 4.1.7 se tiene que $\overline{H}(x,a,\cdot)$ es estrictamente convexa para cada $(x,a) \in \mathbb{K}$, entonces usando el Lema 1.3.3, $\overline{H}(x, f^*(x), \cdot)$ tiene un único ínfimo b_x en $A(x)$. Nótese que

$$\begin{aligned} G(x,f^*(x)) &\leq G(x,b_x) \\ &\leq G(x,b_x) + \frac{\lambda}{2}\left(b_x - f^*(x)\right)^2 \\ &\leq \min_{b \in A(x)} \left\{ G(x,b) + \frac{\lambda}{2}\left(b - f^*(x)\right)^2 \right\} \\ &= H(x, f^*(x)) \\ &\leq G(x,f^*(x)) + \frac{\lambda}{2}\left(f^*(x) - f^*(x)\right)^2 \\ &\leq G(x,f^*(x)), \end{aligned} \tag{4.1}$$

lo cual implica que $P(x, f^*(x)) = \{b_x\} = \{f^*(x)\}$ y que $G(x, f^*(x)) = H(x, f^*(x))$. ■

Definición 4.1.10 *Una correspondencia entre X y A es una aplicación denotada por Γ, que a cada $x \in X$ le asocia un subconjunto $\Gamma(x) = A(x)$ de A.*

Definición 4.1.11 *(a) Una correspondencia $\Gamma : X \twoheadrightarrow A$ es hemicontinua superior (superiormente semicontinua) en $x_0 \in X$, si para cada conjunto abierto B de A tal que $\Gamma(x_0) \subset B$, la preimagen superior de B, $\Gamma^+[B] := \{x \in X : \Gamma(x) \subset B\}$ contiene una vecindad de x_0. Una correspondencia $\Gamma : X \twoheadrightarrow A$ es hemicontinua superior en X si es hemicontinua superior en cada $x_0 \in X$.*

(b) *Una multifunción* $\Gamma : X \twoheadrightarrow A$ *es hemicontinua inferior (inferiormente semicontinua) en* $x_0 \in X$, *si para cada conjunto abierto* B *de* A *tal que* $\Gamma(x_0) \cap B \neq \emptyset$, *la preimagen inferior de* B, $\Gamma^-[B] := \{x \in X : \Gamma(x) \cap B \neq \emptyset\}$ *contiene una vecindad de* x_0. *Una correspondencia* $\Gamma : X \twoheadrightarrow A$ *es hemicontinua inferior en* X *si es hemicontinua inferior en cada* $x_0 \in X$.

(c) *Una correspondencia* $\Gamma : X \twoheadrightarrow A$ *es continua en* $x_0 \in X$ *si es hemicontinua superior e inferior en* $x_0 \in X$. *Una correspondencia* $\Gamma : X \twoheadrightarrow A$ *es continua en* X *si es hemicontinua superior e inferior en cada* $x_0 \in X$.

Una definición equivalente a la anterior es la siguiente

Definición 4.1.12 *Una multifunción* $\Gamma : X \twoheadrightarrow A$ *es*

(a) *superiormente semicontinua (s.s.c.) si el conjunto* $\{x \in X \mid \Gamma(x) \cap F \neq \emptyset\}$ *es cerrado en* X, *para todo conjunto cerrado* $F \subset X$.

(b) *inferiormente semicontinua (i.s.c.) si el conjunto* $\{x \in X \mid \Gamma(x) \cap G \neq \emptyset\}$ *es abierto en* X, *para todo conjunto abierto* $G \subset X$.

(c) *continua si es inferiormente semicontinua y superiormente semicontinua.*

El siguiente resultado es una caracterización de la semicontinuidad inferior, para más detalles véase Proposición D.2 en [25], pág. 182.

Lema 4.1.13 *Sea* $\Gamma : X \twoheadrightarrow A$ *una multifunción, las siguientes afirmaciones son equivalentes*

(a) Γ *es i.s.c.*

(b) *el conjunto* $\{x \in X \mid \Gamma(x) \subset F\}$ *es cerrado en* X, *para todo conjunto cerrado* $F \subset A$.

(c) *si* $x_n \to x$ *en* X *y* $a \in \Gamma(x)$, *entonces existe una sucesión* $a_n \in \Gamma(x_n)$ *tal que* $a_n \to a$.

Un resultado similar es para la caracterización de la semicontinuidad superior, para más detalles véase Lema 2.20 en [17], pág. 5.

Lema 4.1.14 *Sea $\Gamma : X \twoheadrightarrow A$ una multifunción. Supóngase que $\Gamma(x) \neq \emptyset$ para todo $x \in X$ y más aún Γ es compacto valuada. Entonces, las siguientes afirmaciones son equivalentes*

(a) Γ *es s.s.c.*

(b) *si $x_n \to x$, $n \to +\infty$ en X y $a_n \in \Gamma(x_n)$, $n = 1, 2, \cdots$, entonces existe una subsucesión $\{a_{n_k}\}$ de $\{a_n\}$ tal que $a_{n_k} \to a$, $n \to +\infty$.*

Lema 4.1.15 *Supóngase que se cumple la Condición 4.1.3, que G es continua, que los conjuntos $A(x)$ son compactos y que la multifunción $x \twoheadrightarrow A(x)$ es continua, para cada $x \in X$. Entonces H es continua y $P(x, a)$ es no vacío, compacto y la multifunción $a \twoheadrightarrow P(x, a)$ es continua, para cada $x \in X$.*

Demostración. Sea $x \in X$, fijo. Observe que $\overline{H}(x, a, \cdot)$ es continua, ya que es suma de dos funciones continuas. Usando el Lema 4.1.7 $P(x, a)$ es único, para cada $a \in A(x)$. Por lo tanto, usando el teorema del máximo de Berge (véase [24]) se obtiene el resultado. ∎

Nota 4.1.16 *Dado que f^* es única para H (véase Lema 4.1.9), cuando $P(x, f^*(x)) = \{f^*(x)\}$, $x \in X$ usaremos la notación $P(x, f^*(x)) = f^*(x)$, para todo $x \in X$.*

En lo que resta del capítulo se usarán las siguientes notaciones para las derivadas parciales, $H_a(x, a)$, $H_{aa}(x, a)$ denotarán la primera, segunda derivada parcial con respecto a a, respectivamente.

Lema 4.1.17 *Supóngase que se cumple la Condición 4.1.3. Entonces, para cada $x \in X$, existe la derivada parcial $H_a(x, \cdot)$ en el int$(A(x))$,*

$$H_a(x, a) = \lambda(a - P(x, a)).$$

Más aún, para cada $x \in X$

$$|H_a(x, a) - H_a(x, a')| \leq \lambda \mid a - a' \mid . \tag{4.2}$$

para toda $a, a' \in A(x)$.

Demostración. Sean $x \in X$ y $a \in int(A(x))$ fijos y $t > 0$. Por definición

$$
\begin{aligned}
\frac{H(x,a+t)-H(x,a)}{t} &= \frac{\text{mín}_{b\in A(x)}\left\{G(x,b)+\frac{\lambda}{2}(b-a-t)^2\right\}}{t} \\
&- \frac{\text{mín}_{\omega\in A(x)}\left\{G(x,\omega)+\frac{\lambda}{2}(\omega-a)^2\right\}}{t} \\
&\geq \frac{G(x,P(x,a+t)+\dfrac{\lambda}{2}(P(x,a+t)-a-t)^2}{t} \\
&- \frac{G(x,P(x,a+t)+\dfrac{\lambda}{2}(P(x,a+t)-a)^2}{t} \\
&= \frac{\lambda}{2}\frac{(P(x,a+t)-a-t)^2-(P(x,a+t)-a)^2}{t} \\
&= \frac{\lambda}{2}\frac{(P(x,a+t)-P(x,a)+P(x,a)-a-t)^2}{t} \\
&- \frac{\lambda}{2}\frac{(P(x,a+t)-P(x,a)+P(x,a)-a)^2}{t} \\
&= \frac{\lambda}{2}\frac{(P(x,a)-a-t)^2}{t} \quad \text{(véase nota abajo)} \qquad (4.3) \\
&- \frac{\lambda}{2}\frac{(P(x,a)-a)^2}{t} \\
&- \lambda\left(P(x,a+t)-P(x,a)\right),
\end{aligned}
$$

en donde se usó que $H(x,a) \leq G(x,P(x,a+t)) + \dfrac{\lambda}{2}(P(x,a+t)-a)^2$. En este caso, para probar la igualdad (4.3) sólo basta sustituir tomando $\alpha = P(x,a+t)-P(x,a)$, $\beta = P(x,a)-a-t$ y $\gamma = P(x,a)-a$.

Ahora, dado que $P(x,\cdot)$ es continua por el Lema 4.1.15 tomando el límite cuando $t \to 0$ se tiene que $P(x,a+t) \to P(x,a)$ de donde se sigue que

$$
\lim_{t\to 0}\frac{H(x,a+t)-H(x,a)}{t} \geq \lambda(a-P(x,a)).
$$

Por otro lado, como $H(x,a+t) \leq G(x,P(x,a)) + \dfrac{\lambda}{2}\left(P(x,a)-a-t\right)^2$, se tiene que

$$\begin{array}{rcl}
\dfrac{H(x,a+t)-H(x,a)}{t} & = & \dfrac{\text{mín}_{b\in A(x)}\left\{G(x,b)+\frac{\lambda}{2}\left(b-a-t\right)^2\right\}}{t} \\
& - & \dfrac{\text{mín}_{\omega\in A(x)}\left\{G(x,\omega)+\frac{\lambda}{2}\left(\omega-a\right)^2\right\}}{t} \\
& \leq & \dfrac{G(x,P(x,a))+\dfrac{\lambda}{2}\left(P(x,a)-a-t\right)^2}{t} \\
& - & \dfrac{G(x,P(x,a))+\dfrac{\lambda}{2}\left(P(x,a)-a\right)^2}{t} \\
& = & \dfrac{\lambda}{2}\dfrac{\left(P(x,a)-a-t\right)^2-\left(P(x,a)-a\right)^2}{t}.
\end{array}$$

Una vez más, tomando el límite cuando $t \to 0$ se obtiene que

$$\lim_{t\to 0}\frac{H(x,a+t)-H(x,a)}{t} \leq \lambda(a-P(x,a)).$$

Por lo tanto, $H_a(x,a) = \lambda(a-P(x,a))$. Usando lo anterior se tiene que

$$\begin{array}{rcl}
|H_a(x,a)-H_a(x,a')| & = & \lambda|a-a'+P(x,a')-P(x,a)| \\
& \leq & \lambda|a-a'|+\lambda|P(x,a')-P(x,a)|.
\end{array}$$

Ahora, se probará que el segundo término de la igualdad anterior es negativo. Para ésto se usará la monotonicidad de la subdiferencial $\partial G(x,\cdot)$. Para ello sean $a, a' \in A(x)$, $s \in \partial G(x,a)$ y $s' \in \partial G(x,a')$, entonces $(s-s')(a-a') \geq 0$. Para probar ésto, considérense las desigualdades asociadas a los subgradientes con s y s' aplicado a a' y a respectivamente.

$$G(x,a') \geq G(x,a) + s(a'-a)$$

y

$$G(x,a) \geq G(x,a') + s(a'-a).$$

Sumando las dos desigualdades anteriores, se obtiene que

$$(s-s')(a-a') \geq 0.$$

Ahora, nótese que por la optimalidad de $P(x,a)$ y $P(x,a')$ se tiene que

$$0 \in \partial \left(G(x, P(x,a) + \frac{\lambda}{2} (P(x,a) - a)^2 \right)$$

y

$$0 \in \partial \left(G(x, P(x,a') + \frac{\lambda}{2} (P(x,a') - a')^2 \right).$$

Así,

$$H_a(x,a) = \lambda(a - P(x,a)) \in \partial G(x, P(x,a))$$

y

$$H_a(x,a') = \lambda(a' - P(x,a')) \in \partial G(x, P(x,a')).$$

Usando la monotonicidad de $\partial G(x,\cdot)$ se tiene que

$$P(x,a') - P(x,a) \leq 0.$$

Por lo tanto,

$$|H_a(x,a) - H_a(x,a')| \leq \lambda |a - a'|.$$

■

Lema 4.1.18 *Supóngase que se cumple la Condición 4.1.3 y que $P_a(x,a)$ existe. Entonces la segunda derivada parcial $H_{aa}(x,\cdot)$, para cada $x \in X$ y $a \in int(A(x))$, es*

$$H_{aa}(x,a) = \lambda(1 - P_a(x,a)).$$

Demostración. Sea $x \in X$ fijo. Por el Lema 4.1.17 se llega a que

$$H_a(x,a) = \lambda(a - P(x,a)). \tag{4.4}$$

Ahora, usando que $P(x,a)$ es diferenciable con respecto a a y derivando nuevamente la ecuación (4.4) se concluye que

$$H_{aa}(x,a) = \lambda(1 - P_a(x,a)).$$

■

4.2 Aplicación a procesos de decisión de Markov

4.2.1 Modelo perturbado usando la regularización de Moreau-Yosida

Considérese un proceso de decisión de Markov descontado fijo, asociado a un modelo de control de Markov $(X, A, \{A(x)| \, x \in X\}, Q, c)$, al cual le llamaremos proceso original y será denotado por M. Supóngase que X y A son ambos subconjuntos de Borel de $\mathbb{R}$. Es importante mencionar que los resultados que se obtienen en $\mathbb{R}$ se pueden extender de manera análoga a $\mathbb{R}^n$.

En lo que resta de este capítulo, se supondrá que el proceso original M satisface la Condición 1.1.5 (a), (b) y (c). La Condición 1.1.5 (a), (b) y (c) para el proceso M no será mencionada en cada lema o teorema en este capítulo, pero se supone que se cumple. Más aún, dado un modelo M sea f^* la correspondiente política óptima cuya existencia está asegurada en el Lema 1.1.6, la cual no necesariamente es única.

Defínase para $\lambda > 0$ fijo, el siguiente PDM denotado por M_λ con el MCM $(X, A, \{A(x)|x \in X\}, Q, \widehat{c})$, donde $\widehat{c}(x,a) = c(x,a) + \dfrac{\lambda}{2}\left(a - f^*(x)\right)^2$, $x \in X$, $a \in A(x)$ y f^* es una política óptima del modelo original M, fija y c es la función de costo para M. Obsérvese que ambos PDMs M y M_λ coinciden en los componentes del modelo de control de Markov excepto en la función de costo; más aún, el conjunto $\mathbb{F}$ de las políticas estacionarias es el mismo para ambos modelos. A M_λ le llamaremos proceso perturbado basado en la regularización de Moreau-Yosida. En lo que resta del capítulo λ se supondrá $\lambda > 0$ y fijo.

Para M_λ, sea $W(\pi, x)$ la esperanza del costo total descontado cuando la política π es aplicada, dado el estado inicial x, y ω^* denota la correspondiente función de valor óptimo.

Notación 4.2.1 *Recuerde que en PDMs, para cada $(x,a) \in \mathbb{K}$, G y G_n denotan lo siguiente*

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

y

$$G_n(x,a) := c(x,a) + \alpha \int_X V_{n-1}(y) Q(dy|x,a).$$

Definición 4.2.2 *Defínase para f^* del modelo original M y $\lambda > 0$ fijo*

$$H(x, f^*(x)) = \min_{b \in A(x)} \left\{ G(x,b) + \frac{\lambda}{2} \left(b - f^*(x)\right)^2 \right\},$$

$$P(x, f^*(x)) = \arg\min_{b \in A(x)} \left\{ G(x,b) + \frac{\lambda}{2} \left(b - f^*(x)\right)^2 \right\}.$$

Además,

$$H_n(x, f^*(x)) := \min_{b \in A(x)} \left\{ G_n(x,b) + \frac{\lambda}{2} \left(b - f^*(x)\right)^2 \right\}.$$

4.2.2 Convexidad en el modelo perturbado

En esta sección se probará que el modelo perturbado M_λ, bajo condiciones de convexidad se vuelve estrictamente convexo.

Condición 4.2.3

1. *X y A son convexos;*
2. *$(1-\beta)a + a' \in A((1-\beta)x + x')$ para todo $x \in X$, $a \in A(x)$, $a' \in A(x')$ y $\beta \in [0,1]$. Además, también se supone que: si $x, y \in X$, $x < y$, entonces $A(y) \subseteq A(x)$, y $A(x)$ son convexos para cada $x \in X$;*
3. *Q es inducida por la ecuación en diferencias $x_{t+1} = F(x_t, a_t, \xi_t)$, con $t = 0, 1, \ldots$, donde $F : X \times A \times S \to X$ es una función medible y $\{\xi_t\}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas (i.i.d) con valores en $S \subseteq \mathbb{R}$ y con densidad común Δ. En resumen, se supone que $F(\cdot, \cdot, s)$ es una función convexa en $\mathbb{K}$, para cada $s \in S$; y si $x, y \in X$, $x < y$, entonces $F(x, a, s) \leq F(y, a, s)$ para cada $a \in A(x)$ y $s \in S$;*
4. *c es estrictamente convexo en $\mathbb{K}$, y si $x, y \in X$, $x < y$, entonces $c(x, a) \leq c(y, a)$ para cada $a \in A(x)$.*

Condición 4.2.4

1. *Lo mismo que en la Condición 4.2.3 (1);*
2. *$(1-\beta)a + a' \in A((1-\beta)x + x')$ para todo $x \in X$, $a \in A(x)$, $a' \in A(x')$ y $\beta \in [0,1]$. Además, también se supone que $A(x)$ es convexo para cada $x \in X$;*

3. *Q es dada por la relación $x_{t+1} = \gamma x_t + \delta a_t + \xi_t$, $t = 0, 1, \ldots$, en donde $\{\xi_t\}$ son variables aleatorias i.i.d con valores en $S \subseteq \mathbb{R}$ y con densidad Δ, γ y δ son números reales.*

4. *c es estrictamente convexo en $\mathbb{K}$.*

Lema 4.2.5 *Supóngase que una de las Condiciones 4.2.3 ó 4.2.4 se cumplen. Entonces $G(x, \cdot)$ es estrictamente convexa, para cada $x \in X$. Además, la política óptima f^* es única.*

Demostración. Véase [16]. ∎

Notación 4.2.6 *Para cada $(x, a) \in \mathbb{K}$ fijo, se usará la siguiente notación*

$$\overline{H}(x, f^*(x), b) := G(x, b) + \frac{\lambda}{2} (b - f^*(x))^2,$$

para todo $b \in A(x)$.

Teorema 4.2.7 *Supóngase que una de las Condiciones 4.2.3 ó 4.2.4 se cumplen. Entonces $\overline{H}(x, f^*(x), \cdot)$ es una función estrictamente convexa, para cada $x \in X$.*

Demostración. Sea $(x, a) \in \mathbb{K}$ fijo. Usando el Lema 4.1.7. Se obtiene que $\overline{H}(x, a, \cdot)$ es estrictamente convexa. ∎

4.2.3 Existencia de una política óptima única del modelo perturbado

Lema 4.2.8 *Supóngase que la Condición 1.1.5 (c) se cumple. Entonces, existe $\pi \in \Pi$ tal que $W(\pi, x) < \infty$, para cada $x \in X$.*

Demostración. Sea $x \in X$ fijo. Por la Condición 1.1.5 (c), existe $\pi \in \Pi$ tal que $V(\pi, x) < \infty$. Entonces, usando $\pi = f^*$

$$\begin{aligned}
W(\pi, x) &= E_x^\pi \left[\sum_{t=0}^{\infty} \alpha^t \widehat{c}(x_t, a_t) \right] \\
&= E_x^\pi \left[\sum_{t=0}^{\infty} \alpha^t \left(c(x_t, a_t) + \frac{\lambda}{2} (a_t - f^*(x_t))^2 \right) \right] \\
&= E_x^\pi \left[\sum_{t=0}^{\infty} \alpha^t c(x_t, a_t) \right] \\
&= v^*(x) < \infty,
\end{aligned}$$

en donde f^* es la política óptima del modelo original M y está dada como en el Lema 1.1.6. ■

Teorema 4.2.9 *Sea f^* la política óptima del modelo original M como en el Lema 1.1.6, entonces $P(x, f^*(x)) = \{f^*(x)\}$, para cada $x \in X$, es decir, f^* es la única política óptima para el modelo perturbado M_λ. Además, $G(x, f^*(x)) = H(x, f^*(x))$ para cada $x \in X$.*

Demostración. Usando el Lema 4.1.9 se tiene que $P(x, f^*(x)) = \{f^*(x)\}$ y que $G(x, f^*(x)) = H(x, f^*(x))$. Sólo resta probar que f^* es óptima para el proceso perturbado M_λ. De lo anterior se tiene que para cada $x \in X$, $W(f^*, x) = V(f^*, x) = v^*(x)$. Es fácil verificar que para cada $\pi \in \Pi$ y $x \in X$, $W(\pi, x) \geq V(\pi, x) \geq v^*(x)$. Combinando estas igualdades y desigualdades se sigue que, para cada $\pi \in \Pi$ y $x \in X$, $W(\pi, x) \geq W(f^*, x)$, por tanto $w^*(\cdot) = v^*(\cdot) = W(g^*, \cdot)$ y f^* es óptima para M_λ. Sea h una política estacionaria tal que h es óptima para M_λ y $h \neq f^*$. Tomemos $x \in X$ tal que $h(x) \neq f^*(x)$. De (4.1) y usando que $w^*(\cdot) = v^*(\cdot)$ se obtiene que

$$
\begin{aligned}
w^*(x) &= G(x, f^*(x)) \\
&< G(x, h(x)) + \frac{\lambda}{2} (f^*(x) - h(x))^2 \\
&= c(x, h(x)) + \frac{\lambda}{2} (f^*(x) - h(x))^2 + \alpha \int_X v^*(y) Q(dy|x, h(x)) \\
&= c(x, h(x)) + \frac{\lambda}{2} (f^*(x) - h(x))^2 + \alpha \int_X w^*(y) Q(dy|x, h(x)).
\end{aligned}
$$

Así,

$$
w^*(x) < c^*(x, h(x)) + \alpha \int_X w^*(y) Q(dy|x, h(x)),
$$

i.e., $h(x)$ no es óptima, lo cual es una contradicción para la optimalidad de h. ■

Dado que la política óptima f^* del modelo original M es una política óptima única para el proceso perturbado M_λ (véase Lema 4.2.9), cuando $P(x, f^*(x)) = \{f^*(x)\}$, $x \in X$ usaremos la notación $P(x, f^*(x)) = f^*(x)$, para todo $x \in X$.

4.2.4 Diferenciabilidad en el modelo perturbado

Condición 4.2.10 *1. Supóngase que para cada $x \in X$, los conjuntos $A(x)$ son compactos y la multifunción $x \twoheadrightarrow A(x)$ es continua.*

2. $c(\cdot, \cdot)$ es continuo y $\int \upsilon^*(y)Q(dy \mid \cdot, \cdot)$ es una función continua.

Lema 4.2.11 *Supóngase que se cumple la Condición 4.2.10. Entonces en el modelo perturbado M_λ se cumple que υ^* y $P(x, f^*(x)) = f^*(x)$ son continuas, para cada $x \in X$.*

Demostración. Véase [37]. ■

Teorema 4.2.12 *Supóngase que se cumple la Condición 4.2.3 ó la Condición 4.2.4 y la Condición 4.2.10. Entonces si $f^*(x) \in int(A(x))$, se tiene que*

$$H_a(x, f^*(x)) = \lambda(f^*(x) - P(x, f^*(x))) = 0.$$

Demostración. Sea $x \in X$ fijo. Del Lema 4.1.17 se tiene que para cada $a \in int(A(x))$

$$H_a(x, a) = \lambda(a - P(x, a)).$$

Entonces, evaluando en $f^*(x)$ y usando el Teorema 4.2.9 se llega a que

$$H_a(x, f^*(x)) = \lambda(f^*(x) - P(x, f^*(x))) = \lambda(f^*(x) - f^*(x)) = 0.$$

■

4.2.5 Cotas de convergencia para las políticas provenientes del método de iteración de valores

En esta sección se dará de manera explícita una cota de convergencia para las políticas provenientes del método de iteración de valores, como una consecuencia de la perturbación basada en la regularización de Moreau-Yosida cuando los conjuntos de acciones admisibles son compactos.

Condición 4.2.13 *Supóngase que existe una constante $\hat{k}$, $1 < \delta < \frac{1}{\alpha}$ y una función medible $\varpi : X \to [1, \infty)$ tales que*

$$\sup_A \mid c(x, a) \mid \leq \hat{k}\varpi(x), \text{ y}$$

$$\sup_A \int \varpi(y)Q(dy|x,a) \leq \delta\varpi(x),$$

para todo $x \in X$, donde ϖ es una función de peso $\varpi \geq 0$ en X.

Definición 4.2.14 *Sea $\varpi : X \to \mathbb{R}^+$ una función medible, se define*

$$\| u \|_\varpi = \left|\left|\frac{u}{\varpi}\right|\right| = \frac{\sup_{x\in X} \mid u(x) \mid}{\varpi(x)} \tag{4.5}$$

Para una demostración del siguiente lema véase [26], Teorema 8.3.6.

Lema 4.2.15 *Bajo las Condiciones 4.2.10 y 4.2.13, se tiene que existe $N \in \mathbb{N}$ tal que, para todo $n \geq N$ se satisface que*

$$\| V_n - \upsilon^* \|_\varpi \leq \frac{\hat{k}(\alpha\delta)^n}{1-\alpha\delta}, \tag{4.6}$$

para todo $a \in A(x)$ y $x \in X$, donde $\hat{k}$, δ, $\varpi(x)$ son como en la Condición 4.2.13 y $f_n(x)$ son las políticas provenientes del método de iteración de valores.

Lema 4.2.16 *Bajo las Condiciones 4.2.10 y 4.2.13, se tiene que existe $N \in \mathbb{N}$ tal que, para todo $n \geq N$ se satisface que*

$$|V_n(x) - \upsilon^*(x)| \leq \frac{\hat{k}\varpi(x)(\alpha\delta)^n}{1-\alpha\delta}, \tag{4.7}$$

para todo $a \in A(x)$ y $x \in X$, donde $\hat{k}$, δ, $\varpi(x)$ son como en la Condición 4.2.13 y $f_n(x)$ son las políticas provenientes del método de iteración de valores.

Demostración. Por definición y por el Lema 4.2.15

$$\| V_n - \upsilon^* \|_\varpi = \frac{\sup_{x\in X} \mid V_n(x) - \upsilon^*(x) \mid}{\varpi(x)} \leq \frac{\hat{k}(\alpha\delta)^n}{1-\alpha\delta}. \tag{4.8}$$

Ahora, como $\mid V_n(x) - \upsilon^*(x) \mid \leq \sup_{x\in X} \mid V_n(x) - \upsilon^*(x) \mid$ y despejando a $\varpi(x)$ se tiene que

$$|V_n(x) - \upsilon^*(x)| \leq \frac{\hat{k}\varpi(x)(\alpha\delta)^n}{1-\alpha\delta}.$$

■

Lema 4.2.17 *Bajo las Condiciones 4.2.10 y 4.2.13, se tiene que existe $N \in \mathbb{N}$ tal que, para todo $n \geq N$ se satisface que*

$$|G(x, f_n(x)) - G(x, f^*(x))| \leq \frac{\hat{k}\varpi(x)(\alpha\delta)^n}{1-\alpha\delta}(\delta+1), \tag{4.9}$$

para todo $a \in A(x)$ y $x \in X$, donde $\hat{k}$, δ, $\varpi(x)$ son como en la Condición 4.2.13 y $f_n(x)$ son las políticas provenientes del método de iteración de valores.

Demostración. Sea $x \in X$ y $n \geq 1$ fijos. Usando el Lema 1.1.6 (c) y el Lema 4.2.16 se tiene que

$$\begin{aligned}
|G(x, f_n(x)) - G(x, f^*(x))| &= |G(x, f_n(x)) - \upsilon^*(x)| \\
&= |\, c(x, f_n(x)) + \alpha \int_X \upsilon^*(y) Q(dy|x, f_n(x)) \\
&+ V_n(x) - V_n(x) - \upsilon^*(x)| \\
&\leq |\, \alpha \int_X (\upsilon^*(y) - V_{n-1}(y))\, Q(dy|x, f_n(x)) \\
&+ V_n(x) - \upsilon^*(x)| \\
&\leq \alpha \int_X |\upsilon^*(y) - V_{n-1}(y)|\, Q(dy|x, f_n(x)) \\
&+ |V_n(x) - \upsilon^*(x)| \\
&\leq \frac{\hat{k}(\alpha\delta)^n}{1-\alpha\delta} \int_X \varpi(y) Q(dy|x, f_n(x)) + \frac{\hat{k}\varpi(x)(\alpha\delta)^n}{1-\alpha\delta} \\
&\leq \frac{\hat{k}(\alpha\delta)^n}{1-\alpha\delta}\delta\varpi(x) + \frac{\hat{k}\varpi(x)(\alpha\delta)^n}{1-\alpha\delta} \\
&= \frac{\hat{k}\varpi(x)(\alpha\delta)^n}{1-\alpha\delta}(\delta+1).
\end{aligned}$$

■

Lema 4.2.18 *Supóngase que se cumple la Condición 4.2.3 o la Condición 4.2.4 y que también se satisfacen las Condiciones 4.2.10 y 4.2.13. Entonces, para el modelo perturbado M_λ, cuando $f^*(x) \in int(A(x))$ se satisface que*

$$H_{aa}(x, f^*(x)) = \lambda,$$

para todo $x \in X$.

Demostración. Sea $x \in X$ fijo. Por el Teorema 4.1.18 se tiene que $H_{aa}(x,a) = \lambda(a - P_a(x,a))$, para cada $a \in int(A(x))$. Nótese que en PDMs para M_λ, $P(x, f^*(x)) = f^*(x)$ es una función que sólo depende de x y es constante en la variable a. Entonces $P_a(x, f^*(x)) = 0$. Entonces

$$H_{aa}(x, f^*(x)) = \lambda$$

para cada $x \in X$. ∎

Teorema 4.2.19 *Supóngase que las Condiciones 4.2.3 o 4.2.4 y las Condiciones 4.2.10 y 4.2.13 se cumplen. Entonces, para cada $x \in X$ existe $K(x)$ tal que $\mid f_n(x) - f^*(x) \mid \leq K(x)$, donde $K(x) = \sqrt{\dfrac{2\hat{k}\varpi(x)(\alpha\delta)^n}{\lambda(1-\alpha\delta)}(\delta+1)}$. Es decir, $f_n \xrightarrow{p} f^*$ (f_n converge puntualmente a f^*, cuando $n \to \infty$). Más aún,*

$$\mid f_n(x) - f^*(x) \mid \leq \sqrt{\frac{2\hat{k}\varpi(x)(\alpha\delta)^n}{\lambda(1-\alpha\delta)}(\delta+1)}, \tag{4.10}$$

para todo $x \in X$, donde $\hat{k}$, δ, $\varpi(x)$ son como en la Condición 4.2.13 y $f_n(x)$ son las políticas provenientes del método de iteración de valores.

Demostración. Sea $x \in X$ fijo. Primero note que $H(x, f_n(x)) < G(x, f_n(x))$, ya que la Observación 4.1.4 en particular se cumple para $f_n(x) \in A(x)$, es decir,

$$\begin{aligned} H(x, f_n(x)) &= \min_{b\in A(x)} \left\{ G(x,b) + \frac{\lambda}{2}(b - f_n(x))^2 \right\} \\ &\leq c(x, f_n(x)) + \alpha \int_X \upsilon^*(y) Q(dy|x, f_n(x)) + \frac{\lambda}{2}(f_n(x) - f_n(x))^2 \\ &= G(x, f_n(x)). \end{aligned}$$

Ahora, desarrollando la serie de Taylor alrededor de $f^*(x)$ para $H(x, f_n(x))$, donde f_n son las políticas de iteración de valores. Usando que $H_a(x, f^*(x)) = 0$ (ya que, f^* es óptima para M_λ por el Teorema 4.2.9) se tiene que

$$H(x, f_n(x)) = H(x, f^*(x)) + \frac{1}{2!}H_{aa}(x, f^*(x))(f_n(x) - f^*(x))^2, \tag{4.11}$$

debido a que las derivadas de orden tres o más son todas cero. Usando el Teorema 4.2.9 y el Lema 4.2.18 se llega a que la ecuación (4.11) es

$$H(x, f_n(x)) = G(x, f^*(x)) + \frac{1}{2}\lambda(f_n(x) - f^*(x))^2, \tag{4.12}$$

Luego usando que $H(x, f_n(x)) \leq G(x, f_n(x))$ se tiene que

$$\begin{aligned} \frac{\lambda}{2}(f_n(x) - f^*(x))^2 &= H(x, f_n(x)) - H(x, f^*(x)) \\ &\leq G(x, f_n(x)) - G(x, f^*(x)). \end{aligned} \tag{4.13}$$

Usando el Lema 4.2.17 se concluye que

$$\begin{aligned} \frac{\lambda}{2}(f_n(x) - f^*(x))^2 &\leq G(x, f_n(x)) - G(x, f^*(x)) \\ &\leq \frac{\hat{k}\varpi(x)(\alpha\delta)^n}{1 - \alpha\delta}\left(\delta + 1\right). \end{aligned} \tag{4.14}$$

Despejando en la desigualdad (4.14) se llega a que

$$|f_n(x) - f^*(x)| \leq \sqrt{\frac{2\hat{k}\varpi(x)(\alpha\delta)^n}{\lambda\left(1 - \alpha\delta\right)}\left(\delta + 1\right)}. \tag{4.15}$$

Por lo tanto, tomando el límite cuando $n \to \infty$ las políticas de iteración de valores convergen puntualmente a la política óptima (i.e., $f_n(x) \longrightarrow f^*(x)$, para todo $x \in X$). ■

Observación 4.2.20 *Nótese que en el caso de que en la Condición 4.2.13 el costo sea acotado, la convergencia es uniforme.*

El siguiente teorema muestra que la continuidad se obtiene como consecuencia de las cotas de convergencia.

Teorema 4.2.21 *Supóngase que se cumplen las Condiciones 1.1.5, las Condiciones 4.2.10 y 4.2.13 y que la función de costo $c(\cdot, \cdot)$ es continua en $\mathbb{K}$. Entonces f^* es continua, si y sólo si, v^* es continua.*

Demostración. Supongamos que f^* es continua. Entonces,

$$v^*(x) := c(x, f^*(x)) + \alpha \int_X v^*(y) Q(dy|x, f^*(x)), \tag{4.16}$$

es continua, para cada $x \in X$.
Por otro lado, supongamos υ^* continua, y tomemos una sucesión $\{x_n\}$ tal que $x_n \longrightarrow x$, usando la igualdad (4.13) y la desigualdad (4.14) se sigue que

$$\begin{aligned} |f^*(x) - f^*(x_n)|^2 &\leq G(x, f^*(x)) - G(x, f^*(x_n)) \\ &\leq \upsilon^*(x) - \upsilon^*(x_n) \\ &= 0. \end{aligned} \tag{4.17}$$

Por lo tanto, f^* es continua. ■

En este capítulo se encontró otra manera de perturbar la función de costo adicionándole una función del tipo cuadrática, con lo cual en el modelo perturbado M_λ se demostró la unicidad de la política óptima, la convexidad y diferenciabilidad de la función de valor óptimo con lo cual se lograron obtener una cota de convergencia para las políticas provenientes del método de iteración de valores.

Capítulo 5

EJEMPLOS Y APLICACIONES

5.1 Ejemplos usando la perturbación basada en el principio variacional de Ekeland

En esta sección se darán ejemplos que muestran que a pesar de que el modelo original no presenta unicidad en la política óptima, en el modelo perturbado sí hay unicidad de la política óptima para el modelo perturbado.

Ejemplo 5.1.1 *Sean $X=(0,\infty)$, $A=A(x)=(-\infty,0)$, para todo $x\in X$. La dinámica está dada por*

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

$t=0,1,\ldots$. En donde $\xi_0,\xi_1,\ldots$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas (i.i.d) con valores en $S=(-\infty,0]$ y con densidad común y continua denotada por Δ. La función de costo es dada por

$$c(x,a)=x+\mid a\mid.$$

Nótese que en este caso la función de costo es convexa.

Lema 5.1.2 *El Ejemplo 5.1.1 satisface la Condición 1.1.5 (a), (b) y (c).*

Demostración. Primero se probará que el costo es inferiormente compacto, en efecto, sea $x\in X$ y $r\in\mathbb{R}$, entonces

$$A_r(x)=\{a\in A(x)\mid c(x,a)\leq r\}.$$

Considérese los siguientes casos para $x\in X$ fijo

1. Si $r < 0$, entonces

$$A_r(x) = \{a \in A(x) \mid c(x,a) \leq r\} = \emptyset,$$

el cual es trivialmente un conjunto compacto.

2. Si $r \geq 0$, entonces

$$\begin{aligned} A_r(x) &= \{a \in A(x) \mid x + |a| \leq r\} \\ &= \{a \in A(x) \mid |a| \leq r - x\} \\ &= \{a \in A(x) \mid -a \leq r - x\} \\ &= \{a \in A(x) \mid a \geq x - r\} \\ &= \begin{cases} [x-r, 0] & si \quad x - r \leq 0 \\ \\ \emptyset & si \quad x - r \geq 0. \end{cases} \end{aligned}$$

Como $[x-r, 0]$ es un intervalo cerrado y acotado, se tiene que es compacto.

Nótese que en cualquier caso los conjuntos son compactos. Por lo tanto, el costo es inferiorrmente compacto.

Ahora, se demostrará que la ley de transición es fuertemente continua: sea $u : X \to \mathbb{R}$ una función medible y acotada, entonces usando el Teorema del Cambio de Variable (véase [2]) se tiene que

$$\begin{aligned} \int_{(0,\infty)} u(y) Q(dy \mid x, a) &= \int_{(-\infty,0]} u(xe^{a+s}) \Delta(s) ds \\ &= \int_{(0,xe^a]} u(w) \Delta(ln\frac{w}{x} - a)\frac{1}{w} dw \end{aligned}$$

para cada $(x,a) \in \mathbb{K}$. Como u es una función acotada y Δ es una función continua y acotada, entonces usando el Teorema de la Convergencia Dominada (véase [2]), se obtiene que

$$\int_{(0,xe^a]} u(w) \Delta(ln\frac{w}{x} - a)\frac{1}{w} dw.$$

es una función continua en $\mathbb{K}$. Por tanto,

$$\int_{(0,\infty)} u(y) Q(dy \mid \cdot, \cdot)$$

es una función continua en $\mathbb{K}$.

Sólo resta pobar que V es finita valuada. Para ello sea $g(x) = 0$, para cada $x \in X$. Entonces

$$\begin{aligned} V(g,x) &= E_x^g\left[\sum_{t=0}^{\infty}\alpha^t c(x_t,a_t)\right] \\ &= E_x^g\left[\sum_{t=0}^{\infty}\alpha^t\left(x_t+\mid a_t\mid\right)\right] \\ &\leq E_x^g\left[\sum_{t=0}^{\infty}\alpha^t x\right] = \frac{x}{1-\alpha}. \end{aligned}$$

■

El ejemplo anterior (Ejemplo 5.1.1) es presentado en [38], sin embargo en este trabajo se estudia otra manera de perturbar el modelo usando las técnicas expuestas en el Capítulo 3 de esta tesis. El propósito principal de este ejemplo también es mostrar que este trabajo generaliza la teoría presentada en [38].

Ejemplo 5.1.3 *Sea $X = \mathbb{R}$ y $A = A(x) = [-2,\infty)$, para todo $x \in X$. La dinámica está dada por*

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

y la función de costo

$$c(x,a) = \begin{cases} \sqrt{a} & si\ a \geq 0\ y\ x \in X \\ 0 & si\ -\infty < a < 0 \end{cases}.$$

En donde $\xi_0, \xi_1, \ldots$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas (i.i.d) con valores en $S = [0,\infty)$ y con densidad común y continua denotada por Δ. Nótese que en este caso la función de costo no es convexa.

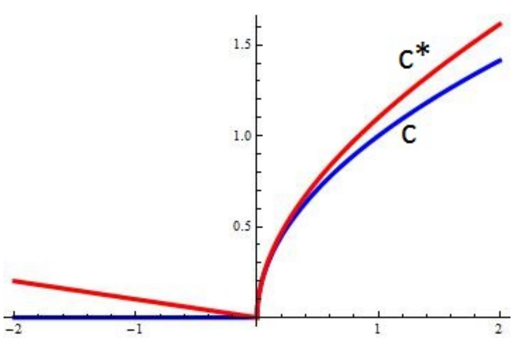

Figura 5.1: **Gráfica de c y c^***

Lema 5.1.4 *El Ejemplo 5.1.3 satisface la Condición 1.1.5 (a), (b) y (c).*

Demostración. Primero se probará que el costo es inferiormente compacto, en efecto, sea $x \in X$ fijo y $r \in \mathbb{R}$. Analicemos los siguientes casos

1. Si $r < 0$, entonces
$$A_r(x) = \{a \in A(x) \mid c(x,a) \leq r\} = \emptyset,$$
el cual es trivialmente un conjunto compacto.

2. Si $r \geq 0$, entonces
$$\begin{aligned} A_r(x) &= \left\{a \in A(x) \mid \sqrt{a} \leq r\right\} \\ &= \left\{a \in A(x) \mid \sqrt{a} \leq r^2\right\} \\ &= \left[-2, r^2\right]. \end{aligned}$$
Observe que $[-2, r^2]$ es un intervalo cerrado y acotado, es decir, es compacto.

Como en ambos casos los conjuntos son compactos. Entonces se concluye que el costo es inferiormente compacto.

Ahora, se probará que la ley de transición es fuertemente continua: sea B un elemento de la sigma álgebra de Borel de X y
$$Q(B \mid x, a) = \int I_B(x + a + s)\Delta(s)ds,$$
para cada $(x, a) \in \mathbb{K}$. Usando el Teorema de Cambio de Variable (véase [2]), se obtiene que
$$Q(B \mid x, a) = \int_B \Delta(u - x - a)du,$$
i.e., $\Delta(u-x-a)$ es una densidad para $Q(\cdot \mid x, a)$ con respecto a la medida de Lebesgue de $\mathbb{R}$. Como $\Delta(\cdot)$ es continua, entonces Q es fuertemente continua. Sólo resta pobar que V es finita. Para ello sea $g(x) = 0$, para cada $x \in X$. Entonces
$$\begin{aligned} V(g, x) &= E_x^g\left[\sum_{t=0}^{\infty} \alpha^t c(x_t, a_t)\right] \\ &= E_x^g\left[\sum_{t=0}^{\infty} \alpha^t \sqrt{a_t}\right] \\ &= 0. \end{aligned}$$

■

En el Ejemplo 5.1.3 se muestra que la perturbación basada en el principio variacional de Ekeland funciona en el caso de funciones de costo que no necesariamente son convexas. Además, también de manera gráfica es posible ver que 0 es la única política óptima del modelo perturbado (véase Figura 5.1 para $\varepsilon = .5$).

Ejemplo 5.1.5 *Sea $X = A = A(x) = \mathbb{R}$, para todo $x \in X$. La dinámica está dada por*

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

y la función de costo para $(x, a) \in \mathbb{K}$ y $\varepsilon > 0$ fijo,

$$c(x,a) = \begin{cases} \mid a \mid & si\ a \in (-\infty, -\varepsilon) \cup (\varepsilon, \infty), \\ 0 & si\ a \in [-\varepsilon, \varepsilon] \end{cases}.$$

En donde $\xi_0, \xi_1, \ldots$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas (i.i.d) con valores en $S = [0, \infty)$ y con densidad común y continua denotada por Δ.

Lema 5.1.6 *El Ejemplo 5.1.5 satisface la Condición 1.1.5 (a), (b) y (c).*

Demostración. Primero se probará que el costo es inferiormente compacto, en efecto, sea $x \in X$ fijo y $r \in \mathbb{R}$, entonces

1. Si $r < 0$, entonces

$$A_r(x) = \{a \in A(x) \mid c(x,a) \leq r\} = \emptyset,$$

el cual es trivialmente un conjunto compacto.

2. Si $r \geq 0$, entonces

$$\begin{aligned} A_r(x) &= \{a \in A(x) \mid |a| \leq r\} \\ &= \{a \in A(x) \mid -r \leq a \leq r\} \\ &= [-r, r]. \end{aligned}$$

Observe que $[-r, r]$ es un intervalo cerrado y acotado, es decir, es compacto.

Como en ambos casos los conjuntos son compactos. Entonces se concluye que el costo es inferiormente compacto.

Ahora, la prueba de que la ley de transición es fuertemente continua: sea B un elemento de la sigma álgebra de Borel de X y

$$Q(B \mid x, a) = \int I_B(x + a + s)\Delta(s)ds$$

para cada $(x, a) \in \mathbb{K}$. Usando el Teorema de Cambio de Variable (véase [2]), se obtiene que

$$Q(B \mid x, a) = \int_B \Delta(u - x - a)du$$

i.e., $\Delta(u-x-a)$ es una densidad para $Q(\cdot \mid x, a)$ con respecto a la medida de Lebesgue de $\mathbb{R}$. Como $\Delta(\cdot)$ es continua, entonces Q es fuertemente continua. Sólo resta probar que V es finita valuada. Para ello sea $g(x) = 0$, para cada $x \in X$. Entonces

$$\begin{aligned} V(g, x) &= E_x^g\left[\sum_{t=0}^{\infty} \alpha^t c(x_t, a_t)\right] \\ &= E_x^g\left[\sum_{t=0}^{\infty} \alpha^t |a_t|\right] \\ &= 0. \end{aligned}$$

■

El Ejemplo 5.1.5 muestra que en el modelo original se pueden tener muchas políticas óptimas, sin embargo en el modelo perturbado sólo se tiene una única política óptima.

5.2 Ejemplos usando la perturbación basada en la regularización de Moreau-Yosida

En esta sección se presentarán ejemplos y aplicaciones de la regularización de Moreau-Yosida. Algunos de estos ejemplos son conocidos en la literatura, sin embargo no existe literatura en donde sean tratados con estos métodos y se dé una cota de convergencia para las políticas provenientes del método de iteración de valores.

Ejemplo 5.2.1 *Sean* $X = A = A(x) = \mathbb{R}$ *para todo* $x \in X$. *La dinámica está dada por*

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

y la función de costo

$$c(x,a) = |x|.$$

En donde ξ_t *es una sucesión de variables aleatorias independientes e idénticamente distribuidas (i.i.d) con valores en* $S = [0, \infty)$. *Nótese que la función de costo es una función no diferenciable en* 0.

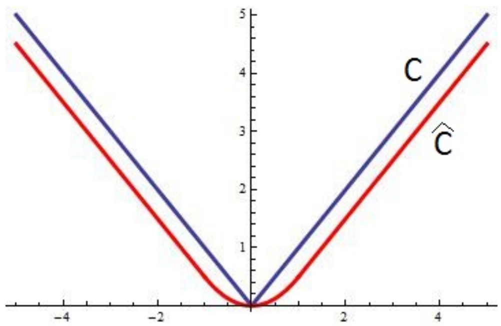

Figura 5.2: **Gráfica de** c **y** $\widehat{c}$

Lema 5.2.2 *En el Ejemplo 5.2.1 se tiene que* $f^*(x) = g^*(x) = 0$ *y que* $\upsilon^*(x) = 0 = \omega^*(x)$.

Demostración. Usando que $\upsilon_0(x) = 0$, se tiene para $n = 1$

$$H_1(x,a) = \min_{b \in A(x)} \left\{ G_1(x,b) + \frac{\lambda}{2}(b-a)^2 \right\},$$

como $G_1(x,a) = c(x,a)$ y $f_1(x) = 0$ se obtiene que

$$H_1(x,0) = \min_{b \in A(x)} \left\{ |x| + \frac{\lambda}{2} b^2 \right\} = |x|,$$

Para poder minimizar H_1 sobre las $a \in A(x)$, primero obsérvese que H_1 es una función diferenciable en todo A, así, derivando e igualando a cero, se obtiene que $f_1(x) = g_1(x) = 0$ y por tanto $\upsilon_1(x) = 0 = \omega_1(x)$. Continuando de la misma forma se obtiene que $f^*(x) = g^*(x) = 0$, de donde se concluye que $\upsilon^*(x) = 0 = \omega^*(x)$. ■

El Ejemplo 5.2.1 muestra de manera gráfica (véase Figura 5.2, para $\lambda = 1$) cómo la regularización de Moreau-Yosida convierte a una función no diferenciable en una función diferenciable. Es importante mencionar que este ejemplo no cumple las condiciones para tener una cota de las políticas provenientes del método de iteración de valores. Es decir, sólo es posible obtener las conclusiones de los Teoremas 4.2.7, 4.2.9 y 4.2.12.

Ejemplo 5.2.3 (Un sistema de inventario producción) *Sea $X = \mathbb{R}$, y $A = A(x) = [0, \infty)$ para todo $x \in X$. La dinámica está dada por*

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

para $t = 0, 1, \ldots$. La función de costo es dada por

$$c(x_t, a_t, \xi_t) = qa_t + h \cdot \text{máx}(0, x_{t+1}) + p \cdot \text{máx}(0, -x_{t+1}), \tag{5.1}$$

en donde,
q:= costo por unidad de producción,
h:=costo unitario por conservar un excesivo inventario, y
p:=costo unitario por una demanda insatisfecha.

Estas unidades de costo son todas positivas, y se supone que $p > q$. Más aún $\xi_0, \xi_1, \ldots,$ son variables aleatorias no negativas, independientes e idénticamente distribuidas, independientes del stock inicial x_0, las cuales toman valores en $S = \mathbb{R}$ y con función de distribución de probabilidad denotada por F, que es, $F(s) = P(\xi_0 \leq s)$, para todo $s \in \mathbb{R}$, con $F(s) = 0$ si $s < 0$. También se supone que la media de la demanda $E(\xi_0) = \int sdF(s)$ es finita.

Lema 5.2.4 *El Ejemplo 5.2.3 satisface la Condición 1.1.5 (a), (b) y (c).*

Demostración. Primero se probará que la función de costo es inferiormente compacta, para ello sea $r \in \mathbb{R}$ y $x \in X$, entonces

$$\begin{aligned} & \{a \in A(x) \mid c(a, r) \leq r\} \\ = \ & \{a \in A(x) \mid qa + h \cdot \text{máx}(0, x) + p \cdot \text{máx}(0, -x) \leq r\} \\ = \ & [0, qa + h \cdot \text{máx}(0, x) + p \cdot \text{máx}(0, -x)], \end{aligned}$$

el cual es un intervalo compacto. Por tanto, la función de costo es inferiormente compacta.

Ahora, se demostrará que la ley de transición es fuertemente continua. Sea $\mu : X \to \mathbb{R}$ una función medible y acotada. Tome $(x_k, a_k) \in \mathbb{K}$ tal que $(x_k, a_k) \to (x, a) \in \mathbb{K}$, $k \to \infty$. Entonces para cada $k = 0, 1, \ldots$, usando el Teorema de Cambio de Variable (véase [2]), se tiene que

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

Por otro lado,

$$\begin{aligned}(-\infty, x+a) &\subset \liminf(-\infty, x_k + a_k] \\ &\subset \limsup(-\infty, x_k + a_k] \\ &\subset (-\infty, x+a].\end{aligned}$$

Por tanto, $\left\{I_{(-\infty,x_k+a_k]}\right\}$ converge a $I_{(-\infty,x+a]}$ casi dondequiera con respecto a la medida de Lebesgue m en $\mathbb{R}$.

Sea $\theta \in \mathbb{R}$ una cota para μ, i.e., $|\mu(x)| \leq \theta$, para todo $x \in X$. Defina $\eta_k(\cdot) := I_{(-\infty,x_k+a_k]}(\cdot)\mu(\cdot)\Delta(x_k + a_k - \cdot)$, $\eta(\cdot) := I_{(-\infty,x+a]}(\cdot)\mu(\cdot)\Delta(x + a - \cdot)$, $\widehat{g}_k(\cdot) := \theta I_{(-\infty,x_k+a_k]}(\cdot)\Delta(x_k + a_k - \cdot)$, $\widehat{g}(\cdot) := \theta I_{(-\infty,x+a]}(\cdot)\Delta(x + a - \cdot)$, para todo $k = 0, 1, \ldots$ y $(x, a), (x_k, a_k) \in \mathbb{K}$. Nótese que $\{\eta_k\}$ y $\{\widehat{g}_k\}$ converge casi dondequiera respecto a m a η y $\widehat{g}$, respectivamente. Además, para cada $k = 0, 1, \ldots$, $|\eta_k|(\cdot) \leq \widehat{g}_k(\cdot)$ y usando el Teorema de Cambio de Variable (véase [2]) se obtiene que

$$\int \widehat{g}_k(l)dl = \int \widehat{g}(l)dl = \theta,$$

para cada $k = 0, 1, \ldots$. Por tanto, usando una versión generalizada del Teorema de la Convergencia Dominada (véase [49], p. 92) se obtiene que $\int \eta_k(l)dl = \int \eta(l)dl$, $k \to \infty$, i.e.,

$$\int_{(-\infty,x_k+a_k]} \mu(l)\Delta(x_k + a_k - l)dl \to \int_{(-\infty,x+a]} \mu(l)\Delta(x + a - l)dl,$$

$k \to \infty$. Por lo tanto, Q es fuertemente continua.

Finalmente, para probar que la función de valor óptimo es finita, tome $f(x) = 0$, para todo $x \in X$ y denote $\widehat{\mu} := E[\xi]$.

Si $x < 0$, entonces

$$c(x, f) = pE\left[\max(0, \xi - x)\right] \leq p(-x) + \widehat{\mu}p. \tag{5.2}$$

Si $x \geq 0$, se tiene que

$$c(x, f) = hE\left[\max(0, x - \xi)\right] + pE\left[\max(0, \xi - x)\right] \leq hx + \widehat{\mu}p. \tag{5.3}$$

Usando (5.2) y (5.3) se concluye que

$$c(x, f) = r'|x| + \widehat{\mu}p,$$

para todo $x \in X$, donde $r' := \text{máx}(p, h)$ y usando inducción se llega a que

$$E_x^f \left[c(x_t, f) \right] \leq r'|x| + t(\widehat{\mu} r') + \widehat{\mu} p, \tag{5.4}$$

para todo $x \in X$ y $t = 0, 1, \ldots$. Ahora, para cada $x \in X$, usando (5.4)

$$V(f, x) \leq \sum_{t=0}^{\infty} \alpha^t \left(r'|x| + t(\widehat{\mu} r') + \widehat{\mu} p \right) = \frac{r'|x|}{1-\alpha} + \frac{\widehat{\mu} r'|x| \alpha}{(1-\alpha)^2} + \frac{\widehat{\mu} p}{1-\alpha} < \infty,$$

por lo tanto $V(\pi, x) < \infty$ para $\pi = \{f, f, \ldots\}$ y $x \in X$. ■

El Ejemplo 5.2.3 es conocido en la literatura como inventarios de producción (véase [9] y [25]), el cual presenta un costo no diferenciable, pero sin embargo usando el método de perturbación antes propuesto el costo perturbado es diferenciable, pero no se da de manera explícita la derivada de este. Es de suma importancia mencionar que este ejemplo no cumple las condiciones para tener una cota de las políticas provenientes del método de iteración de valores. Es decir, sólo es posible obtener las conclusiones de los Teoremas 4.2.7, 4.2.9 y 4.2.12.

Ejemplo 5.2.5 (Lineal) *Sea $X = A = \mathbb{R}$ y $A(x) = [-1, 1]$. La dinámica está dada por*

$$x_{t+1} = x_t + a_t + \xi_t, \ \textit{para cada } t = 0, 1, \ldots, \tag{5.5}$$

en donde $\xi_t \sim N(0, 1)$ son variables aleatorias independientes e idénticamente distribuidas normal estándar y la función de costo está dada por

$$c(x, a) = \mid x \mid + \mid a \mid . \tag{5.6}$$

Solución 5.2.6 *La solución de la política óptima del modelo perturbado se obtendrá mediante el método de iteración de valores.*

$$\begin{aligned} H^1(x, a) &= \min_{b \in A(x)} \left\{ \mid x \mid + \mid b \mid + \frac{\lambda}{2} (b - a)^2 \right\} \\ &= \min_{b \in A(x)} \left\{ \mid x \mid + \mid b \mid + \frac{\lambda}{2} \left(b^2 - 2ab + a^2 \right) \right\} \\ &= \begin{cases} \mid x \mid + \dfrac{\lambda}{2} a^2 & si \ \mid a \mid < 1/\lambda \\ \mid x \mid + \mid a \mid - \dfrac{1}{2\lambda} & si \ \mid a \mid \geq 1/\lambda. \end{cases} \end{aligned}$$

Entonces

$$f_1(x) = g_1(x) = 0$$

y

$$V_1(x) = W_1(x) = \mid x \mid .$$

$$H^2(x,a) \;=\; \min_{b\in A(x)} \left\{ \mid x \mid + \mid b \mid + \alpha E|x+b+\xi| + \frac{\lambda}{2}(b-a)^2 \right\}.$$

Ahora, calculando la $E|x+b+\xi|$, *se tiene que*

$$\begin{aligned}
E|x+b+\xi| &= \int_0^1 |x+b+s|ds \\
&= \int_0^{-(x+b)} -(x+b+s)ds + \int_{-(x+b)}^1 (x+b+s)ds \\
&= \left[(x+b)s + \frac{s^2}{2}\right]_{-(x+b)}^1 + \left[-(x+b)s - \frac{s^2}{2}\right]_0^{-(x+b)} \\
&= (x+b) + \frac{1}{2} + (x+b)^2 - \frac{(x+b)^2}{2} + (x+b)^2 - \frac{(x+b)^2}{2} \\
&= x + b + \frac{1}{2} + (x+b)^2 = x^2 + x + \frac{1}{2} + b^2 + 2bx + b.
\end{aligned}$$

Entonces

$$\begin{aligned}
H^2(x,a) &= \min_{b\in A(x)} \left\{ \mid x \mid + \mid b \mid + \alpha E|x+b+\xi| + \frac{\lambda}{2}(b-a)^2 \right\} \\
&= \min_{b\in A(x)} \left\{ \mid x \mid + \mid b \mid + \alpha \left[(x+b)^2 + x + \frac{1}{2} + b\right] + \frac{\lambda}{2}(b-a)^2 \right\} \\
&= |x| + \alpha x^2 + \alpha x + \frac{\alpha}{2} \\
&+ \min_{b\in A(x)} \left\{ \mid b \mid + \alpha b^2 + 2\alpha bx + \alpha b + \frac{\lambda}{2}(b-a)^2 \right\} \\
&= |x| + \alpha x^2 + \alpha x + \frac{\alpha}{2} \\
&+ \frac{\lambda}{2} \min_{b\in A(x)} \left\{ b^2 + 2\left(\frac{1}{\lambda} - |a|\right)|b| + a^2 + \frac{2\alpha}{\lambda}b^2 + \frac{4\alpha}{\lambda}bx + \frac{2\alpha}{\lambda}b \right\} \\
&= |x| + \alpha x^2 + \alpha x + \frac{\alpha}{2} + \frac{\lambda}{2}a^2 \\
&+ \frac{\lambda}{2} \min_{b\in A(x)} \left\{ \left(\frac{2\alpha+\lambda}{\lambda}\right) b^2 + 2\left(\frac{2\alpha}{\lambda}x + \frac{\alpha}{\lambda} + \frac{1}{\lambda} - |a|\right)|b| \right\}.
\end{aligned}$$

De donde se sigue que

$$
\begin{aligned}
H^2(x,a) &= |x| + \alpha x^2 + \alpha x + \frac{\alpha}{2} + \frac{\lambda}{2}a^2 \\
&+ \frac{\lambda}{2} \min_{\eta \geq 0} \min_{\eta = |b|} \left\{ \left(\frac{2\alpha + \lambda}{\lambda} \right) \eta^2 + 2 \left(\frac{1}{\lambda} (2\alpha x + 2\alpha + 1) - |a| \right) \eta \right\}.
\end{aligned}
$$

Por tanto,

$$H^2(x,a) = |x| + \alpha x^2 + \alpha x + \frac{\alpha}{2} + \frac{\lambda}{2}a^2$$

cuando $\eta = 0$ *y*

$$
\begin{aligned}
|a| &< \frac{1}{\lambda} (2\alpha x + 2\alpha + 1) \\
|a| - \frac{2\alpha}{\lambda} x &< \frac{1}{\lambda} (2\alpha + 1) \\
\frac{\lambda |a| - 2\alpha x}{\lambda} &< \frac{1}{\lambda} (2\alpha + 1) \\
\lambda |a| - 2\alpha x &< (2\alpha + 1) .
\end{aligned}
$$

En otro caso,

$$\left(\frac{2\alpha + \lambda}{\lambda} \right) \eta + 2 \left(\frac{1}{\lambda} (2\alpha x + 2\alpha + 1) - |a| \right) = 0,$$

despejando η *se obtiene*

$$\eta = \frac{\lambda}{2\alpha + \lambda} \left[|a| - \frac{1}{\lambda} (2\alpha x + 2\alpha + 1) \right]$$

y

$$
\begin{aligned}
\lambda |a| - 2\alpha x &\geq 2\alpha + 1 \\
|a| &\geq \frac{1}{\lambda} [2\alpha (x+1) + 1] .
\end{aligned}
$$

$$\begin{aligned}
H^2(x,a) &= |x| + \alpha x^2 + \alpha x + \frac{\alpha}{2} + \frac{\lambda}{2}a^2 \\
&+ \frac{\lambda}{2}\left\{\left(\frac{2\alpha+\lambda}{\lambda}\right)\left(\frac{\lambda}{2\alpha+\lambda}\left[|a| - \frac{1}{\lambda}(2\alpha x + 2\alpha + 1)\right]\right)^2\right. \\
&+ \left. 2\left(\frac{1}{\lambda}(2\alpha x + 2\alpha + 1) - |a|\right)\left(\frac{\lambda}{2\alpha+\lambda}\left[|a| - \frac{1}{\lambda}(2\alpha x + 2\alpha + 1)\right]\right)\right\} \\
&= |x| + \alpha x^2 + \alpha x + \frac{\alpha}{2} + \frac{\lambda}{2}a^2 \\
&+ \frac{\lambda}{2}\left\{\left(\frac{\lambda}{2\alpha+\lambda}\right)\left(|a| - \frac{1}{\lambda}(2\alpha x + 2\alpha + 1)\right)^2\right. \\
&- \left. 2\left(\frac{\lambda}{2\alpha+\lambda}\right)\left(|a| - \frac{1}{\lambda}(2\alpha x + 2\alpha + 1)\right)^2\right\} \\
&= |x| + \alpha x^2 + \alpha x + \frac{\alpha}{2} + \frac{\lambda}{2}a^2 \\
&- \frac{\lambda}{2}\left\{\left(\frac{\lambda}{2\alpha+\lambda}\right)\left(|a| - \frac{1}{\lambda}(2\alpha x + 2\alpha + 1)\right)^2\right\}.
\end{aligned}$$

Desarrollando lo anterior se tiene que

$$\begin{aligned}
H^2(x,a) &= |x| + \alpha x^2 + \alpha x + \frac{\alpha}{2} + \frac{\lambda}{2}a^2 \\
&- \frac{\lambda}{2}\left\{\left(\frac{\lambda}{2\alpha+\lambda}\right)\left(a^2 - 2|a|\frac{1}{\lambda}(2\alpha x + 2\alpha + 1)\right.\right. \\
&+ \left.\left. \frac{1}{\lambda^2}(2\alpha x + 2\alpha + 1)^2\right)\right\} \\
&= |x| + \alpha x^2 + \alpha x + \frac{\alpha}{2} + \frac{\lambda}{2}a^2 \\
&- \frac{\lambda}{2}\left\{\left(\frac{\lambda}{2\alpha+\lambda}\right)\left(a^2 - 2|a|\frac{1}{\lambda}(2\alpha x + 2\alpha + 1)\right.\right. \\
&+ \left.\left. \frac{1}{\lambda^2}\left(4\alpha^2 x^2 + 4\alpha x(2\alpha+1) + (2\alpha+1)^2\right)\right)\right\}.
\end{aligned}$$

Continuando se llega a que

$$\begin{aligned} H^2(x,a) &= |x| + \alpha x^2 + \alpha x + \frac{\alpha}{2} + \frac{\lambda}{2}a^2 \\ &- \frac{\lambda}{2}\left\{\left(\frac{\lambda}{2\alpha+\lambda}\right)\left(a^2 - 2|a|\frac{1}{\lambda}\left(2\alpha x + 2\alpha + 1\right)\right.\right. \\ &+ \left.\left.\frac{1}{\lambda^2}\left(4\alpha^2x^2 + 8\alpha^2 x + 4\alpha x + 4\alpha^2 + 4\alpha + 1\right)\right)\right\}. \end{aligned}$$

Simplificando la ecuación anterior se obtiene que

$$\begin{aligned} H^2(x,a) &= |x| + \frac{\alpha\lambda}{2\alpha+\lambda}x^2 + \frac{\alpha\lambda - 2\alpha^2 - 2\alpha}{2\alpha+\lambda}x + \frac{\lambda^2 + 2\alpha\lambda - 2\lambda}{2(2\alpha+\lambda)}a^2 \\ &+ \frac{\lambda}{2\alpha+\lambda}\left(2\alpha x + 2\alpha + 1\right)|a| + \frac{\alpha\lambda - 2\alpha^2 - 4\alpha - 1}{2(2\alpha+\lambda)}. \end{aligned}$$

Por lo tanto,

$$H^2(x,a) = \begin{cases} |x| + \alpha x^2 + \alpha x + \frac{\alpha}{2} + \frac{\lambda}{2}a^2 & si \quad |a| < \frac{1}{\lambda}\left[2\alpha(x+1)+1\right] \\ \\ |x| + \frac{\alpha\lambda}{2\alpha+\lambda}x^2 + \frac{\alpha\lambda - 2\alpha^2 - 2\alpha}{2\alpha+\lambda}x & \\ +\frac{\lambda^2 + 2\alpha\lambda - 2\lambda}{2(2\alpha+\lambda)}a^2 & \\ +\frac{\lambda}{2\alpha+\lambda}\left(2\alpha x + 2\alpha + 1\right)|a| & \\ +\frac{\alpha\lambda - 2\alpha^2 - 4\alpha - 1}{2(2\alpha+\lambda)} & si \quad |a| \geq \frac{1}{\lambda}\left[2\alpha(x+1)+1\right] \end{cases}$$

con lo que se concluye que

$$f_2(x) = g_2(x) = 0$$

y

$$W_2(x) = V_2(x) = |x| + \alpha x^2 + \alpha x + \frac{\alpha}{2}.$$

De manera análoga se obtiene

$$H^3(x,a) = |x| + \frac{\lambda}{2}a^2 + \alpha(1+\alpha)x^2 + 2\alpha(1+\alpha)x + \alpha(1 + \frac{5}{6}\alpha)$$

si $|a| < \frac{\alpha}{\lambda}\left[2\alpha(x+1) + 2(\alpha+1) + \frac{1}{\alpha}\right]$

$$
\begin{aligned}
H^3(x,a) &= |x| + \frac{2\lambda\alpha(1+\alpha) - \lambda^2}{2\left(\lambda + 2\alpha(1+\alpha)\right)}a^2 \\
&+ \frac{\lambda\left(2\alpha^2(x+1) + 2\alpha(\alpha+1) + 1\right)}{2\left(\lambda + 2\alpha(1+\alpha)\right)}|a| \\
&+ \frac{\alpha\left(\lambda + 2\alpha + \alpha\lambda + 2\alpha^2 - 2\alpha^3\right)}{\lambda + 2\alpha(1+\alpha)}x^2 \\
&+ \frac{\alpha\left(\lambda - \alpha + \alpha\lambda + \alpha^2 - \alpha^3\right)}{\lambda + 2\alpha(1+\alpha)}x \\
&+ \frac{2\alpha\lambda - 4\alpha^2 - \frac{2}{3}\alpha^3 + \frac{5}{3}\lambda\alpha^2 - \frac{2}{3}\alpha^4 - 4\alpha - 1}{2\left(\lambda + 2\alpha(1+\alpha)\right)},
\end{aligned}
$$

si $|a| \geq \frac{\alpha}{\lambda}\left[2\alpha(x+1) + 2(\alpha+1) + \frac{1}{\alpha}\right].$

Entonces

$$f_3(x) = g_3(x) = 0$$

y

$$W_3(x) = V_3(x) = |x| + \alpha(1+\alpha)x^2 + 2\alpha(1+\alpha)x + \alpha\left(1 + \frac{\alpha}{2} + \frac{\alpha}{3}\right).$$

Con lo anterior se obtuvo que H es clase C^2 y luego usando el Teorema 4.2.19 se tiene que

$$|f_n(x) - f^*(x)| \leq \sqrt{\frac{2\hat{k}\varpi(x)(\alpha\delta)^n}{\lambda\,(1-\alpha\delta)}\,(\delta+1)}.$$

Ahora, se hará una aproximación de la política óptima f^* con un error $= 10^{-5}$. Nótese que $\widehat{k} = 1 = \delta$ y $\varpi(x) = 1 + \sigma^2 = 1 + 1 = 2$, para cada $x \in X$. Entonces es suficiente hacer sólo una iteración. Por lo tanto,

$$|f_1(x) - f^*(x)| = |0 - f^*(x)| \leq 10^{-5}.$$

Dicho de otra manera, $f^*(x) = 0$, para todo $x \in X$.

De manera análoga,

$$|V_1(x) - \upsilon^*(x)| = ||x| - \upsilon^*(x)| \leq 10^{-5}.$$

Es decir, $\upsilon^*(x) = |x|$, para todo $x \in X$.

El Ejemplo 5.2.5 muestra cómo la regularización de Moreau-Yosida también funciona como un método para calcular la política óptima del modelo original y del modelo perturbado. Es decir, para este ejemplo se obtienen las conclusiones de los Teoremas 4.2.7, 4.2.9, 4.2.12 y 4.2.19.

Ejemplo 5.2.7 (Lineal Cuadrático) *Sea $X = A = \mathbb{R}$. La dinámica está dada por*

$$x_{t+1} = \gamma x_t + \beta a_t + \xi_t, \ \textit{para cada } t = 0, 1, \ldots, \tag{5.7}$$

y la función de costo

$$c(x, a) = qx^2 + ra^2. \tag{5.8}$$

Condición 5.2.8 *(a) Supóngase que $\gamma\beta \neq 0$, y que q y r son positivos.*

(b) Las variables aleatorias ξ_t, para cada $t = 0, 1, \ldots$ son independientes e idénticamente distribuidas con valores en $S = \mathbb{R}$. Más aún, supóngase que ξ_0 tiene densidad continua Δ, con media 0 y varianza σ^2; es decir,

$$E(\xi_t) = \int sg(s)ds = 0,$$

y

$$E(\xi_t^2) = \int s^2 g(s)ds = \sigma^2 < \infty.$$

(c) Para cada $x \in X$, los conjuntos de control son los intervalos

$$A(x) := \left[- \left|\gamma/\beta\right| x, \left|\gamma/\beta\right| x\right].$$

Lema 5.2.9 *Bajo la Condición 5.2.8, se tiene que en el Ejemplo 5.2.7*

(a) c es inferiormente compacto y estrictamente convexo en $\mathbb{K}$.

(b) La ley de transición, $Q(\cdot \mid x, a)$, $a \in A(x)$ inducida por (5.7) es fuertemente continua.

(c) Se cumple la Condición 4.2.10 (2).

(d) La multifunción $x \twoheadrightarrow A(x)$ es continua, para todo $x \in X$.

Demostración. Para la prueba de (a) y (b) véase [16] Lema 4.9.

Para demostrar (c), es suficiente tomar $\hat{k} = q + r\dfrac{\gamma^2}{\beta^2}$, $\delta = 1$ y $\varpi(x) = x^2 + \sigma^2$.

(d) Primero se probará que la multifunción es i.s.c. para ello, sea $x_n \to x$ en X y $a \in A(x) = [-|\gamma/\beta|\, x, |\gamma/\beta|\, x]$. Si $a_n = -|\gamma/\beta|\, x_n$ ó $a_n = |\gamma/\beta|\, x_n$, entonces $a_n \to -|\gamma/\beta|\, x$ ó $a_n \to |\gamma/\beta|\, x$, cuando $n \to \infty$. De otra forma, si $a \in (-|\gamma/\beta|\, x, |\gamma/\beta|\, x)$, entonces $a = \lambda(-|\gamma/\beta|\, x) + (1-\lambda)\, |\gamma/\beta|\, x$, para $\lambda \in (0,1)$. Entonces $a_n = \lambda(-|\gamma/\beta|\, x_n) + (1-\lambda)\, |\gamma/\beta|\, x_n$, tomando límite cuando $n \to \infty$ se tiene que $a_n \to a$. Por lo tanto, la multifunción $x \twoheadrightarrow A(x)$ es i.s.c.

Ahora, para demostrar que la multifunción es s.s.c., sea $x_n \to x$ en X y $a_n \in A(x_n) = [-|\gamma/\beta|\, x_n, |\gamma/\beta|\, x_n]$. Entonces $-|\gamma/\beta|\, x_n \le a_n \le |\gamma/\beta|\, x_n$, tomando límite cuando $n \to \infty$ se tiene que

$$-|\gamma/\beta|\, x \le \liminf_{n\to\infty} a_n \le \limsup_{n\to\infty} a_n \le |\gamma/\beta|\, x.$$

Usando el Teorema 3.17 en [50], pág. 59 se obtiene que la multifunción $x \twoheadrightarrow A(x)$ es s.s.c.

Con lo anterior se concluye que la multifunción $x \twoheadrightarrow A(x)$ es continua.

■

Para la prueba del siguiente lema véase [25].

Lema 5.2.10 *Bajo la Condición 5.2.8 se tiene que en el Ejemplo 5.2.7 la función de valor óptimo es*

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

y

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

para cada $x \in \mathbb{R}$ y la política óptima es

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

y

$$f_n^*(x) = K_n x,$$

para cada $x \in \mathbb{R}$, donde $K > 0$ es la solución única de la ecuación

$$K = \left[\frac{1 - \alpha K \beta^2}{r + \alpha K \beta^2}\right] \alpha K \gamma^2 + q, \tag{5.9}$$

y $K_n > 0$ es la solución única de la ecuación

$$K_n = \left[\frac{1 - \alpha K_n \beta^2}{r + \alpha K_n \beta^2}\right] \alpha K_n \gamma^2 + q. \tag{5.10}$$

Observación 5.2.11 *Nótese que $f^*(x)$ siempre está en $A(x)$ para todo $x \in X$, ya que si $f^*(x) \in A(x)$ si y sólo si $-|\frac{\gamma}{\beta}|x \leq -\frac{\alpha K \gamma \beta}{r + \alpha K \beta^2} x \leq |\frac{\gamma}{\beta}|x$ si y sólo si $-\frac{\gamma}{\beta} \leq \frac{\alpha K \gamma \beta}{r + \alpha K \beta^2} \leq \frac{\gamma}{\beta}$ equivalentemente $-1 \leq \frac{\alpha K \beta^2}{r + \alpha K \beta^2} \leq 1$, es decir, $-1 - 2\alpha K \beta^2 \leq 0 \leq r$.*

Teorema 5.2.12 *En el Ejemplo 5.2.7 se tiene que*

$$| G(x, f_n(x)) - G(x, f^*(x)) | \leq \frac{2\alpha^n \left(q + r\frac{\gamma^2}{\beta^2}\right)(x + \sigma^2)}{1 - \alpha},$$

para cada $x \in \mathbb{R}$ y

$$| f_n(x) - f^*(x) | \leq 2\sqrt{\frac{\alpha^n \left(q + r\frac{\gamma^2}{\beta^2}\right)(x + \sigma^2)}{\lambda (1 - \alpha)}},$$

para cada $x \in \mathbb{R}$.

Demostración. Sea $\hat{k}$, δ y $\varpi(x)$ son como en la demostración del Lema 5.2.9 (c) y usando el Lema 4.2.17 se tiene que

$$| G(x, f_n(x)) - G(x, f^*(x)) | \leq \frac{\hat{k}\varpi(x)(\alpha\delta)^n}{1 - \alpha\delta}(\delta + 1) = \frac{2\alpha^n \left(q + r\frac{\gamma^2}{\beta^2}\right)(x + \sigma^2)}{1 - \alpha}.$$

Ahora, usando el Teorema 4.2.19

$$| f_n(x) - f^*(x) | \leq \sqrt{\frac{2\hat{k}\varpi(x)(\alpha\delta)^n}{\lambda (1 - \alpha\delta)}(\delta + 1)} = 2\sqrt{\frac{\alpha^n \left(q + r\frac{\gamma^2}{\beta^2}\right)(x + \sigma^2)}{\lambda (1 - \alpha)}}.$$

■

El Ejemplo 5.2.7 es muy conocido en la literatura de PDMs, incluso se han probado que las políticas provenientes del método de iteración de valores convergen a la política óptima, sin embargo en este trabajo se da de manera explícita una cota de convergencia puntual. Es decir, para este ejemplo se obtienen las conclusiones de los Teoremas 4.2.7, 4.2.9, 4.2.12 y 4.2.19.

Capítulo 6

CONCLUSIONES Y PROBLEMAS ABIERTOS

6.1 Conclusiones

Considérese un proceso de decisión de Markov (PDM) descontado fijo, asociado a un modelo de control de Markov (MCM) $(X, A, \{A(x)| x \in X\}, Q, c)$, el cual le llamaremos proceso original y será denotado por M. A este PDM le asignaremos un PDM con las mismas componentes del MCM asociado al proceso original, excepto en la función de costo, la cual se remplazará por otra función de costo c^* o $\widehat{c}$ que difiere del costo original por una función apropiada. A la nueva función de costo le llamaremos costo perturbado. Al PDM con el MCM $(X, A, \{A(x)| x \in X\}, Q, c^*(\widehat{c}))$ le llamaremos PDM perturbado y será denotado por M_ε o M_λ.

En esta tesis se estudiaron los PDMs descontados a tiempo discreto. En este tipo de procesos se dan dos metodologías para perturbar el modelo original M, de tal manera que en el nuevo modelo perturbado $M_\varepsilon(M_\lambda)$ se pueden probar la unicidad de la política óptima, la diferenciabilidad del modelo perturbado M_λ y además de dar cotas de convergencia uniforme para las políticas de iteración de valores.

En el primer capítulo se dieron los preliminares, en donde se aborda las definiciones básicas para PDMs descontados y la versión estándar del Teorema de Ekeland, así como la metodología para perturbar según Moreau-Yosida.

En el segundo capítulo se expusieron los antecedentes a este trabajo que se encontraron en la literatura referente a PDMs descontados. Es importante mencionar que en este trabajo se lograron obtener condiciones más generales que las que se dan en los trabajos expuestos en este capítulo de antecedentes, lo cual hace que este trabajo pase a ser una generalización a cada uno de estos resultados encontrados en cada trabajo.

En el Capítulo 3 se generaliza [16], debido a que aquí se prueba la unicidad en el modelo perturbado M_ε sin usar condiciones de convexidad. Además, se hace una extensión del principio variacional de Ekeland a PDMs, ya que, el resultado que se da para PDMs no es un corolario directo del Teorema de Ekeland. Sin embargo, su prueba clásica de este principio puede adaptarse a PDMs cuando tanto el espacio de estados y de acciones son ambos espacios de Borel.

En el Capítulo 4 se da una forma de cómo aplicar la regularización de Moreau-Yosida a PDMs descontados con espacios de estados y acciones subconjuntos de $\mathbb{R}$. Con esta perturbación se obtiene la unicidad de política óptima del modelo perturbado M_λ, así como la convexidad y la diferenciabilidad en el modelo perturbado M_λ. Es importante mencionar que para demostrar la diferenciabilidad en el modelo perturbado se usaron técnicas de subdiferenciabilidad y superdiferenciabilidad. Además, como una consecuencia de la diferenciabilidad se dan cotas de convergencia para las políticas de provenientes del método de iteración de valores.

Por último se resolvieron una serie de ejemplos usando ambas formas de perturbar el modelo. Cabe mencionar que esto nos muestra el gran potencial que tiene este tipo de perturbaciones aplicadas a procesos de decisión de Markov descontados.

6.2 Trabajos futuros

A lo largo del desarrollo de este trabajo surgieron algunos problemas nuevos, los cuales se pretende resolver en un futuro, tales como:

- Demostrar de manera precisa la existencia de una política ε-óptima; es importante mencionar que Bertsekas en [9] dio una prueba, sin embargo, en este texto se omite la prueba de la medibilidad de dicha política, lo cual no se puede obtener de manera directa).

- Determinar una forma más general de perturbar el modelo, de tal manera que en el perturbado se siga cumpliendo la unicidad, la diferenciabilidad y además también se puedan dar cotas de convergencia uniforme.

- Implementar estas técnicas en algún otro criterio de rendimiento, como puede ser el costo promedio esperado.
- Aplicar este tipo de perturbaciones a modelos sensibles al riesgo.
- También otra línea a donde se podría aplicar ésto es a la teoría de juegos.

Bibliografía

[1] C. D. Aliprantis and O. Burkinshaw: "Principles of Real Analysis". Academic Press, Third Edition, 1998.

[2] R. B. Ash and C. A. Doléans-Dade: "Probability and Measure Theory". Academic Press Elsevier, Second Edition, 2005.

[3] K. Avrachenkov, J. A. Filar, V. Gaitsgory and A. Stillman: "Singularly perturbed linear programs and Markov decision processes". Oper. Res. Lett., Vol. 44, 297-301, 2016.

[4] H. Attouch and H. Riahi: "Stability results for Ekeland's ε-variational principle and cone extremal solutions". Math. Oper. Res., Vol. 18, No. 1, 173-201, 1993.

[5] N. Bäuerle and U. Rieder: "Markov Decision Processes with Applications to Finance". Springer. First Edition, 2011.

[6] H. H. Bauschke, S. M. Moffat, and X. Wang: "Self-dual smooth approximations of convex functions via the proximal average". Fixed-Point Algorithms for Inverse Problems in Science and Engineering Volume 49 of the series Springer Optimization and Its Applications, 23-32, 2011.

[7] R. Bellman: "Dynamic Programming". Princeton University Press, Princeton, 1957.

[8] A. Belloni: "Lecture notes for IAP 2005 course introduction to bundle methods". Version of February 11, 2005. (Se puede conseguir en la dirección: https://faculty.fuqua.duke.edu/∼abn5/LecturesIntroBundle.pdf).

[9] D. P. Bertsekas: "Dynamic Programming: Deterministic and Stochastic Models". Prentice-Hall, NJ, 1987.

[10] D. P. Bertsekas and S. E. Shreve: "Stochastic Optimal Control: The Discrete-Time Case". Athena Scientific, First Edition, 2007.

[11] E. Bishop and R. R. Phelps: "The support functionals of a convex set". In: Proc. Sympos. Pure Math., Vol. VII, (V. L. Klee, Editor.), Amer. Math. Soc., 27–35, 1963.

[12] J. M. Borwein and Q. J. Zhu: "Techniques of Variational Analysis". Springer, New York, 2005.

[13] W. A. Brock and L. J. Mirman: "Optimal economic growth and uncertainty: the discounted case". J. Econ. Theory, Vol. 4, 479-513, 1972.

[14] F. A. Clarke: "Methods of Dynamic and Nonsmooth Optimization". CBMS-NFS Regional Conference Series in Applied Mathematics. First Edition, 1989.

[15] A. Clausen and C. Strub: "Envelope theorems for non-smooth and non-concave optimization". www.carlostrub.ch, 5 Febrero del 2012.

[16] D. Cruz-Suárez, R. Montes-de-Oca and F. Salem-Silva: "Conditions for the uniqueness of optimal policies of discounted Markov decision processes". Math. Methods. Oper. Res., Vol. 60, 415-436, 2004.

[17] D. Cruz-Suárez and R. Montes-de-Oca: "Uniform convergence of the value iteration policies for discounted Markov decision processes". Bol. Soc. Mat. Mexicana, Vol. 12, 133-152, 2006.

[18] H. Cruz-Suárez and R. Montes-de-Oca R: "Discounted Markov control processes induced by deterministic systems". Kybernetika, Vol. 42, 647-664, (2006).

[19] H. Cruz-Suárez and R. Montes-de-Oca: "An envelope theorem and some applications to discounted Markov decision processes". Math. Methods Oper. Res., vol. 67, 299-321, 2008.

[20] A. Dragut: "Structured optimal policies for Markov decision processes: lattice programming techniques", Wiley Encyclopedia of Operations Research and Management Science, edited by James J. Cochran Copyright © John Wiley & Sons, Inc, 2010.

[21] I. Ekeland: "On the variational principle". J. Math. Anal. Appl., Vol. 67, 324-353, 1974.

[22] E. A. Feinberg: "An ε-optimal control of a finite Markov chain with an average reward criterion". Theor. Probab. Appl., Vol. 25, No. 1, 70-81, 1980.

[23] J. Hadamard: "Sur les problemes aux derivees partielles et leur signification Physique". Princeton University Bulletin, 1902.

[24] M. Harris: "Dynamic Economic Analysis". New York: Oxford University Press, 1987.

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

[26] Hernández-Lerma O. and Lasserre J. B., "Further Topics on Discrete-Time Markov Control Processes". Springer-Verlag, 1999.

[27] O. Hernández-Lerma and W. J. Runggaldier: "Monotone approximations for convex stochastic control problems". J. Math. Systems Estim. Control, Vol. 4, No. 1, 99-14, 1994.

[28] C. J. Himmelberg, T. Parthasarathy and F. S. Vanleck: "Optimal plans for dynamic programming problems". Math. Oper. Res., Vol. 1, No. 4, 390-394, 1976.

[29] S. C. Hyeong, C. F. Michael, H. Jiaqiao and I. M. Steven: "Simulation-based Algorithms for Markov Decision Processes". Springer. First Edition, 2007.

[30] A. Jaśkiewicz and A. S. Nowak: "Discounted dynamic programming with unbounded returns: application to economic models". J. Math. Anal. Appl., Vol. 1, 1-13, 2010.

[31] A. Jourania, L. Thibaultb, D. Zagrodnyc: "Differential properties of the Moreau envelope". J. Funct. Anal., Vol. 266, 1185-1237. 2014.

[32] D. M. Kreps and E. L. Porteus: "On the optimality of structured policies in countable stage decision processes. II: positive and negative problems". SIAM J. Appl. Math., Vol. 32, No. 2, 457-466, 1977.

[33] Lai-Jiu Lin and Wei-Shih Du: "Ekeland's variational principle, minimax theorems and existence of nonconvex equilibria in complete metric spaces". J. Math. Anal. Appl., vol. 323, 360-370, 2006.

[34] R. Lucchetti: "Convexity and Well-Posed Problems". CMS Books in Mathematics, Springer, New York 2006.

[35] J. E. Marsden: "Elementary Classical Analysis". WH Freedman, Second Edition, San Francisco, 1993.

[36] H. Moreira and W. Maldonado: "A contractive method for computing the stationary solution of the Euler equation". Economics Bulletin, Vol. 3, No. 1, 1-14, 2003.

[37] R. Montes-de-Oca and E. Lemus-Rodríguez: "An unbounded Berge's minimum theorem with applications to discounted Markov decision processes". Kybernetika (Prague) Vol. 48, No. 2, 268-286, 2012.

[38] R. Montes-de-Oca, E. Lemus-Rodríguez, and F. Salem-Silva: "Non-uniqueness versus uniqueness of optimal policies in convex discounted Markov decision processes". J. Appl. Math. 2013, Art. ID 271279, 5 pages, doi:10.1155/2013/271279.

[39] R. Montes-de-Oca and F. Salem-Silva: "Estimates for perturbations of average Markov decision processes with a minimal state and upper bounded by stochastically ordered Markov chains ". Kybernetika, Vol. 41, Number 5, 757-772, 2005.

[40] R. I. Ortega-Gutiérrez, R. Montes-de-Oca and E. Lemus-Rodríguez: "Uniqueness of optimal policies as a generic property of discounted Markov decision processes: Ekeland's variational principle approach". Kybernetika, Vol. 52, No. 1, 66-75, 2016.

[41] R. I. Ortega Gutiérrez: "Teorema de la medida producto en espacios de probabilidad". Tesis de licenciatura, BUAP, 2008.

[42] R. I. Ortega Gutiérrez: "La ecuación de Euler para la solución de juegos estocásticos descontados". Tesis de maestría, BUAP, 2011.

[43] E. L. Porteus: "On the optimality of structured policies in countable stage decision processes". Manage. Csi., Vol. 22, No. 2, 148-157. 1975.

[44] E. L. Porteus: "Conditions for characterizing the structure of optimal strategies in infinite-horizon dynamic programs". J. Optim. Theory Appl., Vol. 36, No. 3, 419-432, 1982.

[45] M. L. Puterman M. L., "Markov Decision Processes: Discrete Stochastic Dynamic Programming". Wiley, 1994.

[46] U. Rieder: "Structural results for partial observed control models". Methods and Models of Operations Research, Vol. 35, 473-490. 1991.

[47] R. T. Rockafellar: "Convex Analysis". Princeton University Press, 1970.

[48] R. T. Rockafellar and R. J.B. Wets: "Variational Analysis". Springer, New York 2004.

[49] H. L. Royden: "Real Analysis". Macmillan, New York 1998.

[50] W. Rudin: "Principios de Análisis Matemático". McGraw Hill, Tercera Edición, 1980.

[51] W. J. Runggaldier: "On the construction of ε-optimal strategies in partially observed MDPs". Ann. Oper. Res., Vol. 28, 81-96, 1991.

[52] M. Schäl: "Conditions for optimal in dynamic programmming and for the limit of n-stage optimal policies to be optimal". Springer-Verlag, Vol. 32, 179-196, 1975.

[53] M. Spivak: "Calculus". Editorial Reverté, Segunda Edición, 2006.

[54] K. Tanaka, M. Hosino and D. Kuroiwa: "On an ε-optimal policy of discrete time stochastic control processes". Bull. Inform. Cybernet. Vol. 27, 107-119, 1995.

[55] K. Tanaka: "On perturbation of dynamic programming". Nonlinear and convex analysis in economic theory, Springer-Verlag, vol. 419. 275-287, 1995.