Universidad Michoacana
de San Nicolás de Hidalgo

INSTITUTO DE FÍSICA Y MATEMÁTICAS

Optimalidad Ergódica en el Modelo de Mitra-Wan y en Juegos Markovianos

T E S I S

QUE PARA OPTAR POR EL GRADO DE:
DOCTOR EN CIENCIAS

PRESENTA:

LEONARDO RAMIRO LAURA GUARACHI

DIRECTOR DE LA TESIS:

DR. ONÉSIMO HERNÁNDEZ LERMA

Morelia Michoacán; octubre de 2015.

Agradecimiento

Quiero expresar mi profundo agradecimiento al Dr. Onésimo Hernandez Lerma por haberme aceptado como su estudiante y haberme brindado su apoyo durante mi estancia de investigación en el Departamento de Matemáticas del CINVESTAV.

Guardaré en mi memoria, con especial cariño, a la gran familia que he encontrado en México: profesores, compañeros, amigos, ... Sin lugar a dudas, hicieron más amena esta experiencia.

Finalmente, este trabajo no hubiera sido posible sin el financiamiento que el pueblo mexicano me ha otorgado a través del CONACyT.

Para Manuel, Justina,
Rocío, Elliot.

Índice general

Resumen

Uno de los modelos más importantes en el área de la administración sostenible de recursos forestales, que continúa siendo una referencia para la planificación de políticas de administración, ha sido introducido por los autores Mitra y Wan durante la década de los años 80. Bajo ciertas condiciones, este modelo ha permitido conocer algunas de las propiedades cualitativas que una política óptima (no estacionaria) tiene. Así mismo, asegura la existencia de una política estacionaria óptima —que se entiende como el máximo estado sostenible— al que convergen las políticas no estacionarias óptimas. En este trabajo el modelo de Mitra-Wan se estudia como un problema de control óptimo a tiempo discreto y horizonte infinito. Para ello consideraremos algunos de los criterios de optimalidad más significativos: *optimalidad en promedio*, *optimalidad buena*, *optimalidad rebasante* y *optimalidad en sesgo*.

Por otra parte, Shapley introdujo un primer modelo de juegos estocásticos en [70]. En este tipo de juegos existe una única función de pago; el primer jugador trata de maximizar dicha función, en cambio el segundo jugador trata de minimizar. En posteriores investigaciones, se ha estudiando extensivamente las propiedades y caracterizaciones del criterio de optimalidad promedio esperado. Sin embargo, existen pocos tratamientos sobre el resto de los criterios. En nuestro estudio consideramos un modelo general de juegos markovianos de suma cero y de dos jugadores. Para estudiar las estrategias óptimas de los jugadores, extenderemos los criterios de optimalidad mencionados anteriormente y, adicionalmente, introduciremos la *optimalidad promedio F-fuerte* y el criterios de *optimalidad n-descontada*.

Nuestro aporte principal, en ambos modelos, es la descripción completa de las relaciones que existen entre los criterios que estamos considerando. Además, mostramos ciertas propiedades cualitativas que las políticas óptimas poseen, así mismo, proporcionamos algunos ejemplos que ilustran las conclusiones principales.

Palabras clave: sistemas a tiempo discreto; horizonte infinito; modelo de Mitra-Wan; juegos estocásticos; optimalidad ergódica.

Abstract

One of the most important models in the area of sustainable forest resource management, which is still a guiding principle for the policy makers, was introduced by Mitra and Wan in the 1980s. Under certain conditions, this model has revealed some qualitative properties that non-stationary optimal policies have. It also ensures the existence of an optimal stationary policy —which is understood as the maximum sustainable state— where all the optimal non-stationary policies converge. In this work the Mitra-Wan forestry model is studied as an discrete-time optimal control problem in infinite horizon. To this end, we consider some of the most significant optimality criteria: *long-run average (or ergodic) optimality, good optimality, overtaking optimality* and *bias optimality*.

On the other hand, Shapley studied a two-person stochastic zero-sum game in his seminal work [70]. In this class of games there is a single payoff function which one of the players wishes to maximize whereas the other player whishes to minimize. In subsequent studies, the properties and characterization of the expected average optimality criterion were extensively developed. However, there are just a few treatments on the other optimality criteria. In our study, we consider a general model of two-person zero-sum Markov game and investigate the optimal strategies for the players. For that porpuse, we will extend the optimality criteria aforementioned and, additionally, we introduce *average F-strong optimality* and *optimality n-discounted* criteria.

Our main contribution, in both models, is the complete description of the relations between the optimality criteria that we are considering in each model. Moreover, we also show some asymptotic properties for the optimal policies, as well as some examples that illustrate the main results.

Introducción

El interés por comprender de mejor manera los problemas de optimización ha ocupado permanentemente la mente humana. Por poner algunos puntos de referencia en la línea del tiempo, en 1638, Galileo estudió —entre sus últimos trabajos de esa época— la forma geométrica de una cadena suspendida en sus extremos. En 1662, Fermat abordó los problemas de minimización en la óptica. Alrededor de 30 años después, un análisis riguroso de este tipo de problemas se ha iniciado con el nacimiento del cálculo variacional; esto es, desde las primeras investigaciones sobre curvas y superficies minimizantes realizadas por Newton, Leibniz, J. Bernoulli, Euler, Lagrange, entre otros.

Las primeras aplicaciones del cálculo variacional han sido en el área de la física, específicamente lo relacionado con el principio de mínima acción. Entre tanto, los primeros trabajos con aplicaciones en la economía han surgido entre los años 1920 y 1930, con algunos de sus precursores como Ross, Evans, Hotelling y Ramsey.

Posteriormente, en los años 1950, L. S. Pontryagin y sus colaboradores revolucionaron esta área con una nueva teoría que generaliza al cálculo variacional: la teoría de control óptimo y el principio del máximo, que es un conjunto de condiciones necesarias para que una función de control sea óptima. Simultáneamente, otro de los logros importantes fue alcanzado por Richard Bellman al desarrollar un nueva técnica de optimización, la programación dinámica. Este nuevo método amplía el campo de aplicaciones a problemas que involucran incertidumbre.

Actualmente en la literatura existe una gran variedad de modelos dinámicos de optimización. Aquellos sistemas en donde únicamente un individuo influye sobre un parámetro del sistema (problemas de control óptimo) y los sistemas en donde varios agentes (jugadores) influyen de manera independiente (juegos dinámicos). En cada uno de estos modelos los participantes tienen el objetivo de minimizar o maximizar, según sea el caso, una función objetivo. A la vez, estos modelos tienen muchas variantes: pueden ser a tiempo discreto o continuo; con horizonte finito o infinito; determinístico o estocástico; con descuento o sin descuento.

En este trabajo nos ocuparemos de dos modelos a tiempo discreto y horizonte infinito. Una primera parte está dedicada al estudio del modelo de administración forestal Mitra-Wan como un problema de control óptimo [47]. En la segunda parte estudiaremos un modelo general de un juego markoviano de suma cero.

Los modelos de control óptimo a tiempo discreto y horizonte infinito surgieron en los trabajos de Ramsey [68] y Hotelling [34]. Estos modelos se presentan de manera natural en problemas de administración sostenible de recursos naturales. Por ejemplo, explotación forestal y pesquera [49, 63]. En este tipo de problemas es muy razonable que el tiempo sea discreto y el horizonte infinito. El modelo de administración forestal que estudiaremos, introducido por Mitra y Wan [56, 57], es uno de los más representativos de esta área. Recientes trabajos sobre este modelo son, por ejemplo, [42, 43, 45]. Por otra parte, el modelo de Mitra-Wan ha sido planteado a tiempo continuo por medio de la ecuación del transporte [16], en el cual se ha mostrado tener propiedades cualitativas análogas al caso discreto que estudiaremos [47].

Por otra parte, los juegos markovianos de suma cero tienen su origen en el trabajo de L. S. Shapley [70]. Estos juegos han sido ampliamente estudiados bajo el supuesto de que el espacio de estados es finito y los jugadores disponen cada uno de una cantidad finita de acciones. Sin embargo, diversos modelos no cumplen con esa condición. De ahí nace la necesidad de estudiar juegos más generales. Algunos trabajos que abordan este tipo de juegos son [51, 20, 30, 61, 35, 36, 37]

Para estudiar estos dos modelos, arriba mencionados, consideraremos algunos de los criterios de optimalidad más significativos: *optimalidad buena*, *optimalidad en promedio*, *optimalidad rebasante* y *optimalidad en sesgo*. En los capítulos correspondientes a juegos markovianos, consideraremos adicionalmente, entre otros, los criterios de *optimalidad n-descontada*. Nuestro aporte principal, en ambos modelos, consiste en dar una descripción completa de las relaciones que existen entre los criterios de optimalidad que estamos considerando. Adicionalmente, proporcionamos algunas de sus propiedades cualitativas.

La mayoría de estos criterios de optimalidad que consideraremos han sido introducidos originalmente en problemas de crecimiento económico o acumulación de capital y se conocen como criterios de optimalidad ergódica. Cuando se inició el estudio de la "optimalidad promedio", a fines de los años 1950s y principios de la década de los años 1960s, los investigadores hablaban de "convergencia de promedios" así que de manera natural asociaban sus trabajos a la teoría ergódica clásica. Por tal motivo usaban mucho el término "optimalidad ergódica". Ahora bien, uno de los primeros autores en estudiar el criterio de optimalidad buena fue Gale [19]. Por su parte, el criterio de

optimalidad en promedio se estudia (aproximadamente) desde el trabajo de Bellmann [5]. Mientras que el criterio de optimalidad rebasante fue introducido por Gale [19] y von Weizsäcker [79]. En tanto la optimalidad en sesgo, por Veinott [74].

Existen diversos estudios preliminares sobre estos criterios de optimalidad. Para problemas de control óptimo determinístico ver, por ejemplo, [11, 59, 80, 81, 8]. Estos temas se han estudiado de manera más extensa para procesos de control markovianos (PCMs), entre otros, ver [32, 5, 6, 12, 23, 28, 48, 66, 33, 64]. En cambio, para juegos markovianos existe una cantidad muy limitada de trabajos y se han estudiado sólamente algunos de los criterios, ver [30, 20, 36, 37, 38, 39].

Lo que sigue de este trabajo está ordenado del siguiente modo. En el Capítulo 1 estudiaremos el modelo de administración forestal de Mitra-Wan. Una descripción de la estructura de un juego markoviano y sus propiedades se realiza en el Capítulo 2. El Capítulo 3 reúne las propiedades del criterio de optimalidad promedio y sus equivalentes, mientras que en el Capítulo 4 se estudian los criterios sensibles a tiempo finito (optimalidad rebasante, optimalidad en sesgo, optimalidad 0-descontada). Las conclusiones y perspectivas de investigación a futuro se encuentran en el Capítulo 5. Finalmente, el Apéndice A reúne algunas de las propiedades básicas de cadenas de Markov.

Capítulo 1

El modelo forestal de Mitra-Wan

El estudio de la administración de recursos forestales se inició alrededor de los años 1700. Los primeros trabajos de investigaciones que se realizaron tenían como objetivo determinar la *edad óptima de rotación*. Esto es, dada una especie de árboles, determinar la edad óptima a la que se debe cosechar, y establecer un plan sostenible de cosechas. Como es natural, existen diferentes puntos de vista sobre lo que una "política optima"podría significar. Sin embargo, como menciona Samuelson en su análisis [69], una de las soluciones más razonables al problema de la edad de rotación es el dado por Faustman, Presler y Ohlin. Actualmente estos conceptos siguen siendo una referencia para la administración de recursos forestales. Una amplia descripción sobre la edad de rotación se puede ver en Amacher et al. [1].

Partiendo del análisis realizado por Faustman, los autores Mitra y Wan [56, 57] formularon el problema de la administración forestal como un modelo de optimización en donde la función objetivo es la suma de utilidades de los diferentes periodos. Para este análisis usaron los criterios de optimalidad introducidos por Gale [19] y von Weizsäcker [79]. En dicho análisis realizaron una descripción de las propiedades cualitativas de las políticas óptimas. Demostraron que existe un estado estacionario óptimo al que todas las soluciones del problema de optimización convergen. Estos resultados se conocen como teoremas de tipo *autopista* (del inglés *turnpike*). Bajo ciertas condiciones, el estado estacionario óptimo es único y se interpreta como el máximo estado sostenible, ver Kant y Berry [40].

Trabajos recientes sobre el modelo de Mitra-Wan son, por ejemplo, [42, 43, 45]. Extendido a tiempo continuo en [16]. En estos trabajos se estudian bajo condiciones menos restrictivas sobre la función de utilidad [47]. Se estudian la estructura de las soluciones considerando diferentes criterios de optimalidad;

por ejemplo, políticas buenas y políticas rebasantes de Gale [19], optimalidad débil de Brock [9] y las políticas con mínima pérdida de valor.

En este capítulo estudiaremos el modelo de Mitra-Wan como un problema de control óptimo. Consideraremos los criterios de optimalidad ergódicos: optimalidad en promedio, políticas buenas, optimalidad en sesgo, optimalidad rebasante. Ademas de estudiar las propiedades que cada uno de ellos tiene, daremos un mapa completo de las relaciones que existen entre ellos.

1.1. El modelo

Sea $\mathbb{N}_0$ el conjunto de los números enteros no negativos, $\mathbb{R}^n_+$ el conjunto de vectores de $\mathbb{R}^n$ con coordenadas no negativas y, finalmente, sea el simplex

$$\Delta := \{(x_1, ..., x_n) \in \mathbb{R}^n_+ : x_1 + ... + x_n = 1\}.$$

El modelo forestal de Mitra–Wan considera una unidad de área cubierta de manera uniforme por árboles de cierta especie cuyas edades varían de 1 a n años. Supondremos que los árboles después de n años pierden su valor económico o mueren. Sea $t \in \mathbb{N}_0$ un periodo de tiempo. Denotemos por $x_i(t)$ la proporción de la superficie ocupada por árboles de edad $i = 1, ..., n$ en el periodo t. Esto es,

$$x_1(t) + x_2(t) + ... + x_n(t) = 1.$$

Vemos así que la configuración forestal en el periodo t se puede representar por un vector $x(t) = (x_1(t), x_2(t), ..., x_n(t))$ en el simplex Δ

Ahora lo que sigue es representar por otro vector un plan o política de cosecha en cada periodo de tiempo. Al finalizar el periodo t cosecharemos árboles de cada edad. Denotemos por $u_i(t)$ el área que se libera al cosechar árboles de edad $i = 1, ..., n$. El plan de cosecha en el periodo t será el vector

$$u(t) = (u_1(t), u_2(t), ..., u_n(t)), \quad 0 \leq u_i(t) \leq x_i(t). \tag{1.1}$$

Es importante remarcar que podemos suponer que $u_n(t) = x_n(t)$ para todo $t \in \mathbb{N}_0$ ya que los árboles de más de n años pierden su valor económico.

Ahora veamos cómo cambia la configuración forestal de un periodo a otro. El área que se libera al finalizar el periodo t se usará al inicio del periodo $t + 1$ para plantar nuevos árboles de un año. En resumen, el área ocupada por árboles, según su edad, en el periodo $t + 1$ es

$$x_1(t+1) = \sum_{i=1}^{n} u_i(t) \quad \text{y} \quad x_i(t+1) = x_{i-1}(t) - u_{i-1}(t), \quad 2 \leq i \leq n.$$

Es decir, la configuración forestal en el periodo $t+1$ es

$$x(t+1) = \left(\sum_{i=1}^{n} u_i(t), x_1(t) - u_1(t), ..., x_{n-1}(t) - u_{n-1}(t) \right),$$

o en forma matricial:

$$x(t+1) = \begin{pmatrix} 0 & 0 & . & 0 & 0 \\ 1 & 0 & . & 0 & 0 \\ 0 & 1 & . & 0 & 0 \\ . & . & . & . & . \\ 0 & 0 & . & 1 & 0 \end{pmatrix} x(t) + \begin{pmatrix} 1 & 1 & . & 1 & 1 \\ -1 & 0 & . & 0 & 0 \\ 0 & -1 & . & 0 & 0 \\ . & . & . & . & . \\ 0 & 0 & . & -1 & 0 \end{pmatrix} u(t). \quad (1.2)$$

Este es un sistema de control lineal a tiempo discreto con espacio de estados Δ, espacio de control $U := [0,1]^n$ y, tomando en cuenta (1.1), el conjunto de controles admisibles en el estado $x \in \Delta$ es dado por

$$U(x) := [0, x_1] \times [0, x_2] \times \cdots \times [0, x_{n-1}] \times \{x_n\}.$$

Adicionalmente, como es usual, definiremos el conjunto de parejas admisibles estado-control como $\mathbb{K} := \{(x,u) : x \in \Delta, u \in U(x)\}$ y la función del sistema como $f : \mathbb{K} \to \Delta$ dada por $f(x,u) := Ax + Bu$, donde A y B son las matrices de la ecuación (1.2). Por tanto, el sistema de control que estudiaremos es de la forma:

$$x(t+1) = f(x(t), u(t)), \quad t \in \mathbb{N}_0, \quad (1.3)$$

donde $x(0) = x \in \Delta$ es una condición inicial dada.

Dada una condición inicial $x(0) = x \in \Delta$, una *política de control* para este estado inicial es una sucesión de controles $\{u(t)\}_{t=0}^{\infty}$ tal que $u(t) \in U(x(t))$ para todo $t \in \mathbb{N}_0$. El conjunto de políticas de control en este estado se denotado por $\mathcal{U}(x)$ y, para simplificar la notación, un elemento de $\mathcal{U}(x)$ lo escribiremos simplemente como $\mathbf{u}$. Por otra parte, dado un estado inicial $x \in \Delta$, el conjunto de todos los estados al que es posible llegar en un tiempo $T > 0$ se conoce como *órbita positiva* y se denota por $\mathcal{O}^+(x,T)$, esto es

$$\mathcal{O}^+(x,T) = \{y \in \Delta : \text{existen } u(0), ..., u(T-1) \text{ tal que } x(0) = x \text{ y } x(T) = y\}.$$

Análogamente, el conjunto de los estados desde donde se puede llegar al estado $x \in \Delta$ en un tiempo $T > 0$ se conoce como *órbita negativa* y se denota por $\mathcal{O}^-(x,T)$, es decir

$$\mathcal{O}^-(x,T) = \{y \in \Delta : \text{existen } u(0), ..., u(T-1) \text{ tal que } x(0) = y \text{ y } x(T) = x\}.$$

La siguiente proposición, adaptado a nuestro contexto, proviene de [58] y [44].

Proposición 1.1.1. *Para cada estado* $x \in \Delta$, $\mathcal{O}^+(x,n) = \mathcal{O}^-(x,n) = \Delta$.

En el siguiente lema se muestra algunas de las propiedades algebraicas que posee el sistema (1.2).

Lema 1.1.2. *Sea* A' *la transpuesta de la matriz* A. *Para cada* $(x,u) \in \mathbb{K}$ *tenemos:*

(i) $(1,...,n) \cdot (I - A')x = 1$,

(ii) $A'f(x,u) = x - u$ y

(iii) $(1,...,n) \cdot (u - x + f(x,u)) = 1$.

Demostración. Estas propiedades se siguen por cálculo directo. □

Sea $(x,u) \in \mathbb{K}$. Se dice que x es un *estado estacionario* y u un *control estacionario* si $f(x,u) = x$. Al conjunto de todas esas parejas lo denotaremos por

$$F := \{(x,u) \in \mathbb{K} : f(x,u) = x\}.$$

El conjunto de los estados estacionarios se denotará por F_Δ, mientras que el conjunto de los controles estacionarios por F_U, esto es

$$F_\Delta := \{x \in \Delta : \text{ existe } u \in U(x) \text{ tal que } (x,u) \in F\}$$

y

$$F_U := \{u \in U : \text{ existe } x \in \Delta \text{ tal que } (x,u) \in F\}.$$

En el siguiente teorema presentamos una descripción explícita de los dos conjuntos que acabamos de definir.

Teorema 1.1.3. *Se satisface lo siguiente:*

(i) $F_\Delta = \{x \in \Delta : x_1 \geq ... \geq x_n\}$,

(ii) $F_U = \{u \in U : (1,2,...,n) \cdot u = 1\}$.

Demostración. Notemos que $I - A'$ es una matriz invertible y su inversa es

$$(I - A')^{-1} = \begin{pmatrix} 1 & 1 & 1 & ... & 1 \\ 0 & 1 & 1 & ... & 1 \\ . & . & . & ... & . \\ 0 & 0 & 0 & ... & 1 \end{pmatrix}. \tag{1.4}$$

(i). Consideremos $x \in F_\Delta$, entonces existe $u \in U(x)$ tal que $f(x,u) = x$. Del Lema 1.1.2(ii), tenemos que $u = (I - A')x = (x_1 - x_2, ..., x_{n-1} - x_n, x_n) \in U$.

En consecuencia $x_1 \geq ... \geq x_n$. Recíprocamente, dado $x = (x_1, ..., x_n) \in \Delta$ con $x_1 \geq ... \geq x_n$, el vector $u = (I - A')x$ es un control admisible para el estado x y satisface $f(x, u) = x$.

(ii). Dado un control $u \in F_U$, existe un estado $x \in \Delta$ para el cual $f(x, u) = x$. Entonces, por el lema 1.1.2(iii), concluimos que $(1, ..., n) \cdot u = 1$. Recíprocamente, cada control $u \in U$ que satisface $(1, ..., n) \cdot u = 1$ es admisible para el estado $x = (I - A')^{-1}u$ y cumplen la igualdad $f(x, u) = x$. □

Notemos que para un estado estacionario $x \in F_\Delta$ existe un único control estacionario $u = (I - A')x$. Recíprocamente, para cada control estacionario $u \in F_U$ existe un único estado estacionario $x = (I - A')^{-1}u$. Esto demuestra la siguiente caracterización.

Corolario 1.1.4. *El conjunto de estados estacionarios y controles estacionarios queda determinado por* $F = \{(x, u) \in F_\Delta \times F_U : u = (I - A')x\}$.

1.2. El problema de control óptimo

Después de que ya tenemos el modelo planteado y hemos estudiado algunas de sus propiedades, ahora nos ocuparemos del problema de optimización. Supongamos que la producción de madera está determinada por el *vector biomasa* $\xi = (\xi_1, \xi_2, ..., \xi_n) \in \mathbb{R}^n_+$, donde ξ_i representa la producción de los árboles de i-años en una unidad de área . Por tanto, la cantidad de madera colectada en el periodo t está determinada por $\xi \cdot u(t)$. Ahora consideramos una función de beneficio (o de ganancia) $r : \mathbb{K} \to [0, \infty)$ definida por $r(x, u) := \theta(\xi \cdot u)$, donde $\theta : [0, \infty) \to [0, \infty)$ es una función continua, estrictamente cóncava y no decreciente. Dada una política admisible $\{u(t)\}_{t=0}^{\infty}$, uno de nuestros objetivos es estudiar el comportamiento de la serie

$$\sum_{t=0}^{\infty} r(x(t), u(t)).$$

Dependiendo de la política que estemos considerando, esta serie podría divergir. En ese caso será necesario introducir criterios de optimalidad que permitan distinguir entre esas políticas. Primero definiremos un concepto de optimalidad en el conjunto de controles estacionarios.

Definición 1.2.1. *Sea* $(\bar{x}, \bar{u}) \in F$. *Diremos que* $\bar{x}$ *es un* estado estacionario óptimo *y* $\bar{u}$ *un* control estacionario óptimo *si*

$$r(\bar{x}, \bar{u}) = \max\{r(x, u) : (x, u) \in F\}.$$

En la literatura sobre modelos de crecimiento económico óptimo, un estado estacionario óptimo se conoce como una *regla de oro*. Este concepto ha sido estudiado inicialmente por autores como John von Neumann, Maurice Allais y Edmund Phelps. Para mayores referencias ver [65].

En el resto de esta sección demostraremos que, bajo algunas condiciones usuales, existe un único control estacionario óptimo (Teorema 1.2.4). Adicionalmente, caracterizamos dicho control (Proposición 1.2.7).

El tiempo de vida (o periodo útil) de algunas especies de árboles puede dividirse en tres etapas de acuerdo al promedio de producción por edad ξ_i/i. En la primera etapa el promedio de producción crece, en la segunda etapa alcanza su máximo y en la última etapa decrece. En nuestro modelo, con algunos ajustes en el tiempo si es necesario, supondremos que los árboles que estamos considerando satisfacen esa condición.

Supuesto 1.2.2. Existe un único $k \in \{1, 2, ..., n\}$ tal que

$$\frac{\xi_k}{k} = \max\{\frac{\xi_i}{i} : i = 1, ..., n\}.$$

Esta propiedad es un supuesto estándar en la literatura, ver por ejemplo [9, 42, 43, 45, 57]. Una de las principales consecuencias del supuesto es la unicidad del estado estacionario óptimo (Teorema 1.2.4). Caso contrario, si el supuesto no se cumple, el sistema podría tener más de un estado estacionario óptimo. Uno de los trabajos relacionados con este tema es, por ejemplo, [52].

Lema 1.2.3. *Para cada $u \in U$, se satisface la desigualdad*

$$\xi \cdot u \leq (\xi_k/k)(1, 2, ..., n) \cdot u.$$

Con igualdad si y sólo si $u_i = 0$ para cada $i \neq k$.

Demostración. Nótese que

$$\frac{\xi_k}{k}(1, 2, ..., n) \cdot u - \xi \cdot u = \sum_{i=1}^{n}(\frac{\xi_k}{k}i - \xi_i)u_i.$$

Por el Supuesto 1.2.2, cada término es no negativo y la suma es cero si y sólo si $u_i = 0$ para cada $i \neq k$. □

El siguiente teorema es un resultado análogo a la unicidad de la regla de oro en [57].

Teorema 1.2.4. *El control $\bar{u} = (0, ..., 0, 1/k, 0, ..., 0)$, donde $1/k$ es la $k-$ésima coordenada, es el único control estacionario óptimo. Más aun, el único estado estacionario óptimo asociado a $\bar{u}$ es $\bar{x} = (1/k, ..., 1/k, 0, ..., 0)$, donde las primeras k-coordenadas son todos iguales a $1/k$.*

Demostración. Por el Lema 1.2.3 y el Teorema 1.1.3, $\xi \cdot u \leq \xi_k/k$ para cada $u \in F_U$. Notemos que $\bar{u} \in F_U$ y es el único control estacionario que satisface la igualdad $\xi \cdot \bar{u} = \xi_k/k$. Ya que θ es estríctamente cóncava y no decreciente, concluimos que $\bar{u}$ es el único control estacionario óptimo. El Corolario 1.1.4 y (1.4) nos dan el estado estacionario óptimo $\bar{x} = (1/k, ..., 1/k, 0, ..., 0)$. □

Nota 1.2.5. Dado $\tau_0 \geq 0$, puesto que θ es estrictamente cóncava, existe $\lambda = \lambda(\tau_0) > 0$ tal que $\theta(\tau) - \theta(\tau_0) \leq \lambda(\tau - \tau_0)$ para cada $\tau \geq 0$, con igualdad sólo para $\tau = \tau_0$.

Proposición 1.2.6. *Sea $(\bar{x}, \bar{u}) \in F$ como en la Definición 1.2.1. Existe $p \in \mathbb{R}^n_+$ tal que*

$$r(x,u) - r(\bar{x},\bar{u}) \leq p \cdot [x - f(x,u)] \quad \textit{para cada} \quad (x,u) \in \mathbb{K}. \tag{1.5}$$

Demostración. Del Lema 1.1.2 deducimos la igualdad

$$(1, 2, ..., n) \cdot u = 1 + (1, 2, ..., n) \cdot (x - f(x,u)). \tag{1.6}$$

Usando la desigualdad del Lema 1.2.3 concluimos que

$$\begin{aligned} \xi \cdot u &\leq (\xi_k/k)(1, 2, ..., n) \cdot u \\ &= (\xi_k/k)\left[1 + (1, 2, ..., n) \cdot (x - f(x,u))\right] \\ &= (\xi_k/k) + (\xi_k/k)(1, 2, ..., n) \cdot (x - f(x,u)). \end{aligned}$$

Por tanto

$$\xi \cdot u - \xi \cdot \bar{u} \leq \frac{\xi_k}{k}(1, 2, ..., n) \cdot (x - f(x,u)). \tag{1.7}$$

Adicionalmente, tomando $\tau = \xi \cdot u$ y $\tau_0 = \xi \cdot \bar{u} = \xi_k/k$ en la Nota 1.2.5, obtenemos la desigualdad

$$r(x,u) - r(\bar{x},\bar{u}) \leq \lambda(\xi \cdot u - \xi \cdot \bar{u}). \tag{1.8}$$

Finalmente, de (1.7) y (1.8), concluimos que $r(x,u) - r(\bar{x},\bar{u}) \leq p \cdot (x - f(x,u))$ donde $p := \lambda(\xi_k/k)(1, 2, ..., n) \in \mathbb{R}^n_+$. □

Sea p como en la Proposición 1.2.6. Siguiendo la notación de McKenzie [53], definiremos la *función pérdida de valor* $\delta : \mathbb{K} \to \mathbb{R}$ como

$$\delta(x,u) := r(\bar{x},\bar{u}) - r(x,u) + p \cdot (x - f(x,u)). \tag{1.9}$$

De la Proposición 1.2.6, $\delta(x,u) \geq 0$ para todo $(x,u) \in \mathbb{K}$.

Según los autores Diehl et al. [13], la función (1.9) es conocida como la *la función de fase rotada*, y la desigualdad (1.5) como *dualidad fuerte del problema del estado estable.*

Proposición 1.2.7. *Sea $x \in \Delta$ un estado arbitrario. Si $\delta(x,u) = 0$, entonces $u = \bar{u}$.*

Demostración. De (1.6), (1.8) y el Lema 1.2.3,

$$\begin{aligned}
\delta(x,u) &= r(\bar{x},\bar{u}) - r(x,u) + p \cdot (x - f(x,u)) \\
&\geq -\lambda(\xi \cdot u - \xi \cdot \bar{u}) + p \cdot (x - f(x,u)) \\
&= -\lambda(\xi \cdot u - \xi_k/k) + \lambda(\xi_k/k)(1,2,...,n) \cdot (x - f(x,u)) \\
&= -\lambda(\xi \cdot u - \xi_k/k) + \lambda(\xi_k/k)[(1,2,...,n) \cdot u - 1] \\
&= \lambda[(\xi_k/k)(1,2,...,n) - \xi] \cdot u \\
&\geq 0.
\end{aligned}$$

En consecuencia, por el Lema 1.2.3, $\delta(x,u) = 0$ implica que $u_i = 0$ para todo $i = 1,2,...,n$ con excepción de $i = k$.

Por otra parte, las desigualdades (1.7) y (1.8) implican que

$$\delta(x,u) \geq \theta(\xi \cdot \bar{u}) - \theta(\xi \cdot u) + \lambda(\xi \cdot u - \xi \cdot \bar{u}) \geq 0.$$

Luego si $\delta(x,u) = 0$, entonces $\theta(\xi_k/k) - \theta(\xi_k u_k) = \lambda(\xi_k u_k - \xi \cdot \bar{u})$. Puesto que θ es estrictamente cóncava, se sigue que $u_k = 1/k$. □

1.3. Optimalidad promedio

Nuestro objetivo en esta sección es estudiar el criterio de optimalidad promedio. Este criterio es uno de los más usados para problemas de control óptimo con horizonte infinito. Buscaremos las políticas que son óptimas en promedio y las caracterizaremos en términos del control estacionario óptimo.

Sea

$$J_T(x,\mathbf{u}) := \sum_{t=0}^{T-1} r(x(t),u(t)), \tag{1.10}$$

el beneficio total acumulado hasta el periodo T usando la política de control $\mathbf{u} \in \mathcal{U}(x)$. Consideremos el correspondiente beneficio (o ganancia) promedio a largo plazo

$$J(x,\mathbf{u}) := \liminf_{T\to\infty} \frac{1}{T} J_T(x,\mathbf{u}). \tag{1.11}$$

Definición 1.3.1. *Una política de control $\mathbf{u}^* \in \mathcal{U}(x)$ se dice que es* óptimo en promedio *si $J(x,\mathbf{u}^*) = J^*(x)$, donde*

$$J^*(x) := \sup\{J(x,\mathbf{u}) : \mathbf{u} \in \mathcal{U}(x)\}.$$

El siguiente teorema muestra, en particular, que el valor de una política óptima en promedio no depende de la condición inicial.

Teorema 1.3.2. *Sea $(\bar{x}, \bar{u})$ como en la Definición 1.2.1. Entonces $J^*(x) = r(\bar{x}, \bar{u})$ para cada $x \in \Delta$.*

Demostración. Por la Proposición 1.1.1, existe una política de control $\widehat{\mathbf{u}} \in \mathcal{U}(x)$ tal que $\widehat{u}(t) = \bar{u}$ para cada $t \geq n$. La política de control $\widehat{\mathbf{u}}$ es tal que $J(x, \widehat{\mathbf{u}}) = r(\bar{x}, \bar{u})$. Por consiguiente $r(\bar{x}, \bar{u}) \leq J^*(x)$. Por otra parte, de la Proposición 1.2.6,

$$J_T(x, \mathbf{u}) - Tr(\bar{x}, \bar{u}) \leq p \cdot [x(0) - x(T)]. \tag{1.12}$$

Ya que Δ es un espacio compacto, el lado derecho de la desigualdad es acotada. Multiplicando por $1/T$ y tomando el límite $\liminf_{T\to\infty}$, obtenemos $J(x, \mathbf{u}) \leq r(\bar{x}, \bar{u})$ para cada $\mathbf{u} \in \mathcal{U}(x)$; por tanto, $J^*(x) \leq r(\bar{x}, \bar{u})$. □

La desigualdad (1.12) nos da el siguiente resultado.

Proposición 1.3.3. *Sea $\mathbf{u} \in \mathcal{U}(x)$ una política de control arbitraria. La sucesión $\{J_T(x, \mathbf{u}) - Tr(\bar{x}, \bar{u})\}_{T=1}^{\infty}$ es acotada o bien $\liminf_{T\to\infty}[J_T(x, \mathbf{u}) - Tr(\bar{x}, \bar{u})] = -\infty$.*

Una política de control $\mathbf{u} \in \mathcal{U}(x)$ para la cual la sucesión $\{J_T(x, \mathbf{u}) - Tr(\bar{x}, \bar{u})\}_{T=1}^{\infty}$ es acotada se dice que es *buena*. Originalmente esta definición fue introducida por Gale [19].

La siguiente proposición establece la relación que existe entre las políticas buenas y las óptimas en promedio.

Proposición 1.3.4. *Toda política de control buena es también óptima en promedio.*

Demostración. Si $\mathbf{u} \in \mathcal{U}(x)$ es una política de control buena, existe $K > 0$ tal que $-K \leq J_T(x, \mathbf{u}) - Tr(\bar{x}, \bar{u}) \leq K$ para cada $T \in \mathbb{N}$. Por lo tanto

$$J(x, \mathbf{u}) = \lim_{T\to\infty} (1/T) J_T(x, \mathbf{u}) = r(\bar{x}, \bar{u}).$$

□

Sea δ la función pérdida de valor (1.9). Dado $x \in \Delta$ y $\mathbf{u} \in \mathcal{U}(x)$ definimos las funciones

$$\delta_T(x, \mathbf{u}) := \sum_{t=0}^{T-1} \delta(x(t), u(t)) \quad \text{y} \quad \delta(x, \mathbf{u}) := \sum_{t=0}^{\infty} \delta(x(t), u(t)). \tag{1.13}$$

Por la definición (1.9) de $\delta(x,u)$, la sucesión $\{\delta_T(x,\mathbf{u})\}_{T=0}^{\infty}$ es creciente y satisface

$$\delta_T(x,\mathbf{u}) = -[J_T(x,\mathbf{u}) - Tr(\bar{x},\bar{u})] + p\cdot(x - x(T)) \quad \text{para todo } T \in \mathbb{N}. \tag{1.14}$$

Nótese que en (1.14), tomando en cuenta que Δ es compacto, el término $p\cdot(x-x(T))$ es acotado. Por lo tanto, $\delta_T(x,\mathbf{u})$ es acotada si y sólo si $[J_T(x,\mathbf{u}) - Tr(\bar{x},\bar{u})]$ es acotada. Esta observación nos da la siguiente proposición.

Proposición 1.3.5. *Una política de control $\mathbf{u} \in \mathcal{U}(x)$ es buena si y sólo si $\delta(x,\mathbf{u}) < \infty$.*

El siguiente teorema de tipo "turnpike" (autopista) es una versión adaptada a nuestro contexto de uno de los resultados de [57].

Teorema 1.3.6. *Sea $x \in \Delta$. Si $\mathbf{u} \in \mathcal{U}(x)$ es una política de control buena, entonces $\lim_{t\to\infty} u(t) = \bar{u}$ y $\lim_{t\to\infty} x(t) = \bar{x}$.*

Demostración. Primero introducimos una función auxiliar $\phi : \mathbb{K} \to \mathbb{R}$ definida por $\phi(x,u) := r(x,u) - p\cdot(x - f(x,u))$. Dicha función satisface la igualdad $\phi(x,u) = -\delta(x,u) + r(\bar{x},\bar{u})$. Ahora supongamos que $\mathbf{u} = \{u(t)\}_{t=0}^{\infty}$ es una política de control buena. Entonces $\delta(x(t),u(t)) \to 0$ cuando $t \to \infty$, lo cual implica que $\phi(x(t),u(t)) \to r(\bar{x},\bar{u}) = \phi(\bar{x},\bar{u})$. Si (x^*,u^*) es un punto de acumulación de $\{(x(t),u(t))\}_{t=0}^{\infty}$, esto significa que existe una infinidad de términos de la sucesión en cada entorno abierto de (x^*,u^*) y entonces $\phi(x^*,u^*) = \phi(\bar{x},\bar{u})$. Por otra parte, $\phi(x,u) \leq r(\bar{x},\bar{u})$ para cada $(x,u) \in \mathbb{K}$. Adicionalmente, ya que r es estrictamente cóncava, entonces ϕ también lo es. Sabemos que las funciones cóncavas poseen un único máximo en subconjuntos compacto: por tanto, ϕ tiene un único máximo que es $(x^*,u^*) = (\bar{x},\bar{u})$. □

1.4. Optimalidad en sesgo

En esta sección estudiaremos un nuevo criterio de optimalidad que en algún sentido es más selectivo que el criterio de optimalidad en promedio. Por ejemplo, el criterio de optimalidad en promedio no distingue dos políticas de control que tienen la misma ganancia en promedio pero que sus correspondientes pagos en tiempo finito (como en (1.10)) son totalmente diferentes, ver [27, 32, 48]. Para distinguir dichas políticas, primero buscaremos las políticas que sean óptimas en promedio, luego dentro de ellas buscaremos políticas que sean más sensibles al pago en tiempo finito.

Dado un estado inicial $x \in \Delta$ y una política de control admisible $\mathbf{u} = \{u(t)\}_{t=0}^{\infty}$. Definimos la función de *sesgo* como

$$b(x, \mathbf{u}) := \sum_{t=0}^{\infty} [r(x(t), u(t)) - r(\bar{x}, \bar{u})] .$$

El siguiente concepto fue introducido por Veinott [74] y posteriormente ha sido estudiado por varios autores en diversos contextos y modelos; ver por ejemplo [12, 27, 32, 48].

Definición 1.4.1. *Sea $x \in \Delta$ el estado inicial del sistema. Una política de control $\mathbf{u}^* \in \mathcal{U}(x)$ se dice que es* óptima en sesgo*, con respecto a x, si $b(x, \mathbf{u}^*) = b^*(x)$ donde*

$$b^*(x) := \sup\{b(x, \mathbf{u}) : \mathbf{u} \in \mathcal{U}(x)\}.$$

Lema 1.4.2. *Para cada política de control $\mathbf{u} \in \mathcal{U}(x)$, $\lim_{T\to\infty}[J_T(x, \mathbf{u}) - Tr(\bar{x}, \bar{u})] = p \cdot (x - \bar{x}) - \delta(x, \mathbf{u})$. Por tanto la función de sesgo también se puede expresar como $b(x, \mathbf{u}) = p \cdot (x - \bar{x}) - \delta(x, \mathbf{u})$.*

Demostración. La demostración se sigue de la ecuación (1.14) y el Teorema 1.3.6. □

El siguiente teorema es una consecuencia inmediata del Lema 1.4.2.

Teorema 1.4.3. *Una política de control $\mathbf{u}^* \in \mathcal{U}(x)$ es óptima en sesgo si y sólo si $\delta(x, \mathbf{u}^*) = \delta^*(x)$, donde $\delta^*(x) := \inf\{\delta(x, \mathbf{u}) : \mathbf{u} \in \mathcal{U}(x)\}$.*

Proposición 1.4.4. *Una política óptima en sesgo es también buena.*

Demostración. De la Proposición 1.1.1, existe $\hat{\mathbf{u}} \in \mathcal{U}(x)$ tal que $\hat{u}(\tau) = \bar{u}$ para cada $\tau \geq n$. Por tanto $\delta^*(x) \leq \delta(x, \hat{\mathbf{u}}) = \delta_n(x, \hat{\mathbf{u}}) < \infty$. Por lo tanto, si $\mathbf{u}$ es una política de cóntrol óptima en sesgo, entonces $\delta(x, \mathbf{u}) < \infty$, i.e. es buena. □

El siguiente resultado, análogo a la Proposición 1.4.2 de [11], garantiza la existencia de una política de control óptima en sesgo. Este resultado se obtiene por medio de argumentos usuales de continuidad y compacidad.

Proposición 1.4.5. *Para cada $x \in \Delta$ existe una política de control $\mathbf{u}^* \in \mathcal{U}(x)$ tal que $\delta(x, \mathbf{u}^*) = \delta^*(x)$.*

El siguiente criterio que adaptamos a nuestro contexto fue inicialmente introducido por Gale [19] y von Weizsäcker [79].

Definición 1.4.6. *Dado un estado inicial $x \in \Delta$, una política de control $\mathbf{u}^* \in \mathcal{U}(x)$ se dice que es* óptima rebasante *(con respecto a x) si para toda política de control $\mathbf{u} \in \mathcal{U}(x)$,*

$$\limsup_{T\to\infty}[J_T(x,\mathbf{u}) - J_T(x,\mathbf{u}^*)] \leq 0.$$

Teorema 1.4.7. *Para cada estado inicial $x \in \Delta$, las siguientes condiciones son equivalentes:*

(i) $\mathbf{u} \in \mathcal{U}(x)$ *es óptima en sesgo,*

(ii) $\mathbf{u} \in \mathcal{U}(x)$ *es óptima rebasante.*

Demostración. (i) $\Rightarrow$ (ii): Sea $\mathbf{u}^*$ una política de control óptima en sesgo y $\mathbf{u}$ cualquier otra política de control. Primero consideramos el caso $\delta(x,\mathbf{u}) < \infty$. Debido al Lema 1.4.2 y la Proposición 1.4.4, el lado derecho de la igualdad

$$J_T(x,\mathbf{u}) - J_T(x,\mathbf{u}^*) = [J_T(x,\mathbf{u}) - Tr(\bar{x},\bar{u})] - [J_T(x,\mathbf{u}^*) - Tr(\bar{x},\bar{u})]$$

es convergente. Tomando límites obtenemos

$$\limsup_{T\to\infty}[J_T(x,\mathbf{u}) - J_T(x,\mathbf{u}^*)] \leq b(x,\mathbf{u}) - b(x,\mathbf{u}^*) \leq 0,$$

lo cual nos da (ii). Ahora consideremos el caso $\delta(x,\mathbf{u}) = +\infty$. Del Lema 1.4.2 concluimos que $\lim_{T\to\infty}[J_T(x,\mathbf{u}) - J_T(x,\mathbf{u}^*)] = -\infty$ y, en consecuencia, por la Definición 1.4.6, $\mathbf{u}^*$ es óptima rebasante.

(ii) $\Rightarrow$ (i): Supongamos que $\mathbf{u}^*$ es una política óptima rebasante y sea $\mathbf{u} \in \mathcal{U}(x)$. Si $\delta(x,\mathbf{u}) = \infty$, entonces $b(x,\mathbf{u}^*) \geq b(x,\mathbf{u}) = -\infty$ y se sigue (i). En otro caso,

$$b(x,\mathbf{u}) - b(x,\mathbf{u}^*) = \lim_{T\to\infty}[J_T(x,\mathbf{u}) - J_T(x,\mathbf{u}^*)] \leq 0$$

y por tanto $\mathbf{u}^*$ es óptima en sesgo. □

1.5. Ejemplo

Para nuestro ejemplo consideraremos una unidad de área cubierta por cierta clase de árboles para la producción de madera o para algún otro beneficio. Asumiremos que dichos árboles mueren o pierden su valor económico al llegar a la edad de $n = 4$ años. La producción de madera por unidad de área de estos árboles depende de la edad y es como sigue: los árboles de un año aún no producen; los de dos años de edad producen 2 unidades; los árboles

de tres años producen 9 unidades; y los de cuatro años de edad producen 10 unidades de madera. En resumen, el vector biomasa es $\xi = (0, 2, 9, 10)$, y la producción promedio en cada periodo es 0, 1, 3 y 5/2 respectivamente. En este caso la producción promedio alcanza su valor máximo a la edad de tres años. Por lo tanto, por el Teorema 1.2.4, el estado estacionario óptimo es $\bar{x} = (1/3, 1/3, 1/3, 0)$ y el correspondiente control estacionario óptimo es $\bar{u} = (0, 0, 1/3, 0)$.

El sistemas (1.2) tiene la forma

$$x(t+1) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} x(t) + \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{pmatrix} u(t).$$

Por simplicidad, supongamos que el estado inicial es $x(0) = (1, 0, 0, 0)$. Sea $\mathbf{u} = \{u(t)\}_{t=0}^{\infty}$ una política de control o plan de cosecha. La cantidad total de madera recolectada en el periodo t es $\xi \cdot u(t)$. Considerando la función de utilidad $r(x, u) = (\xi \cdot u)^{1/2}$, la cual satisface los supuestos, la cantidad de beneficio acumulado a maximizar es la serie

$$\sum_{t=0}^{\infty} r(\xi \cdot u(t)).$$

y el beneficio en promedio a largo plazo es

$$J(x, \mathbf{u}) := \liminf_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} r(\xi \cdot u(t)).$$

Por el Teorema 1.3.2, el valor promedio óptimo es $J^*(x) := \sup_{\mathbf{u}} J(x, \mathbf{u}) = r(\bar{x}, \bar{u}) = \sqrt{3}$. Ejemplos de políticas de control buenas, por tanto óptimas en promedio, son los siguientes

1. Sea $\mathbf{u}_1 = \{u(t)\}_{t=0}^{\infty}$ una política definida como:

   $x(0) = (1, 0, 0, 0)$ $\quad u(0) = (1/3, 0, 0, 0)$ $\quad r(x(0), u(0)) = 0$

   $x(1) = (2/3, 1/3, 0, 0)$ $\quad u(1) = (1/3, 0, 0, 0)$ $\quad r(x(1), u(1)) = 0$

   $x(t) = \bar{x}$ $\quad u(t) = \bar{u}$ $\quad r(x(t), u(t)) = \sqrt{3}$

   para todo $t \geq 3$.

2. Sea $\mathbf{u}_2 = \{u(t)\}_{t=0}^{\infty}$ definida por:

$$\begin{array}{lll} x(0) = (1,0,0,0) & u(0) = (1/3,0,0,0) & r(x(0),u(0)) = 0 \\ x(1) = (1/3,2/3,0,0) & u(1) = (0,1/3,0,0) & r(x(1),u(1)) = \sqrt{2/3} \\ x(t) = \bar{x} & u(t) = \bar{u} & r(x(t),u(t)) = \sqrt{3} \end{array}$$

para todo $t \geq 3$.

Estas políticas de control, $\mathbf{u}_1$ y $\mathbf{u}_2$, son tal que $J(x,\mathbf{u}_1) = J(x,\mathbf{u}_2) = \sqrt{3}$. Entonces, por el Teorema 1.3.2, ambas son políticas de control óptimas en promedio (ver abajo la Figura 1.2). Más aun, toda política de control que llega en tiempo finito al control estacionario óptimo es óptima en promedio. Por lo tanto, no es difícil ver que hay una infinidad de políticas óptimas en promedio.

Los siguientes ejemplos son políticas de control óptimas en promedio que no son buenas. Esto demuestra que el conjunto de las políticas óptimas en promedio *contiene estrictamente* al conjunto de las políticas buenas.

3. Sea $\mathbf{u}_3 = \{u(t)\}_{t=0}^{\infty}$ definida por:

$$\begin{array}{ll} x(0) = (1,0,0,0) & u(0) = (1,0,0,0) \\ x(1) = (1,0,0,0) & u(1) = (1,0,0,0) \\ x(2) = (1,0,0,0) & u(2) = (\frac{11}{12},0,0,0) \\ x(3) = (\frac{11}{12},\frac{1}{3}-\frac{1}{4},0,0) & u(3) = (\frac{47}{60},0,0,0) \\ x(4) = (\frac{47}{60},\frac{1}{3}-\frac{1}{5},\frac{1}{3}-\frac{1}{4},0) & u(4) = (\frac{37}{60},0,\frac{1}{3}-\frac{1}{4},0). \end{array}$$

En general, los estados son $x(t) = (\frac{1}{3}+\frac{1}{t}+\frac{1}{t+1},\frac{1}{3}-\frac{1}{t+1},\frac{1}{3}-\frac{1}{t},0)$ y los controles $u(t) = (\frac{1}{t}+\frac{1}{t+1}+\frac{1}{t+2},0,\frac{1}{3}-\frac{1}{t},0)$ para todo $t \geq 4$.

Esta política de control satisface que $r(x(t),u(t)) = \sqrt{3-\frac{9}{t}}$ para todo $t \geq 4$. Más aun, $\lim_{t\to\infty} x(t) = \bar{x}$ y $\lim_{t\to\infty} u(t) = \bar{u}$. Puesto que r es una funcion continua, tenemos $\lim_{t\to\infty} r(x(t),u(t)) = r(\bar{x},\bar{u})$. Entonces, $J(x,\mathbf{u}_3) = r(\bar{x},\bar{u})$, lo cual significa que $\mathbf{u}_3$ es óptima en promedio. Por otra parte, de la Nota 1.2.5, existe $\lambda > 0$ tal que $r(x(t),u(t)) - r(\bar{x},\bar{u}) \leq \lambda(\xi\cdot u(t) - \xi\cdot\bar{u}) = -\frac{\lambda}{t}$ para todo $t \geq 4$. Por lo tanto $\sum_{t=0}^{\infty}[r(x(t),u(t)) - r(\bar{x},\bar{u})] = -\infty$, lo cual significa que $\mathbf{u}_3$ no es una política buena.

La Figura 1.1 muestra la gráfica de las funciones de pago correspondientes a los ejemplos considerados arriba.

Para concluir, notemos que la función de sesgo $b(x,\mathbf{u}_1) = -2\sqrt{3} < b(x,\mathbf{u}_2) = \sqrt{2/3} - 2\sqrt{3}$. Esto significa que $\mathbf{u}_1$ no es óptima en sesgo. Además, se puede ver que $\mathbf{u}_2$ sí es óptima en sesgo.

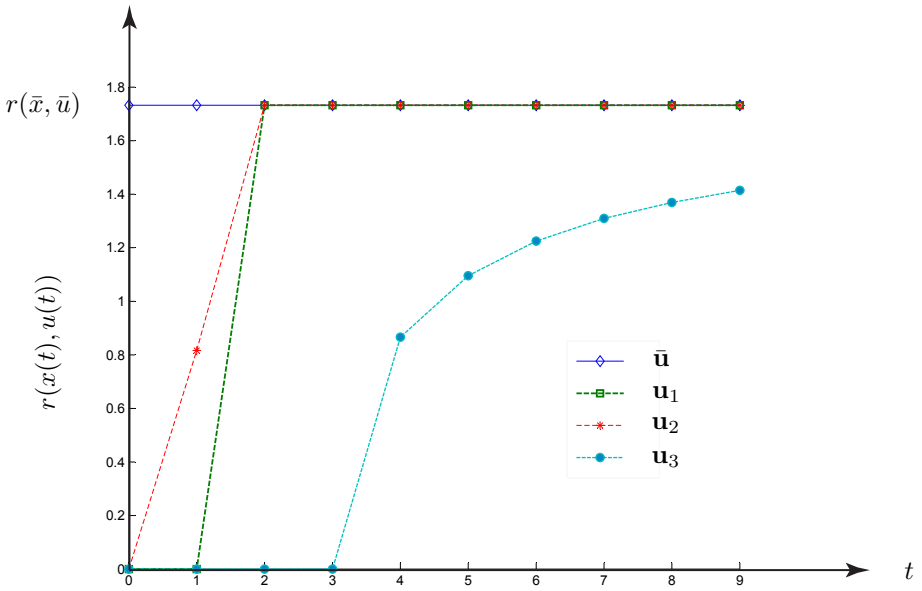

Figura 1.1.

1.6. Conclusiones

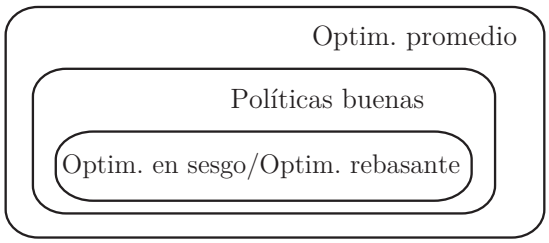

Figura 1.2.

En este capítulo hemos estudiado el modelo forestal de Mitra-Wan como un problema de control óptimo. Hemos considerado cuatro criterios importantes: optimalidad en promedio, optimalidad buena, optimalidad rebasante y optimalidad en sesgo. La Figura 1.2 resume las relaciones de inclusión que existen entre estos criterios: las políticas buenas están estrictamente en el conjunto de políticas optimas en promedio (Teorema 1.3.4); las políticas óptimas en sesgo y las optimas rebasantes son equivalente, y están incluidas estrictamente en el conjunto de las políticas buenas (Teorema 1.4.7).

Finalmente, las soluciones óptimas del problema de control óptimo poseen cierta propiedad de estabilidad conocido como propiedades de tipo autopista: todas las políticas buenas convergen a la política estacionaria óptima (Teorema 1.3.6). Por otra parte, vimos que el valor de las políticas óptimas en promedio es independiente del estado inicial y es igual al valor de la política estacionaria óptima (Teorema 1.3.2).

Capítulo 2

Juegos markovianos de suma cero

Los juegos estocásticos de suma cero han sido estudiados inicialmente por L. S. Shapley [70] y continuado, en particular, en [7, 22]. En estos trabajos el espacio de estados y los conjuntos de acciones de los jugadores son finitos. En ese caso, bajo algunas condiciones adicionales, la existencia del valor de juego y la existencia de estrategias óptimas para los jugadores se demuestran mediante teoremas de punto fijo. Las múltiples aplicaciones – en economía, teoría de colas, biología evolutiva y otros [49, 76, 62] – han motivado la extensión de los juegos estocásticos a espacios más generales. Sin embargo, ejemplos simples, [41, 55, 24], muestran que es necesario imponer algunas condiciones —por ejemplo, continuidad-compacidad, regularidad y estabilidad— para garantizar la existencia del valor del juego y la existencia de estrategias óptimas para los jugadores.

En este capítulo describiremos la estructura general de un juego markoviano de suma cero. Consideraremos las condiciones de continuidad-compacidad, regularidad y estabilidad. Mostraremos las propiedades que tiene el juego bajo estos supuestos.

2.1. Espacios con norma ponderada

En lo que sigue X denotara un *espacio de Borel* (esto es, un subconjunto boreliano de un espacio métrico completo y separable). Su $\sigma-$álgebra de Borel será denotado por $\mathcal{B}(X)$.

Sea X un espacio de Borel, y sea $\mathbb{B}(X)$ el espacio de las funciones medibles con valores reales y acotadas sobre X con la *norma del supremo*

$$\|u\| := \sup_X |u(x)|.$$

Más generalmente, dada una función medible $w : X \to [1, \infty)$, la cual será referida como *función de ponderación o peso*, sea $\mathbb{B}_w(X)$ el espacio de las funciones medibles $u : X \to \mathbb{R}$ con la $w-norma$

$$\|u\|_w := \sup_{x \in X} \frac{|u(x)|}{w(x)} < \infty.$$

Notemos que $\|u\|_w \leq \|u\|$ para todo $u \in \mathbb{B}(X)$. Por lo tanto $\mathbb{B}(X)$ es un subespacio de $\mathbb{B}_w(X)$ y, más aun, ambos son espacios de Banach.

Por otra parte, consideremos también el espacio $\mathbb{M}(X)$ de las medidas con signo μ sobre X con la *norma de variación total*

$$\|\mu\|_{VT} := \sup_{\|u\| \leq 1} \left| \int_X u d\mu \right| = |\mu|(X) < \infty,$$

donde $|\mu| := \mu^+ + \mu^-$ denota la *variación total* de μ, y μ^+, μ^- son la parte positiva y la parte negativa de μ, respectivamente. Análogamente, sea $\mathbb{M}_w(X)$ el espacio de las medidas finitas con signo μ sobre X con la $w-norma$

$$\|\mu\|_w := \sup_{\|u\|_w \leq 1} \left| \int_X u d\mu \right| = \int_X w d|\mu| < \infty. \tag{2.1}$$

De igual manera, $\mathbb{M}(X)$ y $\mathbb{M}_w(X)$ son espacios de Banach y $\mathbb{M}(X) \subset \mathbb{M}_w(X)$. Finalmente, es conveniente mencionar que cada medida $\mu \in \mathbb{M}_w(X)$ se puede identificar con un operador $u \to \mu(u)$ en $\mathbb{B}_w(X)$ definido por

$$\mu(u) := \int_X u(y) \mu(dy), \tag{2.2}$$

con $|\mu(u)| \leq \|u\|_w \|\mu\|_w$.

En lo que sigue denotaremos por $\mathbb{P}(X)$ al espacio de las medidas de probabilidad sobre X con la topología de la convergencia débil. Esto significa que una sucesión de probabilidades μ_n converge débilmente a $\mu \in \mathbb{P}(X)$ si

$$\mu_n(u) \to \mu(u) \quad \forall u \in \mathbb{C}(X),$$

donde $\mathbb{C}(X) \subset \mathbb{B}(X)$ es el espacio de las funciones continuas y acotadas sobre X.

Definición 2.1.1. *Dados dos espacios de Borel X e Y. Una* probabilidad de transición *sobre X dado Y es una función medible $Q : Y \to \mathbb{P}(X)$.*

Para simplificar la notación, escribiremos $Q(B|y)$ en lugar de $Q(y)(B)$ para cada $y \in Y$ y $B \in \mathcal{B}(X)$. Asimismo, denotaremos por $\mathbb{P}(X|Y)$ a la familia

de probabilidades de transición sobre X dado Y. Si $X = Y$, simplemente diremos que Q es una probabilidad de transición sobre X.

Sea Q una probabilidad de transición sobre X. Q induce, de manera natural, dos operadores lineales $u \to Qu$ sobre $\mathbb{B}_w(X)$ y $\mu \to \mu Q$ sobre $\mathbb{M}_w(X)$ del siguiente modo:

$$Qu(x) := \int_X u(y)Q(dy|x), \quad u \in \mathbb{B}_w(X), \quad x \in X, \text{ y} \tag{2.3}$$

$$\mu Q(B) := \int_X Q(B|x)\mu(dx), \quad \mu \in \mathbb{M}_w(X), \quad B \in \mathcal{B}(X). \tag{2.4}$$

Si δ_x es la medida de Dirac en el punto $x \in X$, podemos observar que $\delta_x Q(\cdot) = Q(\cdot|x)$ y $\|\delta_x\|_w = w(x)$. Con esto en mente y tomando en cuenta (2.1), concluimos que los dos operadores tienen la misma norma

$$\begin{aligned} \|Q\|_w &:= \sup\{\|Qu\|_w : \|u\|_w \leq 1\} \qquad (2.5)\\ &= \sup_X \frac{1}{w(x)} \int_X w(y)Q(dy|x) \\ &= \sup_X \frac{1}{w(x)} \|Q(\cdot|x)\|_w \\ &= \sup\{\|\mu Q\|_w : \|\mu\|_w \leq 1\}. \qquad (2.6) \end{aligned}$$

Este valor común se conoce como la $w-$*norma* de Q.

Por último, definimos la *composición* de dos probabilidades de transición Q y R por

$$(QR)(B|x) := \int_X R(B|y)Q(dy|x), B \in \mathcal{B}(X), x \in X. \tag{2.7}$$

Para referirnos a la composición de Q consigo misma, usaremos las siguientes notaciones $Q^0(B|x) := \delta_x(B)$ y $Q^n := QQ^{n-1}, n = 1, 2, ...$

2.2. El modelo del juego

Sea X un espacio de estados. Sean A y B el *espacio de acciones* para el jugador 1 y 2, respectivamente. En cada estado $x \in X$, los jugadores 1 y 2 dispondrán de un subconjunto de *acciones admisibles* $A(x) \subset A$ y $B(x) \subset B$, respectivamente. Asumiremos que cada uno de los conjuntos anteriores son espacios de Borel. Entonces, el conjunto de estados y acciones admisibles

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

es también un espacio de Borel [60]. Por otra parte, en cada estado, los jugadores reciben un pago por las acciones que toman. La *función de pago*

para el jugador $i = 1, 2$ es una función medible $r_i : \mathbb{K} \to \mathbb{R}$. Finalmente, el juego posee una *ley de transición*, que es una probabilidad de transición sobre X dado $\mathbb{K}$, $Q(\cdot|x, a, b)$ con $(x, a, b) \in \mathbb{K}$. En resumen, un juego estocástico de dos personas esta constituido por seis componentes

$$\{X, A(x), B(x), Q(\cdot|x, a, b), r_1(x, a, b), r_2(x, a, b)\}. \tag{2.8}$$

El juego se desarrolla de la siguiente manera. En cada periodo $t \in \mathbb{N}_0$, los jugadores 1 y 2 observan el estado del sistema $x \in X$ y eligen independientemente sus acciones $a \in A(x)$ y $b \in B(x)$, respectivamente. Después de estas acciones, el jugador 1 recibe un pago $r_1(x, a, b)$ y el jugador 2, $r_2(x, a, b)$. Luego el sistema se mueve inmediatamente a un nuevo estado con distribución $Q(\cdot|x, a, b)$. El objetivo de cada uno de los jugadores es encontrar una estrategia que les permita maximizar sus ganancias por el tiempo que dure el juego.

Nota 2.2.1. Nosotros estudiaremos una clase particular de juegos donde el pago que recibe el jugador que gana proviene únicamente del jugador que pierde, esto es $r_1(x, a, b) + r_2(x, a, b) = 0$ para todo $(x, a, b) \in \mathbb{K}$. Esta clase de juegos se conocen como *juegos de suma cero*. En este caso, para estudiarlos basta considerar una única función de pago $r : \mathbb{K} \to \mathbb{R}$ definida por $r(x, a, b) := r_1(x, a, b) = -r_2(x, a, b)$ para todo $(x, a, b) \in \mathbb{K}$. Por otra parte, es importante mencionar que los juegos que no son de suma cero aún no han sido estudiados extensamente y en consecuencia existen muchas preguntas que todavía no se han respondido. Alguno de los pocos trabajos sobre este tema es [39] y las referencias que tiene. Aquí sólo consideraremos el caso de suma cero.

Durante el desarrollo del juego, los jugadores toman sus acciones considerando la historia del juego. Esto es, si el juego se encuentra en el estado x_t, las acciones siguientes, a_t y b_t, serán tomadas evaluando la *historia del juego*

$$h_t := (x_0, a_0, b_0, ..., x_{t-1}, a_{t-1}, b_{t-1}, x_t).$$

El conjunto de las posibles historias del juego hasta el tiempo $t = 1, 2, ...$ es $H_t := \mathbb{K} \times H_{t-1}$ donde $H_0 := X$.

Definición 2.2.2. *Una* estrategia *para el jugador 1 es una sucesión* $\pi^1 := \{\pi_t^1, t \in \mathbb{N}_0\}$ *de probabilidades de transición* $\pi_t^1 : H_t \to \mathbb{P}(A)$ *tal que*

$$\pi_t^1(A(x_t)|h_t) = 1 \text{ para cada } h_t \in H_t \text{ y } t = 0, 1, ...$$

El conjunto de estrategias para el jugador 1 se denota por Π_1. Una estrategia, en general, depende de la historia completa del juego. Es decir,

depende de los estados por los que ha pasado el juego y las decisiones que se ha tomado en cada uno de ellos. Por el contrario, existe una subclase importante de estrategias que no dependen de toda la historia. Esas estrategias solamente dependen del estado en el que se encuentra el juego al momento de decidir. La importancia de estas estrategias se debe a que, en algún sentido, son menos difíciles de calcular. Bajo ciertas condiciones, como sucede en los procesos de control markovianos (PCMs) [32], el equilibrio se logra en este tipo de estrategias. Estas estrategias se definen de la siguiente manera. Sea Φ_1 el conjunto de todas las probabilidades de transición $\varphi : X \to \mathbb{P}(A)$ tal que $\varphi(A(x)|x) = 1$ para cada $x \in X$.

Definición 2.2.3. *Una estrategia $\pi^1 \in \Pi_1$ se dice que es* estacionaria *si existe $\varphi \in \Phi_1$ tal que $\pi_t^1(\cdot|h_t) = \varphi(\cdot|x_t)$ para todo $h_t \in H_t$ y $t \in \mathbb{N}_0$.*

Para no agregar más notaciones, representaremos por Φ_1 al conjunto de las estrategias estacionarias para el jugador 1. Para el jugador 2 se definen de manera análoga, el conjunto de estrategias generales Π_2 y el conjunto de estrategias estacionarias Φ_2.

Consideremos el espacio medible $\Omega := (X \times A \times B)^\infty$ y su $\sigma-$álgebra producto $\mathcal{F}$. Cada elemento de Ω representa un escenario completo del juego donde los jugadores interactúan por un tiempo indefinido.

De acuerdo al teorema Ionescu-Tulcea (ver [6, Capítulo 7]), sabemos que se cumple lo siguiente.

Teorema 2.2.4. *Para cada $(\pi^1, \pi^2) \in \Pi_1 \times \Pi_2$ y cada estado inicial $x \in X$, existe una medida de probabilidad $P_x^{\pi^1\pi^2}$; un proceso estocástico de estados y acciones $\{(x_t, a_t, b_t), t \in \mathbb{N}_0\}$ definido en $(\Omega, \mathcal{F})$ que satisface lo siguiente para cada $V \subset X, C \in \mathcal{B}(A), D \in \mathcal{B}(B), h_t \in H_t$, y $t \in \mathbb{N}_0$:*

$$P_x^{\pi^1\pi^2}(x_0 \in V) = \delta_x(V), \tag{2.9}$$

$$P_x^{\pi^1\pi^2}(a_t \in C, b_t \in D|h_t) = \pi_t^1(C|h_t)\pi_t^2(D|h_t), \tag{2.10}$$

$$P_x^{\pi^1\pi^2}(x_{t+1} \in V|h_t, a_t, b_t) = Q(V|x_t, a_t, b_t). \tag{2.11}$$

Denotaremos por $E_x^{\pi^1,\pi^2}$ al operador esperanza con respecto a la medida de probabilidad $P_x^{\pi^1,\pi^2}$.

Entre otras cosas, el anterior teorema nos dice que los jugadores pueden eligir sus acciones de manera *independiente*, como se ha mencionado anteriormente.

Nota 2.2.5. Los juegos markovianos de dos jugadores se pueden representar mediante un sistema dinámico estocástico a tiempo discreto:

$$x_{n+1} = f(x_n, a_n, b_n, \xi_n) \quad n = 0, 1, 2, ... \tag{2.12}$$

donde x_0 es un estado inicial dado y x_n es el estado del juego al tiempo n, mientras que a_n y b_n son las acciones de los jugadores 1 y 2, respectivamente. Las perturbaciones $\{\xi_i\}_{i=0}^{\infty}$ forman una sucesión de variables aleatorias independientes e idénticamente distribuidas (i.i.d.) con una distribución común μ, e independientes del estado inicial x_0.

Si nuestro modelo está en la forma (2.12), podemos escribirlo como en (2.8) con la ley de transición

$$Q(B|x,a,b) = \mu(\{s \in S | f(x,a,b,s) \in B\}) \tag{2.13}$$

$$= \int I_B(f(x,a,b,s))\mu(ds), \tag{2.14}$$

donde I_B es la función indicadora del conjunto B.

Inversamente, dado un juego markoviano en la forma (2.8), es posible llevar a la forma (2.12); es decir, existe una función medible f y una sucesión de variables aleatorias independientes $\{\xi_i\}_{i=0}^{\infty}$ que satisfacen (2.12). Ver, por ejemplo, [21].

2.3. Supuestos y propiedades

El espacio de estados X y los conjuntos de acciones admisibles $A(x)$ y $B(x)$ son espacios de Borel. Estos espacios podrían ser desde conjuntos finitos hasta espacios continuos. De manera similar, la única propiedad que hemos mencionado de la función de pagos r es la medibilidad. Bajo estas condiciones generales, podrían no existir estrategias óptimas para los jugadores. En esta sección introduciremos las condiciones que impondremos a nuestro modelo. Estas condiciones son una extención natural de las hipótesis usuales de PCMs; ver [23, 28, 32, 33, 73].

Sea $\mathbb{B}_w(\mathbb{K})$ el espacio de funciones medibles $u : \mathbb{K} \to \mathbb{R}$ tal que

$$\|u\|_w := \sup_{(x,a,b)\in\mathbb{K}} \frac{|u(x,a,b)|}{w(x)} < \infty. \tag{2.15}$$

Notemos que $\mathbb{B}_w(X)$ se puede ver como un subespacio de $\mathbb{B}_w(\mathbb{K})$.

Dado $u \in \mathbb{B}_w(\mathbb{K})$ y Q la ley de transición del juego, definimos

$$u(x,\mu,\nu) := \int_{A(x)} \int_{B(x)} u(x,a,b)\nu(db)\mu(da) \tag{2.16}$$

y

$$Q(\cdot|x,\mu,\nu) := \int_{A(x)} \int_{B(x)} Q(\cdot|x,a,b)\nu(db)\mu(da) \tag{2.17}$$

donde μ y ν son medidas de probabilidad sobre $A(x)$ y $B(x)$, respectivamente.

Supuesto 2.3.1 (Continuidad y compacidad)**.** Para cada $(x, a, b) \in \mathbb{K}$:

1. Los conjuntos de acciones admisibles $A(x)$ y $B(x)$ son compactos y no vacíos.

2. La función de pagos $r \in \mathbb{B}_w(\mathbb{K})$ es tal que: $r(x, \cdot, b)$ es semicontinua superiormente en $A(x)$ y $r(x, a, \cdot)$ es semicontinua inferiormente en $B(x)$.

3. Para cada $(x, a, b) \in \mathbb{K}$ y toda función medible y acotada $v : X \to \mathbb{R}$, las funciones $\int_X v(y)Q(dy|x, \cdot, b)$ y $\int_X v(y)Q(dy|x, a, \cdot)$ son continuas en $A(x)$ y $B(x)$, respectivamente. Este supuesto es también válido cuando sustituimos v por w.

Supuesto 2.3.2 (Estabilidad)**.** Existe una medida de probabilidad $\nu \in \mathbb{P}(X)$, un número real $0 < \gamma < 1$ y una función medible $\beta : \mathbb{K} \to [0, 1]$ tal que para cada $(x, a, b) \in \mathbb{K}$ y $D \in \mathcal{B}(X)$ se satisface lo siguiente:

1. $Q(D|x, a, b) \geq \beta(x, a, b)\nu(D)$.

2. $\int_X w(y)Q(dy|x, a, b) \leq \gamma w(x) + \beta(x, a, b)\|\nu\|_w$

3. $\inf_{(\varphi,\psi)\in\Phi_1\times\Phi_2} \int_X \beta(x, \varphi(x), \psi(x))\nu(dx) > 0$.

Supuesto 2.3.3 (Regularidad)**.** Existe una medida $\sigma-$finita λ sobre X con respecto a la cual, para cada par (φ, ψ) en $\Phi_1 \times \Phi_2$, la probabilidad de transición markoviana $Q(\cdot|x, \varphi(x), \psi(x))$ es $\lambda-$irreducible (acerca de la propiedad $\lambda-$irreducibilidad, ver la Definición A.1.1).

Es importante mencionar que, en algunos casos, las condiciones mencionadas anteriormente pueden ser debilitadas. Ver por ejemplo [35] o bien [36]. Por otra parte, un supuesto más fuerte que el 2.3.3, pero fácil de verificar, es el siguiente.

Supuesto 2.3.4. Existe una medida $\sigma-$finita λ sobre X y una función de densidad positiva $g(x, a, b, \cdot)$ tal que

$$Q(D|x, a, b) = \int_D g(x, a, b, y)\lambda(dy).$$

La siguiente proposición es una de las consecuencias mas importantes de los supuestos anteriores. Algunas propiedades adicionales se ven en la Sección A.3 del apéndice.

Proposición 2.3.5. *Bajo los Supuestos 2.3.2 y 2.3.3, para cada par* $(\varphi, \psi) \in \Phi_1 \times \Phi_2$*, el proceso de estados* $\{x_t\}$ *es* Harris recurrente positivo. *En consecuencia, la probabilidad de transición markoviana* $Q(x, \varphi(x), \psi(x))$ *tiene una única* medida de probabilidad invariante $\mu_{\varphi,\psi}$ *sobre* X*, es decir*

$$\mu_{\varphi,\psi}(D) = \int_X Q(D|x, \varphi(x), \psi(x))\mu_{\varphi,\psi}(dx) \tag{2.18}$$

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

No es difícil ver que $\mu_{\varphi,\psi} \in \mathbb{M}_w$ para todo $\varphi \in \Phi_1$ y $\psi \in \Phi_2$. Pues, integrando con respecto a $\mu_{\varphi,\psi}$ ambos lados de la condición (2) del Supuesto 2.3.2, y usando la igualdad (2.18), obtenemos

$$\int_X w(y)\mu_{\varphi,\psi}(dy) \leq \gamma \int_X w(y)\mu_{\varphi,\psi}(dy) + \|\nu\|_w.$$

Por lo tanto, usando (2.1), vemos que $\|\mu_{\varphi,\psi}\|_w \leq \|\nu\|_w/(1-\gamma)$.

Nota 2.3.6. Bajo los mismos supuestos de la Proposición 2.3.5, y de acuerdo al Teorema A.3.3, el proceso de estados $\{x_t\}$ tiene la propiedad de $w-$*ergodicidad geométrica* (A.3.2), esto es, existen números reales $0 < \theta < 1$ y $0 < M$ tal que

$$\left| \int_X u(y)Q^t(dy|x, \varphi(x), \psi(x)) - \int_X u(y)\mu_{\varphi,\psi}(dy) \right| \leq w(x)\|u\|_w M\theta^t \tag{2.19}$$

para cada $u \in \mathbb{B}_w(X), x \in X$ y $t \in \mathbb{N}_0$, donde Q^t denota la probabilidad de transición markoviana de la etapa t.

Lema 2.3.7. *Asumiendo que se satisfacen los Supuestos 2.3.1 y 2.3.2, sea* $(\varphi, \psi) \in \Phi_1 \times \Phi_2$*,* $u \in \mathbb{B}_w(\mathbb{K})$ *y* $x \in X$ *un estado inicial arbitrario. Entonces para todo* $t \in \mathbb{N}_0$ *se satisface*

1. $E_x^{\varphi,\psi} w(x_t) \leq kw(x)$*, donde* $k := 1 + \|\nu\|_w/(1-\gamma)$*,*
2. $|E_x^{\varphi,\psi} u(x_t, a_t, b_t)| \leq \|u\|_w kw(x)$*,*
3. $\lim_{t\to\infty} \frac{1}{t}|E_x^{\varphi,\psi} u(x_t, a_t, b_t)| = 0$.

Demostración. 1) Si $t = 0$, es evidente que $E_x^{\varphi,\psi} w(x) \leq w(x) + \|\nu\|_w$. Supongamos que $t \geq 1$. De (2.11) y el Supuesto 2.3.2 (2)

$$\begin{aligned} E_x^{\varphi,\psi}[w(x_t)|h_{t-1}, a_{t-1}, b_{t-1}] &= \int_X w(y)Q(dy|x_{t-1}, a_{t-1}, b_{t-1}) \\ &\leq \gamma w(x_{t-1}) + \|\nu\|_w. \end{aligned}$$

Tomando esperanzas en ambos lados de la desigualdad, llegamos a la relación $E_x^{\varphi,\psi} w(x_t) \leq \gamma E_x^{\varphi,\psi} w(x_{t-1}) + \|\nu\|_w$. Iterando esta desigualdad obtenemos

$$E_x^{\varphi,\psi} w(x_t) \leq \gamma w(x) + \|\nu\|_w \sum_{i=0}^{t-1} \gamma^j \leq [1 + \|\nu\|_w/(1-\gamma)]w(x).$$

2) De (2.15) sabemos que $|r(x_t, a_t, b_t)| \leq \|u\|_w w(x_t)$ para todo $t = 0, 1, ...,$ entonces

$$|E_x^{\varphi,\psi} r(x_t, a_t, b_t)| \leq \|u\|_w E_x^{\varphi,\psi} w(x_t) \leq \|u\|_w k w(x).$$

3) Se sigue de la anterior desigualdad. □

A continuación introducimos dos conceptos que nos permitirán definir algunos criterios de optimalidad: la *ecuación de Poisson* y el *operador de Bellman*.

Sea $P \in \mathbb{P}(X|X)$ una probabilidad de transición markoviana en X y $c \in \mathbb{B}_w(X)$ una función medible dada. La relación

$$\xi + u(x) = c(x) + \int_X u(y) P(dy|x) \quad \forall x \in X \tag{2.20}$$

se conoce como la *ecuación de Poisson* (E.P.), cuya solución es un par $(\xi, u) \in \mathbb{R} \times \mathbb{B}_w(X)$. Esta solución también se conoce como el *par canónico*.

Para lo que sigue, denotaremos por $\mathbb{A}(x)$ y $\mathbb{B}(x)$ al espacio de medidas de probabilidad sobre los conjuntos $A(x)$ y $B(x)$, respectivamente. Por el Supuesto 2.3.1(1), estos conjuntos son compactos.

Dada una función $u \in \mathbb{B}_w(X)$, r la función de pago del juego y Q la ley de transición, consideramos la función

$$(\varphi, \psi) \mapsto r(x, \varphi(x), \psi(x)) + \int_X u(y) Q(dy|x, \varphi(x), \psi(x)). \tag{2.21}$$

Por el Supuesto 2.3.1, esta función es semicontinua superiormente en la variable φ y es semicontinua inferiormente en la variable ψ. Bajo estas condiciones, el *operador minimax* $T : \mathbb{B}_w(X) \to \mathbb{B}_w(X)$,

$$Tu(x) := \max_{\varphi \in \mathbb{A}(x)} \min_{\psi \in \mathbb{B}(x)} \left\{ r(x, \varphi(x), \psi(x)) + \int_X u(y) Q(dy|x, \varphi(x), \psi(x)) \right\}$$

está bien definido. Más aun, mediante los teoremas minimax [17, 71] y los teoremas de selector medible [60, 36], tenemos la siguiente proposición.

Proposición 2.3.8. *Bajo los Supuestos 2.3.1 y 2.3.2(2), existen estrategias* $\varphi^* \in \Phi_1$ *y* $\psi^* \in \Phi_2$ *tal que para todo* $x \in X$,

$$\begin{aligned} Tu(x) &= r(x, \varphi^*(x), \psi^*(x)) + \int_X u(y) Q(dy|x, \varphi^*(x), \psi^*(x)) \qquad (2.22) \\ &= \max_{\varphi \in \mathbb{A}(x)} \left\{ r(x, \varphi(x), \psi^*(x)) + \int_X u(y) Q(dy|x, \varphi(x), \psi^*(x)) \right\} \\ &= \min_{\psi \in \mathbb{B}(x)} \left\{ r(x, \varphi^*(x), \psi(x)) + \int_X u(y) Q(dy|x, \varphi^*(x), \psi(x)) \right\}. \end{aligned}$$

2.4. Conclusiones

Definimos la estructura general de un juego markoviano de suma cero; el conjunto de estados es un espacio de Borel, los conjuntos de acciones admisibles son compactos y la función de pagos es medible (posiblemente no acotada). En ese marco, definimos la clase de estrategias que los jugadores pueden ejecutar: las estrategias estacionarias. Luego, introducimos condiciones de estabilidad sobre la ley de transición. En seguida, analizamos algunas consecuencias directas de los supuestos: entre los más importantes, el Teorema 2.3.5 muestra la existencia de una medida de probabilidad invariante para cada ley de transición. Más aun, la Nota 2.3.6 muestra que la convergencia hacia la medida de probabilidad invariante es uniforme y exponencial. Finalmente, introducimos la ecuación de Poisson y el operador de Bellman que en los siguientes capítulos nos ayudará a definir algunos criterios de optimalidad.

Capítulo 3

Optimalidad promedio

En este capítulo estudiaremos un primer grupo de criterios de optimalidad ergódicos: *optimalidad en promedio*, *optimalidad canónica*, *optimalidad promedio F-fuerte* y finalmente la *optimalidad* -1*-descontada*. Veremos que todos estos criterios, aunque con definiciones completamente diferentes, son equivalentes. La optimalidad en promedio esperado y la optimalidad canónica han sido estudiados extensivamente para procesos de control markovianos (PCMs), ver por ejemplo, [29, 32]. Una extensión a juegos markovianos de estos criterios se ha realizado en [30, 61, 20]. Por su parte, la optimalidad promedio F-fuerte y la optimalidad -1-descontada han sido estudiados solamente para PCMs en [32] y [33], respectivamente.

Los resultados que mostraremos en este capítulo reúnen algunos de los obtenidos en [30] y las extensiones que realizaremos a los resultados para PCMs de [32, 33].

3.1. Optimalidad promedio

Para cada estado inicial $x \in X$, estrategias $\varphi \in \Phi_1$, $\psi \in \Phi_2$, y todo horizonte finito n, el *pago total esperado* hasta la etapa n para el jugador 1 es

$$J_n(x, \varphi, \psi) := E_x^{\varphi,\psi} \sum_{t=0}^{n-1} r(x_t, a_t, b_t).$$

Al considerar un juego de horizonte infinito, cuando $n \to \infty$, comparar directamente los limites, para estrategias diferentes, podría no ser una buena alternativa. Es posible que estos límites no existan para ciertas estrategias. Para evitar esas dificultades fijaremos nuestra atención en el comportamiento en promedio de la función $J_n(x, \varphi, \psi)$.

El *pago promedio esperado* a largo plazo para el jugador 1 es

$$J(x,\varphi,\psi) := \liminf_{n\to\infty} J_n(x,\varphi,\psi)/n.$$

Este será la función objetivo de los jugadores. El primer jugador buscará estrategias que maximicen mientras que el segundo estrategias que minimicen. El límite inferior se debe a que estamos considerando el peor escenario para el primer jugador.

La función *valor inferior* y *valor superior* del juego se definen, respectivamente, como

$$L(x) := \sup_{\varphi\in\Phi_1}\inf_{\psi\in\Phi_2} J(x,\varphi,\psi) \quad \text{y} \quad U(x) := \inf_{\psi\in\Phi_2}\sup_{\varphi\in\Phi_1} J(x,\varphi,\psi)$$

para todo $x \in X$. Claramente, $L(x) \le U(x)$ para cada $x \in X$. Si $L(x) = U(x)$ para todo $x \in X$, entonces diremos que la función es el *valor* del juego y lo denotaremos por $V(\cdot)$. Como veremos mas adelante, bajo las condiciones que estamos considerando, el juego tiene un único valor. Sin embargo, si se cambian las condiciones, no necesariamente se conserva la propiedad. Para un ejemplo de ese tipo, ver [72].

Definición 3.1.1. *Supongamos que el juego tiene una función de valor $V(\cdot)$. Diremos que un par de estrategias $(\varphi^*,\psi^*) \in \Phi_1 \times \Phi_2$ es un* equilibrio promedio esperado *si*

$$\sup_{\varphi\in\Phi_1} J(x,\varphi,\psi^*) = \inf_{\psi\in\Phi_2} J(x,\varphi^*,\psi) = V(x) \tag{3.1}$$

para cada $x \in X$.

El conjunto de los equilibrios en promedio esperado será denotado por $(\Phi_1 \times \Phi_2)_{ep}$.

Para $(\varphi,\psi) \in \Phi_1 \times \Phi_2$, sea

$$J(\varphi,\psi) := \int_X r(x,\varphi(x),\psi(x))\mu_{\varphi,\psi}(dx) \tag{3.2}$$

donde $\mu_{\varphi,\psi}$ es como en la Proposición 2.3.5.

Lema 3.1.2. *Bajo los Supuestos 2.3.1, 2.3.2 y 2.3.3, obtenemos la desigualdad*

$$|J_n(x,\varphi,\psi) - nJ(\varphi,\psi)| \le \|r\|_w M w(x)/(1-\theta) \tag{3.3}$$

para todo estado inicial $x \in X$ y toda estrategia $(\varphi,\psi) \in \Phi_1 \times \Phi_2$.

Demostración. Esta propiedad se obtiene de la desigualdad (2.19) sustituyendo $u(x)$ por $r(x,\varphi(x),\psi(x))$ en la desigualdad. □

La siguiente proposición muestra que, bajo los supuestos habituales, el pago promedio esperado es independiente del estado inicial del sistema. Más aun, este se puede expresar en términos de la medida de probabilidad invariante de la Proposición 2.3.5.

Proposición 3.1.3. *Bajo los mismos supuestos del Lema 3.1.2. La igualdad*

$$J(x,\varphi,\psi) = \lim_{n\to\infty} \frac{J_n(x,\varphi,\psi)}{n} = J(\varphi,\psi) \tag{3.4}$$

se satisface para todo $x \in X$ *y* $(\varphi,\psi) \in \Phi_1 \times \Phi_2$.

Demostración. Se sigue inmediatamente del Lema 3.1.2. □

Definición 3.1.4. *Una cuádrupla* $(\xi^*, u^*, \varphi^*, \psi^*)$*, que consiste de una constante* $\xi^* \in \mathbb{R}$*, una función medible* $\xi^* : X \to \mathbb{R}$ *y un par de estrategias estacionarias* $(\varphi,\psi) \in \Phi_1 \times \Phi_2$*, se dice que es una* cuádrupla canónica *si para todo* $x \in X$ *se cumple*

$$\xi^* + u^*(x) = r(x,\varphi^*(x),\psi^*(x)) + \int_X u^*(y)Q(dy|x,\varphi^*(x),\psi^*(x)) \tag{3.5}$$

$$= \max_{\varphi\in\mathbb{A}(x)} \left[r(x,\varphi,\psi^*(x)) + \int_X u^*(y)Q(dy|x,\varphi,\psi^*(x)) \right] \tag{3.6}$$

$$= \min_{\psi\in\mathbb{B}(x)} \left[r(x,\varphi^*(x),\psi) + \int_X u^*(y)Q(dy|x,\varphi^*(x),\psi) \right]. \tag{3.7}$$

En este caso diremos que (φ^*,ψ^*) *es un par de* equilibrios canónicos.

El conjunto de los equilibrios canónicos será denotado por $(\Phi_1 \times \Phi_2)_{ca}$.

Notemos que la igualdad 3.5 está relacionado con la ecuación de Poisson 2.20. Por otra parte, las igualdades 3.6 y 3.7 están relacionadas con el operador minimax (2.22). De hecho,

$$\xi^* + u^*(x) = Tu^*(x)$$

para todo $x \in X$.

Teorema 3.1.5. *Si los Supuestos 2.3.1,2.3.2 y 2.3.4 se cumplen, entonces* $(\Phi_1 \times \Phi_2)_{ep} = (\Phi_1 \times \Phi_2)_{ca}$*. Más aun, existe una cuádrupla* $(\xi^*, u^*, \varphi^*, \psi^*)$ *canónica con* $u^* \in \mathbb{B}(X)$ *y se satisface* $V(x) = \xi^*$ *para todo* $x \in X$.

Demostración. Ver Teorema 5.8. de [30]. □

Consideremos la *función de discrepancia* $\Delta : \mathbb{K} \to \mathbb{R}$ dada por

$$\Delta(x,a,b) := r(x,a,b) + \int_X u^*(y)Q(dy|x,a,b) - u^*(x) - \xi^*, \tag{3.8}$$

donde u^* y ξ^* son los del Teorema 3.1.5. Por el Supuesto 2.3.2, es evidente que $\Delta \in \mathbb{B}_w(\mathbb{K})$. Además, teniendo en cuenta la Definición 3.1.4 y el Teorema 3.1.5, es posible caracterizar el equilibrio promedio en términos de esta función.

Proposición 3.1.6. *La estrategia* $(\varphi^*, \psi^*) \in \Phi_1 \times \Phi_2$ *es un equilibrio en promedio si y sólo si*

$$\begin{aligned} \Delta(x, \varphi^*(x), \psi^*(x)) &= \max_{\varphi \in \mathbb{A}(x)} \Delta(x, \varphi, \psi^*(x)) && (3.9)\\ &= \min_{\psi \in \mathbb{B}(x)} \Delta(x, \varphi^*(x), \psi) && (3.10)\\ &= 0. && (3.11) \end{aligned}$$

A continuación definiremos uno de los criterios de optimalidad introducido por Flynn [18]. Dado un par de estrategias y una etapa finita, calcularemos la diferencia que existe entre la acumulación de beneficios usando la estrategia dada y la acumulación óptima hasta esa misma etapa. Luego, nos fijaremos en el comportamiento promedio que tiene esta diferencia al pasar el tiempo. Este criterio ha sido estudiado ampliamente para procesos de decisión markovianos en [28, 32].

Definición 3.1.7. *Diremos que una estrategia* $(\varphi^*, \psi^*) \in \Phi_1 \times \Phi_2$ *es un* equilibrio promedio F-fuerte *si para cada* $x \in X$*,*

$$\lim_{n\to\infty} \frac{1}{n} \left[J_n(x, \varphi^*, \psi^*) - \sup_{\varphi \in \Phi_1} J_n(x, \varphi, \psi^*) \right] = 0, \tag{3.12}$$

y

$$\lim_{n\to\infty} \frac{1}{n} \left[J_n(x, \varphi^*, \psi^*) - \inf_{\psi \in \Phi_2} J_n(x, \varphi^*, \psi) \right] = 0. \tag{3.13}$$

El conjunto de los equilibrios promedio F-fuerte será denotado por $(\Phi_1 \times \Phi_2)_F$.

Teorema 3.1.8. *Una estrategia* $(\varphi^*, \psi^*) \in \Phi_1 \times \Phi_2$ *es equilibrio en promedio si y solamente si* (φ^*, ψ^*) *es un equilibrio promedio F-fuerte.*

Demostración. Sea (φ^*, ψ^*) un equilibrio promedio F-fuerte. De las ecuaciones (3.12) y (3.13) obtenemos

$$\begin{aligned} J(x,\varphi^*,\psi^*) =& \liminf_{n\to\infty} \frac{1}{n} J_n(x,\varphi^*,\psi^*) \\ =& \lim_{n\to\infty} \frac{1}{n}\left[J_n(x,\varphi^*,\psi^*) - \sup_{\varphi\in\Phi_1} J(x,\varphi,\psi^*)\right] \\ &+ \liminf_{n\to\infty} \frac{1}{n} \sup_{\varphi\in\Phi_1} J(x,\varphi,\psi^*) \\ \geq& J(x,\varphi,\psi^*) \end{aligned}$$

y también

$$\begin{aligned} J(x,\varphi^*,\psi^*) =& \liminf_{n\to\infty} \frac{1}{n} J_n(x,\varphi^*,\psi^*) \\ =& \lim_{n\to\infty} \frac{1}{n}\left[J_n(x,\varphi^*,\psi^*) - \inf_{\psi\in\Phi_2} J_n(x,\varphi^*,\psi)\right] \\ &+ \liminf_{n\to\infty} \frac{1}{n} \inf_{\psi\in\Phi_2} J_n(x,\varphi^*,\psi) \\ \leq& J(x,\varphi^*,\psi). \end{aligned}$$

para cada $x \in X$ y $(\varphi, \psi) \in \Phi_1 \times \Phi_2$. De acuerdo a (3.1), concluimos que (φ^*, ψ^*) es un equilibrio promedio.

Recíprocamente, supongamos que (φ^*, ψ^*) es un equilibrio promedio. Entonces, por 3.1.2,

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n}\left[J_n(x,\varphi^*,\psi^*) - \sup_{\varphi\in\Phi_1} J_n(x,\varphi,\psi^*)\right] &= J(x,\varphi^*,\psi^*) - \sup_{\varphi\in\Phi_1} J(x,\varphi,\psi^*) \\ &= 0. \end{aligned}$$

De manera similar

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n}\left[J_n(x,\varphi^*,\psi^*) - \inf_{\psi\in\Phi_2} J_n(x,\varphi^*,\psi)\right] &= J(x,\varphi^*,\psi^*) - \inf_{\psi\in\Phi_2} J(x,\varphi^*,\psi) \\ &= 0. \end{aligned}$$

Por lo tanto (φ^*, ψ^*) es un equilibrio promedio fuerte. □

En resumen, del Teorema 3.1.5 y Teorema 3.1.8, concluimos que el conjunto de equilibrios promedio es no vacío y

$$(\Phi_1 \times \Phi_2)_{ep} = (\Phi_1 \times \Phi_2)_{ca} = (\Phi_1 \times \Phi_2)_F.$$

3.2. Optimalidad con descuento

Dado un número real fijo $0 < \alpha < 1$ y un estado arbitrario $x \in X$, sea $V_\alpha(x, \varphi, \psi)$ el *pago esperado con descuento* α cuando los jugadores usan la estrategia $(\varphi, \psi) \in \Phi_1 \times \Phi_2$, esto es,

$$V_\alpha(x, \varphi, \psi) := \mathbb{E}_x^{\varphi,\psi}\left[\sum_{t=0}^{\infty} \alpha^t r(x_t, a_t, b_t)\right]. \tag{3.14}$$

En general, para una función $u \in \mathbb{B}_w(\mathbb{K})$ escribiremos

$$V_\alpha(x, \varphi, \psi, u) := \mathbb{E}_x^{\varphi,\psi}\left[\sum_{t=0}^{\infty} \alpha^t u(x_t, a_t, b_t)\right]. \tag{3.15}$$

Lema 3.2.1. *Sea $k = 1 + \|\nu\|_w/(1-\gamma)$ como en el Lema 2.3.7. Bajo los Supuestos 2.3.1 y 2.3.2, se tiene*

$$|V_\alpha(x, \varphi, \psi, u)| \leq \frac{k\|u\|_w}{1-\alpha} w(x) \tag{3.16}$$

para todo $u \in \mathbb{B}_w(\mathbb{K})$, $(\varphi, \psi) \in \Phi_1 \times \Phi_2$ y $x \in X$.

Demostración. Del Lema 2.3.7

$$\begin{aligned} |V_\alpha(x, \varphi, \psi, u)| &\leq \mathbb{E}_x^{\varphi,\psi}\left[\sum_{t=0}^{\infty} \alpha^t |u(x_t, a_t, b_t)|\right] \\ &\leq \|u\|_w \sum_{t=0}^{\infty} \alpha^t E_x^{\varphi,\psi} w(x_t) \\ &\leq \|u\|_w w(x) k/(1-\alpha). \end{aligned}$$

□

Ahora definimos la función de valor con descuento α,

$$V_\alpha(x) := \max_{\varphi \in \mathbb{A}(x)} \min_{\psi \in \mathbb{B}(x)} V_\alpha(x, \varphi, \psi).$$

Por los teoremas minimax [17, 71] y el anterior Lema 3.2.1, la función $V_\alpha(x)$ esta bien definida y está en $\mathbb{B}_w(X)$. Adicionalmente, estos mismos argumentos aseguran la existencia del siguiente equilibrio.

Definición 3.2.2. *Diremos que un par de estrategias $(\varphi^*, \psi^*) \in \Phi_1 \times \Phi_2$ es un* equilibrio con descuento α *si*

$$V_\alpha(x, \varphi, \psi^*) \leq V_\alpha(x, \varphi^*, \psi^*) \leq V_\alpha(x, \varphi^*, \psi)$$

para todo $(\varphi, \psi) \in \Phi_1 \times \Phi_2$ y $x \in X$.

3.3. Equilibrio n-descontado fuerte

En esta sección analizaremos otro de los criterios introducidos originalmente por Veinott [74, 75]. Estos criterios, bajo ciertas condiciones, hacen que los juegos markovianos sin descuento y los que tienen descuento sean casos especiales de una familia más amplia. En particular, veremos que el equilibrio promedio se puede obtener mediante una aproximación de juegos markovianos con descuento.

Definición 3.3.1. *Para $n = -1, 0, 1, 2, ...$, diremos que la estrategia $(\varphi, \psi) \in \Phi_1 \times \Phi_2$ es* equilibrio $n-$descontado fuerte *si*

$$\lim_{\alpha \to 1}(1-\alpha)^{-n}[V_\alpha(x, \varphi^*, \psi^*) - \sup_{\varphi \in \Phi_1} V_\alpha(x, \varphi, \psi^*)] = 0, \tag{3.17}$$

$$\lim_{\alpha \to 1}(1-\alpha)^{-n}[V_\alpha(x, \varphi^*, \psi^*) - \inf_{\psi \in \Phi_2} V_\alpha(x, \varphi^*, \psi)] = 0 \tag{3.18}$$

para todo $x \in X$.

Lema 3.3.2. *Bajo los Supuestos 2.3.1, 2.3.2 y 2.3.4, tenemos que*

$$\lim_{\alpha \to 1}(1-\alpha)V_\alpha(x, \varphi, \psi, u) = \mu_{\varphi,\psi}(u)$$

para todo $x \in X$, $u \in \mathbb{B}_w(X)$ y $(\varphi, \psi) \in \Phi_1 \times \Phi_2$.

Demostración. Sean $x \in X$ y $u \in \mathbb{B}_w(X)$ arbitrarios. Por la Nota 2.3.6, tenemos

$$\begin{aligned} |(1-\alpha)V_\alpha(x, \varphi, \psi) - \mu_{\varphi,\psi}(u)| &= (1-\alpha)\left|V_\alpha(x, \varphi, \psi) - \sum_{t=0}^{\infty} \alpha^t \mu_{\varphi,\psi}(u)\right| \\ &\leq (1-\alpha)\sum_{t=0}^{\infty} \alpha^t |E_x^{\varphi,\psi}(x_t) - \mu_{\varphi,\psi}(u)| \\ &\leq (1-\alpha)\frac{M}{1-\alpha\theta}\|u\|_w w(x). \end{aligned}$$

La afirmación se sigue de inmediato al aplicar el límite $\alpha \to 1$.

□

Lema 3.3.3. *Sean Δ la función discrepancia y ξ^*, u^* como en el Teorema 3.1.4. Entonces*

$$\begin{aligned} V_\alpha(x, \varphi, \psi, \Delta) = V_\alpha(x, \varphi, \psi) + \frac{1-\alpha}{\alpha} V_\alpha(x, \varphi, \psi, u^*) - \frac{1}{\alpha}u^*(x) \\ - \xi^*/(1-\alpha) \end{aligned} \tag{3.19}$$

para todo $(\varphi, \psi) \in \Phi_1 \times \Phi_2$ y $x \in X$.

Demostración. De la relación (2.11) sabemos que

$$E_x^{\varphi,\psi}[u^*(x_t)|h_{t-1}, a_{t-1}, b_{t-1}] = \int_X u^*(y)Q(dy|x_{t-1}, a_{t-1}, b_{t-1})$$

para todo $t = 1, 2, ...$ Por lo tanto, aplicando esperanzas en ambos lados llegamos a que

$$E_x^{\varphi,\psi}u^*(x_t) = E_x^{\varphi,\psi}\int_X u^*(y)Q(dy|x_{t-1}, a_{t-1}, b_{t-1}) \tag{3.20}$$

para todo $t = 1, 2, ...$ Finalmente, de (3.14), (3.15) y (3.20) obtenemos lo deseado:

$$\begin{aligned} V_\alpha(x, \varphi, \psi, \Delta) &= V_\alpha(x, \varphi, \psi) + E_x^{\varphi,\psi}\left(\sum_{t=0}^{\infty}\alpha^t\int_X u^*(y)Q(dy|x_t, a_t, b_t)\right) \\ &\quad - V_\alpha(x, \varphi, \psi, u^*) - \xi^*/(1-\alpha) \\ &= V_\alpha(x, \varphi, \psi) + E_x^{\varphi,\psi}\sum_{t=0}^{\infty}\alpha^t u^*(x_{t+1}) - V_\alpha(x, \varphi, \psi, u^*) \\ &\quad - \xi^*/(1-\alpha) \\ &= V_\alpha(x, \varphi, \psi) + \frac{1-\alpha}{\alpha}V_\alpha(x, \varphi, \psi, u^*) - \frac{1}{\alpha}u^*(x) - \xi^*/(1-\alpha). \end{aligned}$$

□

Lema 3.3.4. *Sea $(\varphi^*, \psi^*) \in \Phi_1 \times \Phi_2$ un equilibrio promedio. Bajo los Supuestos 2.3.1, 2.3.2 y 2.3.3, se satisfacen*

1. $\lim\limits_{\alpha\to 1}\sup\limits_{\varphi\in\Phi_1}(1-\alpha)V_\alpha(x, \varphi, \psi^*) \leq \xi^*$

2. $\lim\limits_{\alpha\to 1}\inf\limits_{\psi\in\Phi_2}(1-\alpha)V_\alpha(x, \varphi^*, \psi) \geq \xi^*$

Demostración. Sea $(\varphi^*, \psi^*) \in \Phi_1 \times \Phi_2$ un equilibrio promedio.

1) De la Proposición 3.1.6, $\Delta(x, \varphi, \psi^*(x)) \leq 0$ para todo $\varphi \in \Phi_1$. Por tanto $V_\alpha(x, \varphi, \psi^*, \Delta) \leq 0$. Luego, del Lema 3.3.3, tenemos

$$\sup_{\varphi\in\Phi_1}(1-\alpha)V_\alpha(x, \varphi, \psi^*) \leq \xi^* + \frac{1-\alpha}{\alpha}u^*(x) - \inf_{\varphi\in\Phi_1}\frac{(1-\alpha)^2}{\alpha}V_\alpha(x, \varphi, \psi^*, u^*).$$

Finalmente, considerando la Proposición 3.1.5 y el Lema 3.2.1, aplicamos el límite $\alpha \to 1$ en ambos lados de la desigualdad y obtenemos lo deseado.

2) Se demuestra de manera análoga al caso anterior. □

Teorema 3.3.5. *Asumamos que los Supuestos 2.3.1, 2.3.2 y 2.3.4 se cumplen. Entonces, una estrategia $(\varphi, \psi) \in \Phi_1 \times \Phi_2$ es un equilibrio promedio esperado si y sólo si es equilibrio -1-descontado fuerte.*

Demostración. Sea $(\varphi^*, \psi^*) \in \Phi_1 \times \Phi_2$ un equilibrio promedio. Del Teorema 3.1.5 y los Lemas 3.3.2 y 3.3.4,

$$\begin{aligned}\lim_{\alpha\to 1}(1-\alpha)[V_\alpha(x,\varphi^*,\psi^*) - \sup_{\varphi\in\Phi_1} V_\alpha(x,\varphi,\psi^*)] &= \lim_{\alpha\to 1}(1-\alpha)V_\alpha(x,\varphi^*,\psi^*) \\ &\quad - \lim_{\alpha\to 1}\sup_{\varphi\in\Phi_1}(1-\alpha)V_\alpha(x,\varphi,\psi^*) \\ &\geq \mu_{\varphi^*,\psi^*}(r) - \xi^* \\ &= 0.\end{aligned}$$

Por lo tanto $\lim_{\alpha\to 1}(1-\alpha)[V_\alpha(x,\varphi^*,\psi^*) - \sup_{\varphi\in\Phi_1} V_\alpha(x,\varphi,\psi^*)] = 0$. La otra igualdad se obtiene de manera similar.

Recíprocamente, sea $(\varphi^*, \psi^*) \in \Phi_1 \times \Phi_2$ un equilibrio -1-descontado. De la igualdad (3.17) y los Lemas 3.3.2 y 3.3.4(1),

$$\begin{aligned}|\mu_{\varphi^*,\psi^*}(r) - \xi^*| &\leq |\lim_{\alpha\to 1}(1-\alpha)V_\alpha(x,\varphi^*,\psi^*) - \xi^*| \\ &\leq |\lim_{\alpha\to 1}(1-\alpha)[V_\alpha(x,\varphi^*,\psi^*) - \sup_{\varphi\in\Phi_1} V_\alpha(x,\varphi,\psi^*)]| \\ &\quad + |\lim_{\alpha\to 1}\sup_{\varphi\in\Phi_1}(1-\alpha)V_\alpha(x,\varphi,\psi^*) - \xi^*| \\ &= \xi^* - \lim_{\alpha\to 1}\sup_{\varphi\in\Phi_1}(1-\alpha)V_\alpha(x,\varphi,\psi^*) \\ &\leq \xi^* - \lim_{\alpha\to 1}(1-\alpha)V_\alpha(x).\end{aligned}$$

De manera similar, usando la igualdad (3.18) y los Lemas 3.3.2 y 3.3.4(2), obtenemos $|\mu_{\varphi^*,\psi^*}(r) - \xi^*| \leq \lim_{\alpha\to 1}(1-\alpha)V_\alpha(x) - \xi^*$. Por lo tanto

$$\mu_{\varphi^*,\psi^*}(r) = \xi^* = \lim_{\alpha\to 1}(1-\alpha)V_\alpha(x). \tag{3.21}$$

Por el Teorema 3.1.5, concluimos que (φ^*, ψ^*) es un equilibrio promedio. □

3.4. Conclusiones

En este capítulo hemos introducido cuatro criterios de optimalidad: optimalidad promedio, optimalidad canónica, optimalidad promedio F-fuerte y optimalidad -1-descontado fuerte. Llegamos a la conclusión de que estos cuatro criterios son equivalentes (Teorema 3.1.5, 3.1.8, 3.3.5), aunque con diferentes definiciones e interpretaciones. Por otra parte, también demostramos

que el valor promedio del juego es independiente del estado inicial (Teorema 3.1.5). Más aun, la relación (3.21) muestra que que el valor promedio del juego puede ser aproximado por la función de valor descuento. En la literatura a este grupo de criterios se le conocen como criterios no selectivos, esto debido a que el límite en promedio ignora los pagos en etapas finitas. Por el contrario, en el siguiente capítulos consideraremos algunos de los criterios que sí son sensibles a etapas finitas.

Capítulo 4

Criterios sensibles

En este capítulo estudiaremos tres criterios adicionales: *optimalidad rebasante*, *optimalidad en sesgo* y *optimalidad 0-descontado*. Estos criterios han sido ampliamente estudiados para procesos de control markovianos (PCMs) a tiempo discreto y tiempo continuo, ver por ejemplo, [48, 27, 32, 28, 33]. Bajo condiciones usuales, se sabe que para PCMs la optimalidad rebasante y la optimalidad en sesgo son equivalentes [32]. Sin embargo, esa relación no se preserva en juegos markovianos a tiempo continuo [67]. Del mismo modo, mostraremos que no se satisface para juegos markovianos a tiempo discreto. Por otra parte, el criterio de optimalidad 0-descontada se ha estudiado para PCMs en [33] pero aún no para juegos markovianos a tiempo discreto.

4.1. Optimalidad rebasante

Definición 4.1.1. *Un par de estrategias* $(\varphi^*, \psi^*) \in \Phi_1 \times \Phi_2$ *se dice que es un* equilibrio rebasante *si para cada* $x \in X$ *y* $(\varphi, \psi) \in \Phi_1 \times \Phi_2$,

$$\liminf_{n\to\infty}[J_n(x, \varphi^*, \psi^*) - J_n(x, \varphi, \psi^*)] \geq 0, \tag{4.1}$$

y

$$\limsup_{n\to\infty}[J_n(x, \varphi^*, \psi^*) - J_n(x, \varphi^*, \psi)] \leq 0. \tag{4.2}$$

A continuación veremos dos criterios que también guardan una relación con la optimalidad rebasante: optimalidad promedio D-fuerte y optimalidad en costo de oportunidad. Estos dos criterios fueron introducidos originalmente por Dutta [15] y Flynn [18], respectivamente.

Para cada $(\varphi, \psi) \in \Phi_1 \times \Phi_2$, definamos las funciones de pago para los jugadores 1 y 2:

$$D_1(x, \varphi, \psi) = \liminf_{n\to\infty}\left[J_n(x, \varphi, \psi) - n \sup_{\varphi\in\Phi_1} J(x, \varphi, \psi)\right],$$

$$D_2(x,\varphi,\psi) = \limsup_{n\to\infty}\left[J_n(x,\varphi,\psi) - n\inf_{\psi\in\Phi_2} J(x,\varphi,\psi)\right].$$

En este criterio los jugadores comparan su beneficio/costo en cada etapa con el valor óptimo de la función objetivo promedio. En este caso están interesados en encontrar estrategias que mejoren ese promedio.

Definición 4.1.2. *El par de estrategias* $(\varphi^*,\psi^*)\in\Phi_1\times\Phi_2$ *es un* equilibrio promedio D-fuerte *si para cada* $x\in X$*,*

$$D_1(x,\varphi^*,\psi^*) = \sup_{\varphi\in\Phi_1} D_1(x,\varphi,\psi^*) \tag{4.3}$$

y

$$D_2(x,\varphi^*,\psi^*) = \inf_{\phi\in\Phi_2} D_2(x,\varphi^*,\psi). \tag{4.4}$$

El siguiente teorema nos muestra que los equilibrios rebasantes se pueden obtener con este criterio.

Teorema 4.1.3. *Si* (φ^*,ψ^*) *es un equilibrio rebasante, entonces* (φ^*,ψ^*) *es un equilibrio promedio D-fuerte.*

Demostración. Supongamos que (φ^*,ψ^*) es un equilibrio rebasante, entonces

$$\begin{aligned}
D_1(x,\varphi^*,\psi^*) =& \liminf_{n\to\infty}\left[J_n(x,\varphi^*,\psi^*) - n\sup_{\varphi\in\Phi_1} J(x,\varphi,\psi^*)\right]\\
\geq& \liminf_{n\to\infty}\left[J_n(x,\varphi^*,\psi^*) - J_n(x,\varphi,\psi^*)\right]\\
&+\liminf_{n\to\infty}\left[J_n(x,\varphi,\psi^*) - n\sup_{\varphi\in\Phi_1} J(x,\varphi,\psi^*)\right]\\
\geq& D_1(x,\varphi,\psi^*).
\end{aligned}$$

$$\begin{aligned}
D_2(x,\varphi^*,\psi^*) =& \limsup_{n\to\infty}\left[J_n(x,\varphi^*,\psi^*) - n\inf_{\psi\in\Phi_2} J(x,\varphi^*,\psi)\right]\\
\leq& \limsup_{n\to\infty}\left[J_n(x,\varphi^*,\psi^*) - J_n(x,\varphi^*,\psi)\right]\\
&+\limsup_{n\to\infty}\left[J_n(x,\varphi^*,\psi) - n\inf_{\psi\in\Phi_2} J(x,\varphi^*,\psi)\right]\\
\leq& D_2(x,\varphi^*,\psi).
\end{aligned}$$

Por lo tanto, (φ^*,ψ^*) es un equilibrio promedio D-fuerte. □

El recíproco del anterior teorema no se satisface, a menos que se impongan condiciones adicionales.

Teorema 4.1.4. *Supongamos que* $(\varphi^*, \psi^*) \in \Phi_1 \times \Phi_2$ *es un equilibrio promedio D-fuerte y, además, satisface que*

$$D_1(x, \varphi, \psi^*) = \limsup_{n\to\infty} \left[J_n(x, \varphi, \psi^*) - n \sup_{\varphi\in\Phi_1} J(x, \varphi, \psi^*) \right],$$
$$D_2(x, \varphi^*, \psi) = \liminf_{n\to\infty} \left[J_n(x, \varphi^*, \psi) - n \inf_{\psi\in\Phi_2} J(x, \varphi^*, \psi) \right]$$

para todo $x \in X$ *y* $(\varphi, \psi) \in \Phi_1 \times \Phi_2$. *Entonces* (φ^*, ψ^*) *es un equilibrio rebasante.*

Demostración. Supongamos que (φ^*, ψ^*) es un equilibrio promedio D-fuerte que satisface las hipótesis del teorema, entonces

$$\begin{aligned}
\liminf_{n\to\infty}[J_n(x, \varphi^*, \psi^*) - J_n(x, \varphi, \psi^*)] \geq & \\
&\geq \liminf_{n\to\infty} \left[J_n(x, \varphi^*, \psi^*) - n \sup_{\varphi\in\Phi_1} J(x, \varphi, \psi^*) \right] \\
&\quad - \limsup_{n\to\infty} \left[J_n(x, \varphi, \psi^*) - n \sup_{\varphi\in\Phi_1} J(x, \varphi, \psi^*) \right] \\
&= D_1(x, \varphi^*, \psi^*) - D_1(x, \varphi, \psi^*) \geq 0.
\end{aligned}$$

Por otra parte,

$$\begin{aligned}
\limsup_{n\to\infty}[J_n(x, \varphi^*, \psi^*) - J_n(x, \varphi^*, \psi)] \leq & \\
&\leq \limsup_{n\to\infty} \left[J_n(x, \varphi^*, \psi^*) - n \inf_{\psi\in\Phi_2} J(x, \varphi^*, \psi) \right] \\
&\quad - \liminf_{n\to\infty} \left[J_n(x, \varphi^*, \psi) - n \sup_{\psi\in\Phi_2} J(x, \varphi^*, \psi) \right] \\
&= D_2(x, \varphi^*, \psi^*) - D_2(x, \varphi^*, \psi) \leq 0.
\end{aligned}$$

Así, concluimos que (φ^*, ψ^*) es un equilibrio rebasante. □

Ahora, para introducir el siguiente criterio primero definimos nuevas funciones objetivo para cada uno de los jugadores:

$$OC_1(x, \varphi, \psi) := \liminf_{n\to\infty} \left[J_n(x, \varphi, \psi) - \sup_{\varphi\in\Phi_1} J_n(x, \varphi, \psi) \right],$$

$$OC_2(x, \varphi, \psi) := \limsup_{n\to\infty} \left[J_n(x, \varphi, \psi) - \inf_{\psi\in\Phi_2} J_n(x, \varphi, \psi) \right].$$

El criterio que en seguida definiremos, compara en cada etapa el beneficio acumulado con el optimo beneficio acumulado hasta esta esa misma etapa. Los jugadores buscan estrategias que optimicen esa diferencia.

Definición 4.1.5. *Diremos que el par de estrategias* $(\varphi^*, \psi^*) \in \Phi_1 \times \Phi_2$ *es un* equilibrio-CO *(costo de oportunidad) si para cada* $x \in X$*,*

$$OC_1(x, \varphi^*, \psi^*) = \sup_{\varphi \in \Phi_1} OC_1(x, \varphi, \psi^*) \tag{4.5}$$

y

$$OC_2(x, \varphi^*, \psi^*) = \inf_{\psi \in \Phi_2} OC_2(x, \varphi^*, \psi). \tag{4.6}$$

De nuevo,el siguiente teorema nos dice que los equilibrios rebasantes también forman parte de las soluciones para el criterio que estamos considerando.

Teorema 4.1.6. *Si* (φ^*, ψ^*) *es un equilibrio rebasante, entonces* (φ^*, ψ^*) *es un equilibrio-CO.*

Demostración. Supongamos que (φ^*, ψ^*) es un equilibrio rebasante, entonces

$$\begin{aligned} OC_1(x, \varphi^*, \psi^*) =& \liminf_{n\to\infty} \left[J_n(x, \varphi^*, \psi^*) - \sup_{\varphi \in \Phi_1} J_n(x, \varphi, \psi^*) \right] \\ \geq& \liminf_{n\to\infty} \left[J_n(x, \varphi^*, \psi^*) - J_n(x, \varphi, \psi^*) \right] \\ &+ \liminf_{n\to\infty} \left[J_n(x, \varphi, \psi^*) - \sup_{\varphi \in \Phi_1} J_n(x, \varphi, \psi^*) \right] \\ \geq& OC_1(x, \varphi, \psi^*) \end{aligned}$$

y

$$\begin{aligned} OC_2(x, \varphi^*, \psi^*) =& \limsup_{n\to\infty} \left[J_n(x, \varphi^*, \psi^*) - \inf_{\psi \in \Phi_2} J_n(x, \varphi^*, \psi) \right] \\ \leq& \limsup_{n\to\infty} \left[J_n(x, \varphi^*, \psi^*) - J_n(x, \varphi^*, \psi) \right] \\ &+ \limsup_{n\to\infty} \left[J_n(x, \varphi^*, \psi) - \inf_{\psi \in \Phi_2} J_n(x, \varphi^*, \psi) \right] \\ \leq& OC_2(x, \varphi^*, \psi) \end{aligned}$$

para todo $(\varphi, \psi) \in \Phi_1 \times \Phi_2$. Por lo tanto (φ^*, ψ^*) es un equilibrio-CO. □

El recíproco de este resultado se cumple, similar a la optimalidad promedio D-fuerte, solamente bajo ciertas condiciones adicionales.

Teorema 4.1.7. *Supongamos que* $(\varphi^*, \psi^*) \in \Phi_1 \times \Phi_2$ *es un equilibrio-CO y que, además, satisface*

$$\begin{aligned} OC_1(x, \varphi, \psi^*) =& \limsup_{n\to\infty} \left[J_n(x, \varphi, \psi^*) - \sup_{\varphi \in \Phi_1} J_n(x, \varphi, \psi^*) \right], \\ OC_2(x, \varphi^*, \psi) =& \liminf_{n\to\infty} \left[J_n(x, \varphi^*, \psi) - \inf_{\psi \in \Phi_2} J_n(x, \varphi^*, \psi) \right] \end{aligned}$$

para todo $x \in X$ *y* $(\varphi, \psi) \in \Phi_1 \times \Phi_2$. *Entonces* (φ^*, ψ^*) *es un equilibrio rebasante.*

Demostración. Si (φ^*, ψ^*) es un equilibrio-CO que cumple las hipótesis del teorema, vemos que

$$\begin{aligned}\liminf_{n\to\infty}[J_n(x,\varphi^*,\psi^*) - J_n(x,\varphi,\psi^*)] \geq \\ \geq \liminf_{n\to\infty}\left[J_n(x,\varphi^*,\psi^*) - \sup_{\varphi\in\Phi_1} J_n(x,\varphi,\psi^*)\right] \\ - \limsup_{n\to\infty}\left[J_n(x,\varphi,\psi^*) - \sup_{\varphi\in\Phi_1} J_n(x,\varphi,\psi^*)\right] \\ = OC_1(x,\varphi^*,\psi^*) - OC_1(x,\varphi,\psi^*) \geq 0\end{aligned}$$

y

$$\begin{aligned}\liminf_{n\to\infty}[J_n(x,\varphi^*,\psi) - J_n(x,\varphi^*,\psi^*)] \geq \\ \geq \liminf_{n\to\infty}\left[J_n(x,\varphi,\psi^*) - \inf_{\psi\in\Phi_2} J_n(x,\varphi^*,\psi)\right] \\ - \limsup_{n\to\infty}\left[J_n(x,\varphi^*,\psi^*) - \sup_{\psi\in\Phi_2} J_n(x,\varphi^*,\psi)\right] \\ = OC_2(x,\varphi^*,\psi) - OC_2(x,\varphi^*,\psi^*) \geq 0\end{aligned}$$

para todo $(\varphi, \psi) \in \Phi_1 \times \Phi_2$. Por lo tanto (φ^*, ψ^*) es un equilibrio rebasante. □

4.2. Optimalidad en sesgo

Para cada $(\varphi, \psi) \in \Phi_1 \times \Phi_2$ definiremos su *función de sesgo* por

$$h(x,\varphi,\psi) := \lim_{n\to\infty}[J_n(x,\varphi,\psi) - nJ(\varphi,\psi)] \tag{4.7}$$

donde $J(\varphi, \psi)$ es como en la Proposición 3.1.3. Es importante notar que del Lema 3.1.2 sabemos que $|h(x,\varphi,\psi)| \leq \|r\|_w M w(x)/(1-\theta)$. Por lo tanto, la función de sesgo está bien definida y pertenece al espacio $\mathbb{B}_w(X)$.

La siguiente proposición es una caracterización de la función de sesgo como una solución de una cierta ecuación de Poisson.

Proposición 4.2.1. *Para cada par de estrategias* $(\varphi, \psi) \in \Phi_1 \times \Phi_2$, *el par* $(J(\varphi,\psi), h(\cdot,\varphi,\psi))$ *es la única solución a la ecuación de Poisson*

$$\begin{aligned}J(\varphi,\psi) + h(x,\varphi,\psi) =& r(x,\varphi(x),\psi(x)) \\ &+ \int_X h(y,\varphi,\psi)Q(dy|x,\varphi(x),\psi(x))\end{aligned} \tag{4.8}$$

y que satisface la condición $\int_X h(x,\varphi,\psi)\mu_{\varphi,\psi}(dx) = 0$. *Más aun, para cada solución* u^* *del Teorema 3.1.4,* $h(x,\varphi,\psi) = u^*(x) - \mu_{\varphi,\psi}(u^*)$.

Demostración. Ver la Proposición 10.2.3 en [28]. □

Nota 4.2.2. De (2.19) y (4.7) tenemos la relación

$$J_n(x,\varphi,\psi) = J(\varphi,\psi)n + h(x,\varphi,\psi) + O(\theta^n) \tag{4.9}$$

cuando $n \to \infty$ para todo $x \in X$ y $(\varphi,\psi) \in \Phi_1 \times \Phi_2$

Definición 4.2.3. *Un par de estrategias* $(\varphi^*,\psi^*) \in \Phi_1 \times \Phi_2$ *se dice que es* equilibrio en sesgo *si*

$$h(x,\varphi,\psi^*) \leq h(x,\varphi^*,\psi^*) \leq h(x,\varphi^*,\psi) \tag{4.10}$$

para cada $x \in X$ *y todo equilibrio promedio* $(\varphi,\psi) \in \Phi_1 \times \Phi_2$.

Teorema 4.2.4. *Demos por hecho que los Supuestos 2.3.1, 2.3.2 y 2.3.3 se cumplen. Si el par de estrategias* $(\varphi^*,\psi^*) \in \Phi_1 \times \Phi_2$ *es un equilibrio rebasante entonces es también un equilibrio en sesgo.*

Demostración. Sea $(\varphi^*,\psi^*) \in \Phi_1 \times \Phi_2$ un equilibrio rebasante. De (4.9) obtenemos

$$\begin{aligned} J_n(x,\varphi,\psi^*) - J_n(x,\varphi^*,\psi^*) =& [J(\varphi,\psi^*) - J(\varphi^*,\psi^*)]n \\ &+ h(x,\varphi,\psi^*) - h(x,\varphi^*,\psi^*) \end{aligned}$$

y

$$\begin{aligned} J_n(x,\varphi^*,\psi) - J_n(x,\varphi^*,\psi^*) =& [J(\varphi^*,\psi) - J(\varphi^*,\psi^*)]n \\ &+ h(x,\varphi^*,\psi) - h(x,\varphi^*,\psi^*). \end{aligned}$$

De la Definición 4.1.1 deducimos que

$$\liminf_{n\to\infty}[(J(\varphi,\psi^*) - J(\varphi^*,\psi^*))n + h(x,\varphi,\psi^*) - h(x,\varphi^*,\psi^*)] \leq 0 \tag{4.11}$$

y

$$\limsup_{n\to\infty}[(J(\varphi^*,\psi) - J(\varphi^*,\psi^*))n + h(x,\varphi^*,\psi) - h(x,\varphi^*,\psi^*)] \geq 0 \tag{4.12}$$

para todo $(\varphi,\psi) \in \Phi_1 \times \Phi_2$ y $x \in X$. Dividiendo por n y aplicando límites concluimos que (φ^*,ψ^*) es un equilibrio promedio. Por otra parte, si (φ,ψ) es es un equilibrio promedio entonces $J(\varphi,\psi^*) = J(\varphi^*,\psi^*) = J(\varphi^*,\psi)$. Por tanto, las desigualdades 4.11 y 4.12 muestran que (φ^*,ψ^*) es un equilibrio en sesgo. □

El recíproco del teorema anterior solamente se cumple de manera parcial.

Teorema 4.2.5. *Asumamos que los Supuestos 2.3.1, 2.3.2 y 2.3.3 se cumplen. Si el par de estrategias* $(\varphi^*, \psi^*) \in \Phi_1 \times \Phi_2$ *es un equilibrio en sesgo entonces es un equilibrio rebasante en el conjunto de los equilibrios promedio.*

Demostración. Sea $(\varphi, \psi) \in \Phi_1 \times \Phi_2$ un equilibrio promedio. De la relación (4.9), podemos ver que

$$J_n(x, \varphi, \psi) = V^* n + h(x, \varphi, \psi) + O(\theta^n) \text{ for all } x \in X.$$

Por lo tanto, de acuerdo a la Definición, 4.10

$$\liminf_{n\to\infty}[J_n(x, \varphi^*, \psi^*) - J_n(x, \varphi, \psi^*)] = h(x, \varphi^*, \psi^*) - h(x, \varphi, \psi^*) \geq 0$$

y

$$\limsup_{n\to\infty}[J_n(x, \varphi^*, \psi^*) - J_n(x, \varphi, \psi^*)] = h(x, \varphi^*, \psi^*) - h(x, \varphi^*, \psi) \leq 0$$

para todo $x \in X$. Lo cual demuestra que (φ^*, ψ^*) es un equilibrio rebasante. □

El siguiente ejemplo nos ilustra la diferencia entre el conjunto de los equilibrios en sesgo y los rebasantes.

Ejemplo 4.2.6. Consideremos el espacio de estados $X = \{0, 1\}$, el conjunto de acciones $A = \{0, 1\}$ para el jugador 1 y $B = \{0, 1\}$ para el jugador 2. Supongamos que en el estado cero los jugadores están restringidos a usar únicamente la acción 0, es decir $A(0) = \{0\}$ y $B(0) = \{0\}$. En el estado 1 ambos jugadores disponen de las dos acciones, $A(1) = \{0, 1\}$ y $B(1) = \{0, 1\}$. La función de pagos para el jugador 1 es como sigue: $r(0, 0, 0) = 4$, $r(1, 0, 0) = 1$, $r(1, 0, 1) = 0$, $r(1, 1, 0) = -2$ y $r(1, 1, 1) = 2$. Mientras que la ley de transición del juego es dado por: $Q(0|0, 0, 0) = Q(1|1, 0, 1) = Q(1|1, 1, 0) = Q(1|1, 1, 1) = 1/2$ y $Q(0|1, 0, 0) = 1/4$. No es difícil verificar que este juego cumple con los Supuestos 2.3.1, 2.3.2 y 2.3.3.

Sea $(\varphi, \psi) \in \Phi_1 \times \Phi_2$ una estrategia estacionaria. Debido a que los jugadores solamente tienen una opción en el estado 0, $\varphi(0|0) = \psi(0|0) = 1$. En cambio, en el estado 2, $\varphi(0|1) = \alpha$ y $\psi(0|1) = \beta$ para algún par $(\alpha, \beta) \in [0, 1] \times [0, 1]$. Con lo anterior en mente, el pago esperado (2.16) se reduce a

$$r(0, \varphi(0), \psi(0)) = 4 \text{ y } r(1, \varphi(1), \psi(1)) = 5\alpha\beta - 2\alpha - 4\beta + 2. \qquad (4.13)$$

Mientras la matriz de la ley de transición es dado por

$$Q(\varphi,\psi)=\begin{pmatrix} 1/2 & 1/2 \\ (2-\alpha\beta)/4 & (2+\alpha\beta)/4 \end{pmatrix}, \tag{4.14}$$

y por tanto la medida de probabilidad invariante es

$$\mu_{\varphi,\psi}\{0\}=\frac{2-\alpha\beta}{4-\alpha\beta} \quad \text{y} \quad \mu_{\varphi,\psi}\{1\}=\frac{2}{4-\alpha\beta}. \tag{4.15}$$

De (3.4) y (3.2), la función de pago promedio total es

$$J(\varphi,\psi)=\frac{6\alpha\beta-4\alpha-8\beta+12}{4-\alpha\beta}.$$

Analizando dicha función encontramos que el valor del juego es

$$V(x)=\sup_{0\leq\alpha\leq 1}\inf_{0\leq\beta\leq 1}J(\alpha,\beta)=2$$

y el conjunto de los equilibrios promedio es

$$(\Phi_1\times\Phi_2)_{\text{ae}}=\{(1,\beta),1/2\leq\beta\leq 1\}.$$

Por otra parte, de la Proposición 4.2.1, la función de sesgo para un equilibrio promedio $(1,\beta)\in(\Phi_1\times\Phi_2)_{\text{ae}}$ esta dado por

$$h(x,1,\beta)=\begin{cases} \frac{8}{4-\beta} \text{ si } x=0, \\ \frac{4(\beta-2)}{4-\beta} \text{ si } x=1. \end{cases} \tag{4.16}$$

Por lo tanto, minimizando en β, vemos que el único equilibrio en sesgo es la estrategia $(1,1/2)$.

Por otra parte, sabemos del Teorema 4.2.5 sabemos que todo equilibrio rebasante es un equilibrio en sesgo, pero el recíproco en general no es verdadero. De hecho, el juego que estamos considerando no tiene equilibrios rebasantes. Para verificar esta afirmación comparemos el único equilibrio en sesgo con otra estrategia, por ejemplo $(1,0)$. Vemos que

$$\limsup_{T\to\infty}[J_T(x,1,1/2)-J_T(x,1,0)]=h(x,1,1/2)-h(x,1,0)=2/7.$$

Por tanto no cumple la condición para ser un equilibrio rebasante (4.2).

4.3. Equilibrio n-descontado

En esta sección, nuevamente, definiremos una familia más amplia de criterios que involucran descuento. Estos criterios de decisión se han introducido originalmente para estudiar modelos económicos dinámicos; ver Veinott [74, 75]. Mostraremos que algunos de los criterios definidos en este capítulo también quedan como un caso particular de este.

Definición 4.3.1. *Para $n = -1, 0, 1, 2, ...$, diremos que la estrategia (φ^*, ψ^*) en $\Phi_1 \times \Phi_2$ es un* equilibrio $n-$descontado *si*

$$\liminf_{\alpha\to 1}(1-\alpha)^{-n}[V_\alpha(x,\varphi^*,\psi^*) - V_\alpha(x,\varphi,\psi^*)] \geq 0 \tag{4.17}$$

$$\limsup_{\alpha\to 1}(1-\alpha)^{-n}[V_\alpha(x,\varphi^*,\psi^*) - V_\alpha(x,\varphi^*,\psi)] \leq 0 \tag{4.18}$$

para todo $x \in X$ y $(\varphi, \psi) \in \Phi_1 \times \Phi_2$.

El siguiente teorema nos muestra que los equilibrios en sesgo, incluyendo a los equilibrios rebasantes (Teorema 4.2.4), se pueden obtener como solución a un caso particular de equilibrios que involucran descuento.

Teorema 4.3.2. *Bajo los Supuestos 2.3.1, 2.3.2 y 2.3.4, todo equilibrio en sesgo es un equilibrio* 0*-descontado.*

Demostración. Sea (φ^*, ψ^*) un equilibrio en sesgo. Del Lema 3.3.3,

$$\begin{aligned} V_\alpha(x,\varphi^*,\psi^*) - V_\alpha(x,\varphi,\psi^*) = {} & V_\alpha(x,\varphi^*,\psi^*,\Delta) - V_\alpha(x,\varphi,\psi^*,\Delta) \\ & + \frac{1-\alpha}{\alpha}[V_\alpha(x,\varphi,\psi^*,u^*) - V_\alpha(x,\varphi^*,\psi^*,u^*)]. \end{aligned}$$

Sabemos que, en particular, la estrategia (ψ^*, u^*) es también un equilibrio promedio: entonces $\Delta(x, \psi^*, u^*) = 0$ y $\Delta(x, \varphi, \psi^*) \leq 0$. Por lo tanto

$$V_\alpha(x,\varphi^*,\psi^*) - V_\alpha(x,\varphi,\psi^*) \geq \frac{1-\alpha}{\alpha}[V_\alpha(x,\varphi,\psi^*,u^*) - V_\alpha(x,\varphi^*,\psi^*,u^*)].$$

Por otra parte, del Lema 3.3.2 y la Proposición 4.2.1,

$$\lim_{\alpha\to 1}\frac{1-\alpha}{\alpha}[V_\alpha(x,\varphi,\psi^*,u^*) - V_\alpha(x,\varphi^*,\psi^*,u^*)] = h(x,\varphi^*,\psi^*) - h(x,\varphi,\psi^*).$$

En consecuencia

$$\lim_{\alpha\to 1}[V_\alpha(x,\varphi^*,\psi^*) - V_\alpha(x,\varphi,\psi^*)] \geq h(x,\varphi^*,\psi^*) - h(x,\varphi,\psi^*) \geq 0.$$

De manera similar, $\lim_{\alpha\to 1}[V_\alpha(x,\varphi^*,\psi^*) - V_\alpha(x,\varphi,\psi^*)] \leq 0$. En conclución, (ψ^*, u^*) es un equilibrio 0-descontado. □

El recíproco del anterior teorema no necesariamente se satisface, pero se cumple la siguiente inclusión.

Teorema 4.3.3. *Bajo los Supuestos 2.3.1, 2.3.2 y 2.3.4, todo equilibrio* 0*-descontado fuerte es un equilibrio en sesgo.*

Demostración. Sea (φ^*, ψ^*) un equilibrio 0-descontado fuerte. En particular, (φ^*, ψ^*) es un equilibrio -1-descontado. Por el Teorema 3.3.5, (ψ^*, u^*) es un equilibrio promedio. Es decir, $\Delta(x, \psi^*, u^*) = 0$. Por otra parte, del Lema 3.3.3, para todo $\varphi \in \Phi_1$

$$\begin{aligned} V_\alpha(x, \varphi^*, \psi^*) - \sup_{\varphi \in \Phi_1} V_\alpha(x, \varphi, \psi^*) &\leq V_\alpha(x, \varphi^*, \psi^*) - V_\alpha(x, \varphi, \psi^*) \\ &= \frac{1-\alpha}{\alpha}[V_\alpha(x, \varphi, \psi^*, u^*) - V_\alpha(x, \varphi^*, \psi^*, u^*)] \\ &\quad - V_\alpha(x, \varphi, \psi^*, \Delta). \end{aligned}$$

En particular, sea $\varphi \in \Phi_1$ tal que (φ, ψ^*) es un equilibrio en sesgo. En este caso tendremos $V_\alpha(x, \varphi, \psi^*, \Delta) = 0$. La desigualdad anterior se reduce a

$$V_\alpha(x, \varphi^*, \psi^*) - \sup_{\varphi \in \Phi_1} V_\alpha(x, \varphi, \psi^*) \leq \frac{1-\alpha}{\alpha}[V_\alpha(x, \varphi, \psi^*, u^*) - V_\alpha(x, \varphi^*, \psi^*, u^*)].$$

Aplicando límites en ambos lados obtenemos

$$\lim_{\alpha \to 1}[V_\alpha(x, \varphi^*, \psi^*) - \sup_{\varphi \in \Phi_1} V_\alpha(x, \varphi, \psi^*)] \leq h(x, \varphi^*, \psi^*) - h(x, \varphi, \psi^*).$$

Ahora bien, si suponemos que (φ^*, ψ^*) no es un equilibrio en sesgo, entonces $h(x, \varphi^*, \psi^*) - h(x, \varphi, \psi^*) < 0$. Esto implica que

$$\lim_{\alpha \to 1}[V_\alpha(x, \varphi^*, \psi^*) - \sup_{\varphi \in \Phi_1} V_\alpha(x, \varphi, \psi^*)] < 0.$$

Lo cual es una contradicción a la hipótesis inicial. □

4.4. Conclusiones

En este capítulo consideramos la optimalidad en sesgo, optimalidad rebasante, optimalidad 0-descontada, optimalidad promedio D-fuerte y optimalidad en costo de oportunidad (CO). La relación que existe entre estos criterios se ilustra en la Figura 4.1. En procesos de control markovianos, la optimalidad en sesgo también implica la optimalidad rebasante; el Teorema 4.2.5 nos dá una implicación parcial. Sin embargo, el Ejemplo 4.2.6 muestra que, en definitiva, estos criterios no son equivalente en juegos markovianos.

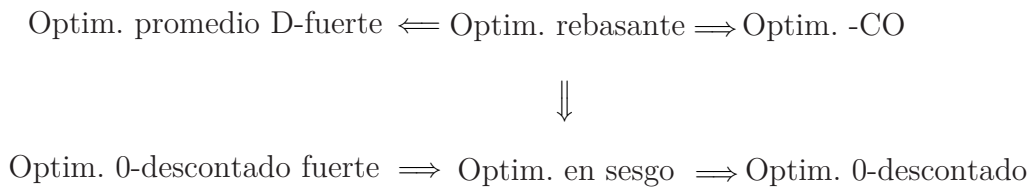
Optim. promedio D-fuerte
Optim. rebasante
Optim. -CO
Optim. 0-descontado fuerte
Optim. en sesgo
Optim. 0-descontado

Figura 4.1.

Capítulo 5

Conclusiones y problemas abiertos

En una primera parte de este trabajo hemos estudiado el modelo forestal de Mitra-Wan bajo cuatro criterios importantes: optimalidad en promedio, optimalidad buena, optimalidad rebasante y optimalidad en sesgo. Concluimos que las políticas buenas están incluidas dentro del conjunto de políticas optimas en promedio (Fig. 1.2). Por otra parte, todas las trayectoria correspondientes a las políticas buenas convergen a la política estacionaria óptima (Teorema 1.3.6). El valor de las políticas óptimas en promedio es independiente del estado inicial y es igual al valor de la política estacionaria óptima (Teorema 1.3.2). También hemos demostrado que las políticas óptimas en sesgo, las optimas rebasantes y las que tienen mínima pérdida de valor son equivalentes (Teorema 1.4.7). Todos estos resultados nos dan un mapa claro de las relaciones que existen entre los diferentes criterios de optimalidad y algunas de sus propiedades.

Una de las direcciones naturales para una posterior investigación es caracterizar estructuras generales de modelos de control óptimo (no necesariamente lineales) donde aun se preservan estas propiedades. Por otra parte, hemos visto que todas las políticas buenas satisfacen las propiedades de tipo autopista, pero no es claro si las políticas optimas en promedio también poseen esa propiedad.

Para juegos markovianos hemos considerado dos grupos de criterios de optimalidad: los criterios de optimalidad en promedio y los criterios de optimalidad que son sensibles a tiempo finito. En el primer grupo conformado por equilibrios en promedio esperado, equilibrios canónicos, promedio F-fuerte, 1-descontado. Demostramos que todos estos criterios son equivalentes (Teoremas 3.1.5, 3.1.8 y 3.3.5). De manera similar al modelo de Mitra-Wan, el valor promedio es independiente del estado inicial (Teorema 3.1.5). El segun-

do grupo esta compuesto por el criterio de optimalidad rebasante, en sesgo, 1-descontado. Al contrario del modelo de Mitra-Wan y procesos de control markoviano, la optimalidad en sesgo no es equivalente a la optimalidad rebasante; el conjunto de los equilibrios rebasantes está estrictamente contenido en el conjunto de equilibrios en sesgo (Teorema 4.2.4 y Ejemplo 4.2.6). Por otra parte, el conjunto de los equilibrios en sesgo está contenido en el conjunto de los equilibrios 0-descontados (Teorema 4.3.2).

Ahora bien, con respecto a las propiedades de tipo turnpike para juegos markovianos, existen pocos trabajos resultados y es una de las tareas a realizar posteriormente. Otra tarea pendiente es investigar los mismos criterios de optimalidad pero para juegos de suma no necesariamente cero.

Apéndice A

Teoremas de ergodicidad

El propósito de este apéndice es presentar algunos teoremas importantes sobre la ergodicidad de cadenas de Markov. Este tipo de teoremas ha sido estudiado por mucho autores, en particular por T. Harris [25] y por S. Meyn y R. Tweedie [54, 55]. Estos teoremas aseguran que, entre otras cosas, dada una probabilidad de transición con ciertas condiciones, existe una única medida invariante y, además, se puede estimar la rapidez de convergencia hacia esta medida. Recientemente, M. Hairer y J. Mattingly [24] dieron una demostración más simplificada de estos resultados. Mayores de detalles sobre este tipo de teoremas se pueden ver en los trabajos [2, 10, 31, 4, 3].

A.1. Conceptos de recurrencia

Sea $(X, \mathcal{B})$ un espacio medible, donde X es un espacio métrico separable y $\mathcal{B}$ es el $\sigma-$álgebra de Borel.

Consideremos una cadena de Markov $\{x_t, t = 0, 1, 2, ...\}$ en X con probabilidad de transición $P(B|x)$, es decir

$$P(B|x) = \text{Prob}(x_{t+1} \in B|x_t = x)$$

para todo $t = 0, 1, ..., B \in \mathcal{B}$ y $x \in X$. Sea $P^t(B|x)$ la probabilidad de que la cadena iniciando en x llega al conjunto B en la etapa t. La sucesión de probabilidades de transición $\{P^t(B|x), x \in X, B \in \mathcal{B}\}$ es tal que para cada $t, n = 0, 1, 2, ...$

$$P^t(B|x) = P(x_{t+n} \in B|x_j, j \leq n; x_n = x).$$

La independencia de P^t de los valores $x_j, j \leq n$, es la propiedad de Markov, y la independencia de P^t y n es la propiedad de homogeneidad en el tiempo.

Dado $B \in \mathcal{B}$, el *tiempo de ocupación* del conjunto B de la cadena se define como

$$\eta_B := \sum_{t=1}^{\infty} I_B(x_t),$$

donde I_B denota la función indicadora de B. Es decir, η_B es el "número de veces"que la cadena está en B.

Si μ es una medida de probabilidad sobre X, denotaremos por E_μ a la esperanza condicional dado que la distribución del estado inicial x_0 es μ. En particular, si μ se concentra en un sólo punto $x \in X$, la esperanza se denota por E_x.

Definición A.1.1. *Sea λ una medida $\sigma-$finita sobre $\mathcal{B}$. Se dice que la cadena de Markov $\{x_t, t = 0, 1, 2, ...\}$ (o la probabilidad de transición P) es:*

1. $\lambda-$irreducible *si $\lambda(B) > 0$ implica que*

$$E_x(\eta_B) = \sum_{t=1}^{\infty} P^t(B|x) > 0 \text{ para todo } x \in X,$$

2. Harris-recurrente *si $\lambda(B) > 0$ implica que*

$$P_x(\nu_B > 0) = P_x(x_t \in B \text{ para algún } t \geq 1) = 1 \text{ para todo } x \in X.$$

Nota A.1.2. Si existe una medida $\sigma-$finita λ sobre X y una función de densidad estrictamente positiva $g(x, \cdot)$ tal que

$$P(D|x) = \int_D g(x, y)\lambda(dy),$$

entonces se ve inmediatamente que P es $\lambda-$irreducible.

A.2. Medidas invariantes

Definición A.2.1. *Sea $\{x_t, t = 0, 1, ...\}$ una cadena de Markov con probabilidad de transición P. Una medida $\sigma-$finita μ en $\mathcal{B}(X)$ se dice que es* invariante *para P si*

$$\mu(B) = \int_X P(B|x)\mu(dx) \tag{A.1}$$

para todo $B \in \mathcal{B}$.

Una probabilidad de transición se dice que es *Harris recurrente positivo* si es Harris recurrente y admite una medida de probabilidad invariante (necesariamente única). En [55, Teorema 13.3.3] y [29, Teorema 1.1] se dan algunas caracterizaciones de este tipo de cadenas de Markov.

Los siguientes resultados muestran algunas de las propiedades que poseen las cadenas de Markov que tienen una medida de probabilidad invariante. La demostración de estos resultados se puede ver en [31] y sus referencias.

Teorema A.2.2. *Si $\{x_t, t = 0, 1, ...\}$ es una cadena de Markov $\lambda-$irreducible con medida de probabilidad invariante μ y ν es su distribución inicial, entonces, para cada función medible no negativa v tal que $\mu(v) := \int v d\mu < \infty$ (ver 2.2),*

1. $\lim_{n\to\infty} \frac{1}{n} E_x \left[\sum_{t=0}^{n-1} v(x_t)\right] = \mu(v)$ μ*–c.s. y*

2. $\lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} v(x_t) = \mu(v)$ P_ν*–c.s.*

Teorema A.2.3. *Supongamos que P es Harris-recurrente. Entonces existe una medida invariante no trivial μ y una tripleta (n, ν, β) que consiste de un entero $n \geq 1$, una medida de probabilidad ν, y una función medible $0 \leq \beta \leq 1$ tal que*

1. *$P^n(B|x) \geq \beta(x)\nu(B)$ para todo $B \in \mathcal{B}$ y $x \in X$;*

2. *$\nu(\beta) > 0$; y*

3. *$0 < \mu(\beta) < \infty$.*

A.3. Teoremas de ergodicidad

En esta sección vemos algunos resultados sobre la convergencia de la sucesión $\{P^t, t = 1, 2, ...\}$ hacia la medida de probabilidad invariante. Estas propiedades han sido estudiadas en [41, 55]. Algunas extensiones y aplicaciones se pueden ver en [23, 50, 61, 28]. Recientes trabajos en este tema son [2, 25, 10, 4, 3].

Definición A.3.1. *Sea $w \geq 1$ una función de ponderación y P una probabilidad de transición sobre X tal que $\|P\|_w < \infty$ (ver 2.5). Se dice que P es $w-$*geométricamente ergódica *si existe una m.p. $\mu \in \mathbb{M}_w(X)$ y constantes no negativas M y $\rho < 1$, tal que*

$$\|P^t - \mu\|_w \leq M\rho^t \tag{A.2}$$

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

El límite μ en (A.2) es necesariamente único. Más aun, μ es una medida de probabilidad invariante pues

$$\begin{aligned}\|\mu P - \mu\|_w &\leq \|P^t - \mu\|_w + \|P^t - \mu P\|_w \\ &\leq \|P^t - \mu\|_w + \|(P^{t-1} - \mu)P\|_w \\ &\leq \|P^t - \mu\|_w + \|P^{t-1} - \mu\|_w \|P\|_w \\ &\leq M\rho^t(1 + \rho^{-1}),\end{aligned}$$

cuando $t \to \infty$, obtenemos $\mu P = \mu$.

Nota A.3.2. Considerando (2.5), la desigualdad (A.2) se puede expresar en forma explícita como

$$\left| \int_X u(y) Q^t(dy|x) - \int_X u(y)\mu(dy) \right| \leq w(x)\|u\|_w M\theta^t \tag{A.3}$$

para todo $u \in \mathbb{B}_w(X)$, $x \in X$, y $t = 0, 1, ...$

Finalmente, mencionamos el siguiente teorema que nos da condiciones suficientes para que una probabilidad de transición tenga la propiedad (A.2). Para la demostración ver [55, Sección 15] o bien [24, 3, 4].

Teorema A.3.3. *Sea $w \geq 1$ una función de ponderación y P una probabilidad de transición en X. Supongamos que existen: una medida $\sigma-$finita λ sobre X para el cual P es irreducible; una medida de probabilidad $\nu \in \mathbb{P}(X)$; un número real positivo $\gamma < 1$; y una función medible $\beta : X \to [0, 1]$ para los cuales*

1. *$P(D|x) \geq \beta(x)\nu(D)$,*
2. *$Pw(x) \leq \gamma w(x) + \beta(x)\|\nu\|_w$,*
3. *$\int_X w(y)\nu(dy) < \infty$,*

entonces P es $w-$geométricamente ergódica.

Bibliografía

[1] G. S. Amacher, M. Ollikainen y E. Koskela [2009], Economics of Forest Resources, MIT Press, Cambridge, Mass.

[2] P. Baxendale [2011], *T. E. Harris's contributions to recurrent Markov processes and stochastic flows*, Ann. Probab. 39(2), 417–428.

[3] W. Bednorz [2013], *The Kendall's theorem and its application to the geometric ergodicity of Markov chains*, arXiv:1301.1481.

[4] W. Bednorz, K. Latuszynski y R. Latala [2008], *A regeneration proof of the central limit theorem for uniformly ergodic Markov chains*, Elect. Comn. in Probab. 13, 85–98.

[5] R. Bellman [1957], *A Markovian decision process*, J. Math. Mech. 6, 679–684.

[6] D. P. Bertsekas y S. E. Shreve [1978], Stochastic Optimal Control: The discrete Time Case, Academic Press, New York.

[7] D. Blackwell y T. S Ferguson [1968], *The big match*, Ann. Math. Stat. 39, 159–163.

[8] J. Blot y N. Hayek [2014], Infinite-Horizon Optimal Control in the Discrete-Time Framework, SpringerBriefs in Optimization.

[9] W. A. Brock [1970], *On existence of weakly maximal programmes in a multi-cector economy*, Rev. Econ. Stud. 37, 275–280.

[10] B. Cloez y M. Hairer [2015], *Exponential ergodicity for Markov processes with random switching*, Bernoulli 21(1), 505–536.

[11] R. A. Dana y C. Le Van [2006], *Optimal growth without discounting*, En R. A. Dana, C. Le Van, T. Mitra y K. Nishimura, editores, Handbook of Optimal Growth: The Discrete Time Horizon,Volume I, Springer-Verlag, Berlin.

[12] E. V. Denardo [1970], *Computing a bias optimal policy in a discrete-time Markov decision problem*, Operations Research 18, 279–289.

[13] M. Diehl, R. Amrit y J. B. Rawlings [2011], *A Lyapunov function for economic optimizing model predictive control,* IEEE Trans. Autom. Control 56(3), pp. 703–707.

[14] R. Dorfman, P. Samuelson y R. Solow [1958]. Linear Programing and Economic Analysis. New York: McGraw-Hill.

[15] P. K. Dutta [1991], *What do discounted optima converge to? A theory of discount rate asymptotics in economic models*, J. Econom. Theory 55, 64–94.

[16] G. Fabbri, S. Faggian y G. Freni [2015], *On the Mitra-Wan forest management problem in continuous time*, J. Econ. Theory 157, 1001 – 1040.

[17] K. Fan [1953], *Minimax theorems*, Proc. Nat. Acad. Sci. USA 39, 42–47.

[18] J. Flynn [1980], *On optimality criteria for dynamic programs with long finite horizons*, J. Math. Anal. Appl. 76, 202–208.

[19] D. Gale [1967], *On optimal development in a multi-sector economy*, Rev. Econ. Stud. 34, 1–18.

[20] M.K. Ghosh y A. Bagchi [1998], *Stochastic games with average payoff criterion*, Appl. Math. Optim. 38, 283–301.

[21] I. I. Gihman y A. V. Skorohod [1979], Controlled Stochastic Processes, Springer-Verlag, New York.

[22] D. Gillette [1957], *Stochastic games with zero stop probabilities*, Ann. Math. Stud. 39, 179–187.

[23] E. Gordienko y O. Hernández-Lerma [1995], *Average cost Markov control policies with weighted norms: existence of canonical policies*, Appl. Math. (Warsaw) 23, 199–218.

[24] M. Hairer y J. C. Mattingly [2011], *Yet another look at Harris' ergodic theorem for Markovian chains*, en: Seminar on Stochastic Annalysis, Random Fields and Applications VI, vol. 63 of Progr. Probab., 109–117, Birkhäuser/Springer Basel AG, Basel.

[25] T. E. Harris [1956], *The existence of stationary measures for certain Markov processes*, en: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. II, 113–124, University of California Press, Berkeley y Los Angeles.

[26] A. Haurie [1976], *Optimal control on an infinite time horizon: the turnpike approach*, J. Math. Econ. 3(1), 81–102.

[27] M. Haviv y M. L. Puterman [1998], *Bias optimality in controlled queueing systems*, Journal of Applied Probability 35, 136–150.

[28] O. Hernández-Lerma y J. Lasserre [1998], Further Topics on Discrete-Time Markov Control Processes, Springer-Verlag, New York.

[29] O. Hernández-Lerma y J. Lasserre [2000], *Further criteria for positive Harris recurrence of Markov chains*, Proceedings of the American Mathematical Society, 129, 1521-1524.

[30] O. Hernández-Lerma y J. Lasserre [2001], *Zero-sum stochastic games in Borel spaces: average payoff criteria*, SIAM J. Control Optim., 39(5), 1520-1539.

[31] O. Hernández-Lerma, R. Montes-de-Oca y R. Cavazos-Cadena [1991], *Recurrence conditions for MDPs with Borel state space: A survey*, Ann. Oper. Res., 28, 29-46.

[32] O. Hernández-Lerma y O. Vega-Amaya [1998], *Infinite-horizon Markov control processes with undiscounted cost criteria: From average to overtaking optimality*, Appl. Math. (Warsaw), 25, 153–178.

[33] N. Hilgert y O. Hernández-Lerma [2003], *Bias optimality versus strong 0-discount optimality in Markov control processes with unbounded costs*, Acta Applicadae Mathematicae, 77, 215-235.

[34] H. Hotelling [1931], *The economics of exhaustive resources*, J. Polit. Econ. 39(2), 137–175.

[35] A. Jaśkiewicz [2009], *Zero-sum ergodic semi-Markov games with weakly continuous transition probabilities*, J. Optim. Theory Appl., 141, 321–347.

[36] A. Jaśkiewicz y A. S. Nowak [2006], *Zero-sum ergodic stochastic games with feller transition probabilities*, SIAM J. Control Optim. 45(3), 773–789.

[37] A. Jaśkiewicz y A. S. Nowak [2011], *Stochastic games with unbounded payoffs: applications to robust control in economics*, Dyn. Games Appl., 1, 253–279.

[38] A. Jaśkiewicz y A. S. Nowak [2014], *Robust Markov control processes*, J. Math. Anal. Appl. , 420, 1337–1353.

[39] A. Jaśkiewicz y A. S. Nowak [2015], *On pure stationary almost Markov Nash equilibria in nonzero-sum ARAT stochastic games*, Math. Meth. Oper. Res. 81, 169-179.

[40] S. Kant y R. Albert Berry (Eds.) [2005], Economics, Sustainability, and Natural Resources: Economics of Sustainable Forest Management. Springer, Dordrecht.

[41] N.V. Kartashov [1985], *Inequalities in theorems of ergodicity and stability for Markov chains with a common phase space*, Theory Probab. Appl. 30, 247–259.

[42] M. Ali Khan y A. Piazza [2010], *On uniform convergence of undiscounted optimal programs in the Mitra-Wan forestry model: The strictly concave case*, Int. J. Econ. Theory 6, 57–76.

[43] M. Ali Khan y A. Piazza [2011], *An overview of turnpike theory: Towards the discounted deterministic case*, Adv. Math. Econ. 14, 39–67.

[44] M. Ali Khan y A. Piazza [2011], *Classical turnpike theory and the economics of forestry*, J. Econ. Behav. Organ 79, 194–210.

[45] M. Ali Khan y A. Piazza [2012], *On the Mitra-Wan forestry model: A unified analysis*, J. Econ. Theory 147, 230–260.

[46] V. Kolokoltsov y W. Yang [2012]. *The turnpike theorems for Markov games*, Dyn. Games and Appl. 2, 294–312.

[47] L. R. Laura-Guarachi y O. Hernández-Lerma [2015], *The Mitra-Wan forestry model: a discrete-time optimal control problem*, Natural Resource Modeling 28(2), 152–168.

[48] M. E. Lewis y M. L. Puterman [2002]. *Bias optimality*, en: E. A. Feinberg y A. Shwartz (Eds.), Handbook of Markov Decision Processes, Internat. Ser. Oper. Res. Management Sci. 40, Kluwer, Boston, 89-111.

[49] N. V. Long [2010], A survey of dynamic games in economics, World Scientific Publishing Co. Pte. Ltd., vol. 1.

[50] F. Luque-Vásquez y O. Hernández-Lerma [1999], *Semi-Markov control models with average costs*, Appl. Math.(Warsaw) 26, 315–331.

[51] A. Maitra y T. Parthasarathy [1970], *On stochastic games*, J. Optim. Theory Appl. 5, 289–300.

[52] M. A. Mamedov [2003], *A turnpike theorem for continuous-time control systems when the optimal stationary point is not unique*, Abstract and Applied Analysis 11, 631–650.

[53] L. W. McKenzie [1986], *Accumulation programs of maximum utility and the von Neuwmann facet*, en: J. N. Wolfe (Ed.), Value, Capital, and Growth, Edinburgh University Press, 353-383.

[54] S. Meyn y R. Tweedie [1992], *Stability of Markovian processes. I. Criteria for discrete–time chains*, Adv. in Appl. Probab. 24(3), 542-574.

[55] S. Meyn y R. Tweedie [1993], Markov Chains and Stochastic Stability. Comunications and Control Engineering Series, Springer-Verlag London Ltd., London.

[56] T. Mitra y H. Y. Wan [1985], *Some theoretical results on the economics of forestry*, Rev. Econ. Stud. 52(2), 263–282.

[57] T. Mitra y H. Y. Wan [1986], *On the Faustmann solution to the forest management problem*, J. Econ. Theory 40, 229–249.

[58] T. Mitra [2005]. *Intergenerational equity and the forest management problem*, en: S. Kant y R. A. Berry (Eds.), Economics, Sustainability, and Natural Resources: Economics of Sustainable Forest Management, Springer, Berlin, 317–173.

[59] D. Monderer y S. Sorin [1993], *Asymptotic properties in dynamic programming*, Internat. J. Game Theory 22, 1–11.

[60] A. S. Nowak [1985], *Measurable selection theorems for minimax stochastic optimization problems*, SIAM J. Control Optim. 42, 466–476.

[61] A. S. Nowak [1999], *Optimal strategies in a class of zero-sum ergodic stochastic games*, Math. Methodos Oper. Res. 50, 399–419.

[62] A. S. Nowak [2007], *On stochastic games in economics*, Math. Methodos Oper. Res. 66, 513–530.

[63] A. S. Nowak [2008], *Equilibrium in a dynamic game of capital accumulation with the overtaking criterion*, Economic Letters 99, 233–237.

[64] A. S. Nowak y O. Vega-Amaya [1999], *A counterexample on overtaking optimality*, Mathematical Methods of Operations Research 49, 435–439.

[65] E. Phelps [1967], Golden Rules of Economic Growth, North-Holland Publishing Company.

[66] A. B. Piunovskiy [2013]. Examples in Markov Decision Processes, Imperial College Press, London.

[67] T. Prieto-Rumeau y O. Hernández-Lerma [2005], *Bias and overtaking equilibria for zero-sum continuous-time Markov games*, Math. Meth. Oper Res 61, 437 – 454.

[68] F.P. Ramsey [1928], *A mathematical theory of savings*, Econom. J. 38, 543–559.

[69] P.A. Samuelson [1976], *Economics of forestry in an evolving society*, Economic Inquiry 14, 466–492.

[70] L.S. Shapley [1953], *Stochastic games*, Proc. Natl. Acad. Sci. USA 39, 1095–1100.

[71] M. Sion [1958], *On general minimax theorems*, Pacific J. Math. 8, 171–176.

[72] F. Thuijsman [2003], *The Big Match an the Paris Match*, en: A. Neyman y S. Sorin (Eds.), Stochastic Games and Applicaions, NATO Science Series C, Mathematical and Physical Sciences, Vol. 570, Kluwer Academic Publishers, Dordrecht, 195 – 204.

[73] O. Vega-Amaya [2003], *The average cost optimality equation: A fixed point approach*, Bol. Soc. Mat. Mexicana 9, 185–195.

[74] A. F. Veinott [1966], *On finding optimal policies in discrete dynamic programming with no discounting*, Ann. Math. Statist. 37, 1284–1294.

[75] A. F. Veinott [1969], *Discrete dynamic programming with sensitive discount optimality criteria*, Ann. Math. Statist. 40, 1635–1660.

[76] N. Vieille [2002], *Stochastic games: recent results,* en: R. J. Aumann, S. Hart (Eds.), Handbook of Game Theory with Economic Applications, North-Holland, Amsterdam, 1833–1850.

[77] G. Vigeral [2013], *A zero-sum stochastic game with compact action sets and no asymptotic value*, Dyn. Games Appl. 3, 172–186.

[78] Q. Wei y X. Chen [2015], *Constrained stochastic games with the average payoff criteria* , Operations Research Letters 43, 83–88.

[79] C.C. von Weizsäcker [1965], *Existence of optimal programs of accumulation for an infinite horizon*, Rev. Econ. Stud. 32, 85–104.

[80] A. J. Zaslavski [1995], *Optimal programs on infinite horizon 1*, SIAM J. Control Optim. 33, 1643–1660.

[81] A. J. Zaslavski [1995], *Optimal programs on infinite horizon 2*, SIAM J. Control Optim. 33, 1661–1686.