UNIVERSIDAD DE SONORA

División de Ciencias Exactas y Naturales

Programa de Posgrado en Matemáticas

Zero-sum Markov games on Borel spaces with non-constant discount factors and unbounded payoff

T E S I S

Que para obtener el grado académico de:

Maestro en Ciencias (Matemáticas)

Presenta:

L.M. Estephania Pivac Alcaraz

Director de tesis: Dr. Jesús Adolfo Minjárez Sosa

Hermosillo, Sonora, México Diciembre de 2017

UNIVERSIDAD DE SONORA

"El Saber de Mis Hijos
Hará Mi Grandeza"

El saber de mis hijos
hará mi grandeza

Estvacl

ESTEPHANIA PIVAC
ALCARAZ

211221128

Acta: 47

Foja: 47

Libro: 1

DR. JESÚS ADOLFO MINJÁREZ SOSA, Coordinador del Programa de Maestro en Ciencias Matemáticas de la Universidad de Sonora, hace constar que las firmas que anteceden corresponden al jurado que intervino en el examen de grado.

Hermosillo, Sonora, a 15 de diciembre de 2017

Adolfo Minjárez

DR. JESÚS ADOLFO MINJÁREZ SOSA

Coordinador de programa

ACTA DE EXAMEN DE GRADO

En la ciudad de Hermosillo, Sonora, siendo las 11:00 horas del día 15 de diciembre de 2017, se reunieron en la Sala de Video Conferencia del Departamento de Matemáticas de la Universidad de Sonora, los integrantes del jurado:

Dr. David González Sánchez
Dr. Fernando Luque Vásquez
Dr. Héctor Jasso Fuentes
Dr. Jesús Adolfo Minjárez Sosa

bajo la presidencia del primero y fungiendo como secretario el último, para realizar el examen de grado del programa de Maestro en Ciencias Matemáticas, a:

ESTEPHANIA PIVAC ALCARAZ

quien de acuerdo a la opción de examen de grado presentó un trabajo de investigación titulado

"Zero-sum Markov games on Borel spaces with non-constant discount factors and unbounded payoff"

El jurado, después de debatir entre sí reservada y libremente, emitió el siguiente dictamen:

APROBADA POR UNANIMIDAD

y para constancia se levantó la presente acta.

Dr. David González Sánchez Presidente	Adolfo Minjárez **Dr. Jesús Adolfo Minjárez Sosa** Secretario
Dr. Fernando Luque Vásquez Sinodal	**Dr. Héctor Jasso Fuentes** Sinodal externo

DIVISIÓN DE CIENCIAS EXACTAS Y NATURALES, COORDINACION POSGRADO EN MATEMATICAS

Folio: 7557

"Lo que deseo, es que todo sea redondo y no haya de ningún modo ni principio ni fin en la forma, sino que haga un conjunto armonioso de vida."
—Vincent van Gogh

Con inmenso amor,
a mi mamá.

Agradecimientos

Quiero comenzar agradeciendo primeramente a Dios, por permitirme cumplir uno más de mis sueños. Por estar siempre cuidando de mi camino y por ser la luz que guía mi vida.

Gracias a mi familia, que son mi motor principal. Por ustedes soy, y por ustedes seré. Los amo inmensamente, ustedes son lo más preciado que tengo en esta vida y gracias a ustedes he llegado hasta donde estoy. Gracias a mi mamá, que es el más grande ángel que Dios puso en mi camino, sin lugar a dudas por ti soy lo que soy. Gracias a mis hermanas Carolina, Vesna Camile y Leonela, por llenar mis días de alegría y risas. Gracias a mi papá por todo su esfuerzo para ayudarme a lograr este objetivo. Gracias a mi abuelita por todo su amor y a mi primo Giovanni por estar siempre para mí.

También quiero agradecer a Pedro Morghen por ser y por estar. Gracias por todo el tiempo juntos, por cuidar siempre de mí, por escucharme y por todo tu amor.

Gracias a todos mis compañeros y amigos que estuvieron presentes en esta etapa. En particular quiero agradecer a Carolina Martínez por su bonita amistad, y por todos los momentos que hemos pasado juntas.

Quiero agradecer de manera muy especial a mi director de tesis, el Dr. Adolfo Minjárez Sosa, por su infinita paciencia, por el tiempo dedicado a este trabajo, por ayudarme en todo momento, por su amistad, por sus consejos, enseñanzas y por su gran apoyo. Gracias por permitirme crecer como matematica e investigadora.

Gracias a mis sinodales por el tiempo dedicado a la revisión y corrección de este trabajo de tesis, por todos sus consejos y sugerencias.

También quiero agradecer a todos mis profesores por sus enseñanzas, por guiarme en mi crecimiento tanto profesional como personal. Toda mi admiración y respeto a ustedes.

Quiero agradecer a la Universidad de Sonora y al Consejo Nacional de Ciencia y Tecnología (CONACyT) por el apoyo económico brindado y por todas las oportunidades que me han dado, que me permitieron concluir mis estudios de posgrado, y sin las cuales no hubiera sido posible.

Finalmente quiero agradecer a todas aquellas personas que he olvidado mencionar, y que de una u otra manera han dejado una huella en mi vida.

Estephania Pivac Alcaraz

Hermosillo, Sonora, México a 15 de diciembre de 2017.

Contents

Introduction

This work deals with three different classes of discrete-time zero-sum discounted Markov games with non-constant discount factors, namely, games

1. where the state process $\{x_n\}$ and the discount process $\{\alpha_n\}$ evolve according to a coupled transition law Q.
2. with state-actions dependent discount factors of the form
$$\alpha(x_n, a_n, b_n),$$
where a_n and b_n represent the actions of players 1 and 2, at time n, respectively;
3. with random state-actions dependent discount factors of the form
$$\alpha(x_n, a_n, b_n, \xi_{n+1}),$$
where $\{\xi_n\}$ is the discount factors' disturbance process, which is a sequence of independent and identically distributed random variables.

In general, a two-person zero-sum game is a two-player game where the profit of one player represents the cost of the other. Then, whereas the goal of player 1 is to maximize her payoff, the goal of player 2 is to minimize his cost.

The main objective is to prove the existence of a value of the game, as well as optimal strategies for both players, for each class of the above games.

The natural motivation in considering non-constant discount factors comes from the applications in economic and financial models where, in general, the discount factors are functions of the interest rates which in turn are uncertain. Such uncertainty can be caused by different facts, e.g., the amount of currency in circulation, and/or actions of the players, and furthermore, random noises. In these cases we have non-constant discount factors, for which the usual theory on discounted Markov games with constant discount factors is not applicable.

This work is structured in three chapters, each corresponding to each type of game with non-constant discount factors we are dealing with.

In the first one, we consider that the payoffs are exponentially discounted with cumulative random discount rates. That is, a payoff R in stage n is equivalent to a payoff $R\exp(-S_n)$ at time 0, where $S_n = \sum_{k=0}^{n-1} \alpha_k$ if $n \geq 1$, $S_0 = 0$. Hence, the payoff at stage n takes the form
$$e^{-S_n} r(x_n, \alpha_n, a_n, b_n),$$

where r represents the one-stage payoff function for players, i.e., r is a reward for player 1 and a cost for player 2. This kind of games is studied by analyzing the joint process $\{(x_t, \alpha_t)\}$, formed by the state and discount processes, which evolves according to a joint transition law Q, and whose performance index takes the form

$$E\left[\sum_{t=0}^{n} e^{-S_t} r(x_t, \alpha_t, a_t, b_t)\right].$$

We analyze the case $n \in \mathbb{N}$ and $n = \infty$.

In Chapter 2 we study discrete-time zero-sum discounted Markov games with state-actions dependent discount factors. In this case we assume that α is a measurable function of the form

$$\alpha(x_n, a_n, b_n), \tag{1}$$

which plays the following role in the evolution of the game. At time $t = 0$ when the game is in state x_0, players 1 and 2 select actions a_0 and b_0, respectively. Then player 1 receives from player 2 a payoff $r(x_0, a_0, b_0)$ and the game moves to a new state x_1 according to a transition law. Once the game is in state x_1 players select actions a_1 and b_1 and player 1 receives from player 2 a discounted payoff $\alpha(x_0, a_0, b_0)r(x_1, a_1, b_1)$. Next the game moves to a new state x_2 and the process is repeated. In general, the payoffs are discounted with multiplicative discount rates, that is, at stage $n \in \mathbb{N}$, player 1 receives from player 2 a discounted payoff of the form

$$\Gamma_n r(x_n, a_n, b_n), \tag{2}$$

where,

$$\Gamma_n := \prod_{k=0}^{n-1} \alpha(x_k, a_k, b_k) \;\text{ if }\; n \in \mathbb{N}, \quad \text{and} \quad \Gamma_0 = 1.$$

Thus the goal of player 1 is to maximize the total expected discounted payoff defined by the accumulation of the one-stage payoffs (2) over an infinite horizon, whereas the goal of player 2 is to minimize such payoff of the form

$$E\left[\sum_{n=0}^{\infty} \Gamma_n r(x_n, a_n, b_n)\right]. \tag{3}$$

To illustrate the application of this class of games, we present an example of a class of games with random horizon.

Finally, in Chapter 3, we study discrete-time zero-sum discounted Markov games with random state-actions dependent discount factors of the form

$$\tilde{\alpha}(x_n, a_n, b_n, \xi_{n+1}),$$

where $\{\xi_n\}$ is a sequence of independent and identically distributed random variables, with common distribution θ. The interpretation of the discount factor function $\tilde{\alpha}$ is similar to that of (1) in the previous class of games. The difference is that we are now randomizing the discount factors, which constitutes a class of games more general than the previous ones. The performance index takes the form

$$E\left[\sum_{n=0}^{\infty} \prod_{k=0}^{n-1} \tilde{\alpha}(x_k, a_k, b_k, \xi_{k+1}) r(x_n, a_n, b_n)\right]. \tag{4}$$

Our approach is to prove that (4) is equivalent to a performance index of the form (3), with state-actions-dependent discount factors functions

$$\alpha_\theta(x,a,b) = \int \tilde{\alpha}(x,a,b,s)\theta(ds).$$

Then, we apply the results corresponding to (3). This kind of games are illustrated with a class of semi-Markov games.

The work is based mainly on the papers [4–6, 11, 13]. For instance, MDPs with random discount factors are analyzed in [4,5] as we do in Chapter 1. Hence, our results extend such papers to the Markov games. On the other hand, the results in Chapters 2 and 3 are taken from works [6, 11, 13].

Chapter 1

Markov games with random discount factors

1.1 Introduction

In this chapter we study a class of discrete-time zero-sum Markov games under an optimality criterion with random discount factors. We begin introducing the game model we are concerned with, as well as the necessary elements to define the corresponding game problem. We also present certain assumptions under which we prove the existence of an optimal pair of strategies. Finally, we present an example to illustrate one potential application of the developed theory.

In order to introduce this class of games and the randomized discount factor, we will assume that the payoffs are exponentially discounted on time, that is, a payoff R at stage t is equivalent to a payoff Re^{-S_t} at time 0, where $S_t = \sum_{i=0}^{t-1} \alpha_i$ if $t \geq 1$, $S_0 = 0$, and $\alpha_t > 0$ represents the discount factor imposed at time t.

1.2 Game model

Consider the following zero-sum two-person game model with random discount factor

$$\mathcal{GM} := (\mathbf{X}, \mathbf{\Gamma}, \mathbf{A}, \mathbf{B}, \mathbb{K}_{\mathbf{A}}, \mathbb{K}_{\mathbf{B}}, Q, r) \tag{1.1}$$

where:

- The state space $\mathbf{X}$ is a non-empty Borel space.
- The discount factor set is $\mathbf{\Gamma} := [\alpha^*, \infty)$, $\alpha^* > 0$.
- The action sets $\mathbf{A}$ and $\mathbf{B}$ for players 1 and 2, respectively, are both non-empty Borel spaces.
- The constraint sets $\mathbb{K}_{\mathbf{A}}$ and $\mathbb{K}_{\mathbf{B}}$ are non-empty Borel subsets of $\mathbf{X} \times \mathbf{\Gamma} \times \mathbf{A}$ and $\mathbf{X} \times \mathbf{\Gamma} \times \mathbf{B}$, respectively. For each $(x, \alpha) \in \mathbf{X} \times \mathbf{\Gamma}$,

  $$A(x, \alpha) := \{a \in \mathbf{A} \; : \; (x, \alpha, a) \in \mathbb{K}_{\mathbf{A}}\}$$

  and

  $$B(x, \alpha) := \{b \in \mathbf{B} \; : \; (x, \alpha, b) \in \mathbb{K}_{\mathbf{B}}\}$$

represent the admissible action (or control) sets for player 1 and 2, respectively, when the system is at state x, and the discount factor α is imposed. The set

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

of admissible state-actions (or controls) quadruplets is a Borel subset of $\mathbf{X} \times \mathbf{\Gamma} \times \mathbf{A} \times \mathbf{B}$.

- The transition law Q is a stochastic kernel (s.k.) on $\mathbf{X} \times \mathbf{\Gamma}$ given $\mathbb{K}$, which denotes the joint distribution law of the state-discount process.
- The one-stage payoff function $r : \mathbb{K} \to \mathbb{R}$ is a measurable function on $\mathbb{K}$.

Interpretation. The game model $\mathcal{GM}$ in (**??**) represents a game that evolves as follows. At each stage $t = 0, 1, \ldots$, players 1 and 2 observe the current game state $x_t = x \in \mathbf{X}$, take into account the imposed discount factor $\alpha_t = \alpha$, and independently choose actions $a_t = a \in A(x, \alpha)$ and $b_t = b \in B(x, \alpha)$, respectively. Then two things happen: (i) player 1 immediately receives a payoff $r(x, \alpha, a, b)$ from player 2, and (ii) the game moves to a new state x_{t+1} and a new discount factor α_{t+1} is imposed according to the transition law $Q(\cdot|x, \alpha, a, b)$. Once the transition to the new state and the new discount factor has occurred, the players choose new actions, and the process is repeated over and over again.

1.2.1 Strategies

The actions chosen by players at each stage are selected by rules known as strategies which are defined as follows.

We define the space of admissible histories up to time t by $\mathbb{H}_0 := \mathbf{X} \times \mathbf{\Gamma}$ and $\mathbb{H}_t := \mathbb{K}^t \times \mathbf{X} \times \mathbf{\Gamma}$, $t \geq 1$. A generic element h_t of $\mathbb{H}_t$ is denoted by

$$h_t := (x_0, \alpha_0, a_0, b_0, \ldots, x_{t-1}, \alpha_{t-1}, a_{t-1}, b_{t-1}, x_t, \alpha_t),$$

where $(x_i, \alpha_i, a_i, b_i) \in \mathbb{K}$ for $i = 0, 1, \ldots, t-1$, and $(x_t, \alpha_t) \in \mathbf{X} \times \mathbf{\Gamma}$, which represents the history of the game up to time t.

For each $(x, \alpha) \in \mathbf{X} \times \mathbf{\Gamma}$, let $\mathbb{A}(x, \alpha) := \mathbb{P}(A(x, \alpha))$ and $\mathbb{B}(x, \alpha) := \mathbb{P}(B(x, \alpha))$ (see Appendix B for details). We define the sets of stochastic kernels

$$\Phi^1 := \{\varphi^1 \in \mathbb{P}(\mathbf{A}|\mathbf{X} \times \mathbf{\Gamma}) : \varphi^1(\cdot|x, \alpha) \in \mathbb{A}(x, \alpha) \ \ \forall (x, \alpha) \in \mathbf{X} \times \mathbf{\Gamma}\}$$
$$\Phi^2 := \{\varphi^2 \in \mathbb{P}(\mathbf{B}|\mathbf{X} \times \mathbf{\Gamma}) : \varphi^2(\cdot|x, \alpha) \in \mathbb{B}(x, \alpha) \ \ \forall (x, \alpha) \in \mathbf{X} \times \mathbf{\Gamma}\}.$$

Definition 1.2.1. *A strategy for player 1 is a sequence* $\pi^1 = \{\pi^1_t\}$ *of stochastic kernels* $\pi^1_t \in \mathbb{P}(\mathbf{A}|\mathbb{H}_t)$ *such that:*

$$\pi^1_t(A(x_t, \alpha_t)|h_t) = 1, \quad \forall \ h_t \in \mathbb{H}_t, \quad t \in \mathbb{N}_0. \tag{1.2}$$

We denote by Π^1 *the family of all strategies for player 1.*

Definition 1.2.2. *A strategy* $\pi^1 = \{\pi^1_t\} \in \Pi^1$ *for player 1 is called:*

(a) a Markov strategy if $\pi^1_t \in \Phi^1$ *for all* $t \in \mathbb{N}_0$*, that is, each* π^1_t *depends only on the current state and the discount factor* (x_t, α_t) *of the system. The set of all Markov strategies for player 1 is denoted by* Π^1_M*.*

(b) a stationary (Markov) strategy if $\pi_t^1(\cdot|h_t) = \varphi^1(\cdot|x_t, \alpha_t)$ $\forall$ $h_t \in \mathbb{H}_t$, $t \in \mathbb{N}_0$, for some stochastic kernel φ^1 in Φ^1, so that $\pi^1 = \{\varphi^1, \varphi^1, \ldots\} := \{\varphi^1\}$. The set of all stationary strategies for player 1 is denoted by Π_S^1.

We have, of course, the following relations

$$\Pi_S^1 \subset \Pi_M^1 \subset \Pi^1.$$

The sets of all strategies Π^2, Markov strategies Π_M^2 and stationary strategies Π_S^2 corresponding to player 2 are defined similarly.

1.2.2 The game process

Let $(\Omega, \mathcal{F})$ be the measurable space that consists of the sample space $\Omega := \mathbb{H}_\infty = \times_{t=0}^{\infty}(\mathbf{X} \times \mathbf{\Gamma} \times \mathbf{A} \times \mathbf{B})$ and its corresponding product σ-algebra $\mathcal{F}$. Elements of Ω are sequences of the form $\omega = (x_0, \alpha_0, a_0, b_0, x_1, \alpha_1, a_1, b_1, \ldots)$, $(x_t, \alpha_t) \in \mathbf{X} \times \mathbf{\Gamma}$ and $a_t \in \mathbf{A}$, $b_t \in \mathbf{B}$, for all $t \in \mathbb{N}_0$. Notice that Ω contains the space of all admissible histories $(x_0, \alpha_0, a_0, b_0, x_1, \alpha_1, a_1, b_1, \ldots)$ with $(x_t, \alpha_t, a_t, b_t) \in \mathbb{K}$ for all $t \in \mathbb{N}_0$.

For each pair of strategies $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$ and each initial probability measure ν in $\mathbf{X} \times \mathbf{\Gamma}$, there exists a unique probability measure $P_\nu^{\pi^1,\pi^2}$ in $(\Omega, \mathcal{F})$ which satisfies $P_\nu^{\pi^1,\pi^2}(\mathbb{H}_\infty) = 1$, and for $t \in \mathbb{N}_0$, $C \in \mathcal{B}(\mathbf{X} \times \mathbf{\Gamma})$, $A \in \mathcal{B}(\mathbf{A})$, $B \in \mathcal{B}(\mathbf{B})$,

$$\begin{aligned} P_\nu^{\pi^1,\pi^2}[(x_0, \alpha_0) \in C] &= \nu(C); && (1.3) \\ P_\nu^{\pi^1,\pi^2}[a_t \in A, b_t \in B | h_t] &= \pi_t^1(A|h_t)\pi_t^2(B|h_t); && (1.4) \\ P_\nu^{\pi^1,\pi^2}[(x_{t+1}, \alpha_{t+1}) \in C | h_t, a_t, b_t] &= Q(C|x_t, \alpha_t, a_t, b_t). && (1.5) \end{aligned}$$

We denote by $E_\nu^{\pi^1,\pi^2}$ the expectation operator with respect to $P_\nu^{\pi^1,\pi^2}$.

If ν is concentrated in $(x, \alpha) \in \mathbf{X} \times \mathbf{\Gamma}$, then we write $P_{(x,\alpha)}^{\pi^1,\pi^2}$ and $E_{(x,\alpha)}^{\pi^1,\pi^2}$ instead of $P_\nu^{\pi^1,\pi^2}$ and $E_\nu^{\pi^1,\pi^2}$, respectively.

The stochastic process $\{x_n\}$ defined on $(\Omega, \mathcal{F}, P_{(x,\alpha)}^{\pi^1,\pi^2})$ is called game process.

1.3 Difference equation models

A particular case of a game model (1.1) is constituted when the dynamic of the game is determined by a stochastic difference equations. For instance, suppose that the discount process evolves in time according to the difference equation:

$$\alpha_{t+1} = G(\alpha_t, \eta_t),$$

for $t = 0, 1, \ldots$, where G is a known continuous function, $\{\eta_t\}$ is the discount random disturbance process, which is formed by independent and identically distributed (or i.i.d. for short) random variables, with common density ρ, in $\mathbb{R}^k$, respectively.

Thus we can represent the game model by

$$\mathcal{GM}_{DE} := (\mathbf{X}, \mathbf{\Gamma}, \mathbf{A}, \mathbf{B}, \mathbb{K}_A, \mathbb{K}_B, Q_1, \mathbb{R}^k, \rho, G, r)$$

where the stochastic kernel Q_1 on $\mathbf{X}$ given $\mathbb{K}$, represents the state process transition law.

In this case, G defines the s.k.:

$$Q_2(\Gamma|\alpha) := \int_{\mathbb{R}^K} 1_\Gamma[G(\alpha, s)]\rho(ds), \quad \Gamma \in \mathcal{B}(\mathbf{\Gamma}),$$

which represents the discount process transition law.

Then, the joint s.k. Q, that represents the state-discount process transition law, is defined as follows:

$$Q(C|x_t, \alpha_t, a_t, b_t) := Q_1 \times Q_2(C|x_t, \alpha_t, a_t, b_t) \quad \forall C \in \mathcal{B}(\mathbf{X} \times \mathbf{\Gamma})$$

where,

$$\begin{aligned} Q(X \times L|x_t, \alpha_t, a_t, b_t) &= Q_1(X|x_t, \alpha_t, a_t, b_t) \cdot Q_2(L|\alpha_t) \\ &= \int_X \int_{\mathbb{R}^k} 1_L[G(\alpha, s)]\rho(ds) Q_1(dx|x_t, \alpha_t, a_t, b_t) \end{aligned}$$

$X \in \mathcal{B}(\mathbf{X})$, $L \in \mathcal{B}(\mathbf{\Gamma})$.

Hence, we can write the model of this game $\mathcal{GM}_{DE}$ in terms of Q, and we obtain a game model as $\mathcal{GM}$ in (1.1).

1.4 Optimality criterion

As we have mentioned before, we assume that the payoffs are exponentially discounted with cumulative random discount rates. That is, a payoff R attained at stage t is equivalent to a payoff Re^{-S_t} at time 0, where $S_t := \sum_{i=0}^{t-1} \alpha_i$ if $t \geq 1$, and $S_0 = 0$. In this sense, when the players 1 and 2 use the strategies $\pi^1 \in \Pi^1$ and $\pi^2 \in \Pi^2$, respectively, given the initial state $x_0 = x$ and the initial discount factor $\alpha_0 = \alpha$, we define for each $n \in \mathbb{N}$

- the *expected discounted payoff up to the n-th stage* (with random discount factor) by

$$V_n(x, \alpha, \pi^1, \pi^2) := E_{(x,\alpha)}^{\pi^1, \pi^2} \left[\sum_{t=0}^{n-1} e^{-S_t} r(x_t, \alpha_t, a_t, b_t) \right]; \tag{1.6}$$

- the *total expected discounted payoff (with random discount factors)* as

$$V(x, \alpha, \pi^1, \pi^2) := E_{(x,\alpha)}^{\pi^1, \pi^2} \left[\sum_{t=0}^{\infty} e^{-S_t} r(x_t, \alpha_t, a_t, b_t) \right]. \tag{1.7}$$

Observe that $\{e^{-S_t}\}$ is a sequence of random variables (not necessarily independent) that represent the discount factor at stage t. Moreover, if $\alpha_t = \alpha$ for every $t \geq 0$ and some $\alpha \in (0, \infty)$, the performance index is reduced to the usual β-discounted criterion with $\beta = e^{-\alpha}$.

1.4.1 Game value

Definition 1.4.1. *For each $n \in \mathbb{N}$, the* lower and upper value of the game in n stages *are given as:*

$$L_n(x,\alpha) := \sup_{\pi^1 \in \Pi^1} \inf_{\pi^2 \in \Pi^2} V_n(x,\alpha,\pi^1,\pi^2) \tag{1.8}$$

and

$$U_n(x,\alpha) := \inf_{\pi^2 \in \Pi^2} \sup_{\pi^1 \in \Pi^1} V_n(x,\alpha,\pi^1,\pi^2), \tag{1.9}$$

respectively, for each initial state-discount pair $(x,\alpha) \in \mathbf{X} \times \mathbf{\Gamma}$.

Notice that, in general, $U_n(\cdot,\cdot) \geq L_n(\cdot,\cdot)$; nevertheless, if $U_n(x,\alpha) = L_n(x,\alpha)$ holds for every $(x,\alpha) \in \mathbf{X} \times \mathbf{\Gamma}$, then the common function is called *the value of the game in n stages* and it is denoted by V_n^*.

Definition 1.4.2. *Consider a game in n stages. If the discounted game has a value V_n^*, then:*

i) A strategy $\pi_*^1 \in \Pi^1$ is said to be optimal for player 1 *if*

$$V_n^*(x,\alpha) = \inf_{\pi^2 \in \Pi^2} V_n(x,\alpha,\pi_*^1,\pi^2).$$

ii) A strategy $\pi_*^2 \in \Pi^2$ is said to be optimal for player 2 *if*

$$V_n^*(x,\alpha) = \sup_{\pi^1 \in \Pi^1} V_n(x,\alpha,\pi^1,\pi_*^2).$$

Thus, the pair (π_*^1,π_*^2) is said to be an optimal pair of strategies.

The lower value $L(x,\alpha)$, the upper value $U(x,\alpha)$, the value of the game $V^*(\cdot,\cdot)$, and the optimal strategies of *the discounted game* (with infinite horizon) are defined similarly.

A pair of strategies $(\pi_*^1,\pi_*^2) \in \Pi^1 \times \Pi^2$ is called a *saddle point* if

$$V_n(x,\alpha,\pi^1,\pi_*^2) \leq V_n(x,\alpha,\pi_*^1,\pi_*^2) \leq V_n(x,\alpha,\pi_*^1,\pi^2), \tag{1.10}$$

$\forall(\pi^1,\pi^2) \in \Pi^1 \times \Pi^2,\ (x,\alpha) \in \mathbf{X} \times \mathbf{\Gamma}$. And observe that $(\pi_*^1,\pi_*^2) \in \Pi^1 \times \Pi^2$ is an optimal pair if and only if it is a saddle point.

1.5 Assumptions and preliminary results

In order to present our assumption and results in a more accessible form, we use the following notation.

Let $\mathbf{Y} := \mathbf{X} \times \mathbf{\Gamma}$, and $y_t := (x_t,\alpha_t)$. Notice that using this notation $\mathbb{K} = \{(y,a,b) :\ y \in \mathbf{Y}, a \in A(y),\ b \in B(y)\}$, and $Q \in \mathbb{P}(Y|\mathbb{K})$ represents the transition law of the process $\{y_t\}$.

For notational convenience, we write the game model (1.1) in the form

$$\mathcal{GM} = (\mathbf{Y},\mathbf{A},\mathbf{B},\mathbb{K}_A,\mathbb{K}_B,Q,r). \tag{1.11}$$

For each $y \in \mathbf{Y}$, $\mathbb{A}(y) := \mathbb{P}(A(y))$ and $\mathbb{B}(y) := \mathbb{P}(B(y))$. Observe that the multifunctions $y \to \mathbb{A}(y)$ and $y \to \mathbb{B}(y)$ are measurable, with values in compact sets if $A(y)$ and $B(y)$ are compact.

For probability measures $\varphi^1(\cdot|y) \in \mathbb{A}(y)$ and $\varphi^2(\cdot|y) \in \mathbb{B}(y)$, $y \in \mathbf{Y}$, we write $\varphi^i(y) := \varphi^i(\cdot|y)$, $i = 1, 2$. In addition, for a measurable function $u : \mathbb{K} \to \mathbb{R}$

$$u(y, \varphi^1, \varphi^2) = u(y, \varphi^1(y), \varphi^2(y)) := \int_{A(y)} \int_{B(y)} u(y, a, b)\varphi^1(da|y)\varphi^2(db|y). \tag{1.12}$$

For instance, for $y \in \mathbf{Y}$ and $Y \in \mathcal{B}(\mathbf{Y})$ we have

$$r(y, \varphi^1, \varphi^2) := \int_{A(y)} \int_{B(y)} r(y, a, b)\varphi^1(da|y)\varphi^2(db|y),$$

and

$$Q(Y|y, \varphi^1, \varphi^2) := \int_{A(y)} \int_{B(y)} Q(Y|y, a, b)\varphi^1(da|y)\varphi^2(db|y).$$

The existence of a value of the game as well as a pair of optimal strategies is analyzed under the following conditions.

Assumption 1.5.1. *The game model $\mathcal{GM}$ (1.11) satisfies the following:*

(a) For each state $y \in \mathbf{Y}$, the admissible actions sets $A(y)$ and $B(y)$ are compact.

(b) For each $(y, a, b) \in \mathbb{K}$, $r(y, \cdot, b)$ is upper semicontinuous (u.s.c.) on $A(y)$, and $r(y, a, \cdot)$ is lower semicontinuous (l.s.c.) on $B(y)$.

(c) For each $(y, a, b) \in \mathbb{K}$ and each bounded measurable function v on $\mathbf{Y}$, the functions

$$\int_{\mathbf{Y}} v(z)Q(dz|y, \cdot, b) \quad \text{and} \quad \int_{\mathbf{Y}} v(z)Q(dz|y, a, \cdot) \tag{1.13}$$

are continuous on $A(y)$ and $B(y)$, respectively.

(d) There exists a constant $M > 0$ and a measurable function $w : \mathbf{Y} \to [1, \infty)$ such that

$$|r(y, a, b)| \leq Mw(y), \tag{1.14}$$

and the functions

$$\int_{\mathbf{Y}} w(z)Q(dz|y, \cdot, b) \quad \text{and} \quad \int_{\mathbf{Y}} w(z)Q(dz|y, a, \cdot) \tag{1.15}$$

are continuous on $A(y)$ and $B(y)$, respectively.

(e) There exists a positive constant β such that $1 < \beta < e^{\alpha^*}$, and for all $(y, a, b) \in \mathbb{K}$

$$\int_{\mathbf{Y}} w(z)Q(dz|y, a, b) \leq \beta w(y). \tag{1.16}$$

Definition 1.5.2. *For each measurable function $u : \mathbf{Y} \to \mathbb{R}$ we define the w-norm as*

$$\|u\|_w := \sup_{y \in \mathbf{Y}} \frac{|u(y)|}{w(y)}.$$

Let $\mathbb{B}_w$ be the Banach space of all real-valued measurable functions defined on $\mathbf{Y}$ with finite w-norm.

For $u \in \mathbb{B}_w(\mathbf{Y})$ we define the Shapley operator

$$Tu(y) := \sup_{\varphi^1 \in \mathbb{A}(y)} \inf_{\varphi^2 \in \mathbb{B}(y)} H(u; y, \varphi^1, \varphi^2), \quad y \in \mathbf{Y}, \tag{1.17}$$

where,

$$H(u; y, a, b) := r(y, a, b) + e^{-\alpha} \int_{\mathbf{Y}} u(z) Q(dz|y, a, b), \quad (y, a, b) \in \mathbb{K}. \tag{1.18}$$

Later, we will be able to assure that the infimum and supremum are attained, thus we can replace inf and sup by min and max, respectively. Therefore, we will have

$$Tu(y) := \max_{\varphi^1 \in \mathbb{A}(y)} \min_{\varphi^2 \in \mathbb{B}(y)} H(u; y, \varphi^1, \varphi^2), \quad \forall u \in \mathbb{B}_w(\mathbf{Y}). \tag{1.19}$$

In order to prove the existence of optimal strategies in both, the finite and infinite horizon cases, we first introduce some previous results.

Lemma 1.5.3. *Suppose that Assumptions 1.5.1(c) and 1.5.1 (d) hold. Then the function $u'(y, a, b) := \int u(z) Q(dz|y, a, b)$ is continuous in $a \in A(y)$ and $b \in B(y)$ for every $y \in \mathbf{Y}$ and every function $u \in \mathbb{B}_w(\mathbf{Y})$.*

Proof. Let u be a function in $\mathbb{B}_w(\mathbf{Y})$, so that $|u(y)| \leq mw(y)$ for all $y \in \mathbf{Y}$, where $m := \|u\|_w$. Then $u_m := u + mw$ is a nonnegative function in $\mathbb{B}_w(\mathbf{Y})$, and so it is the limit of a nondecreasing sequence of measurable bounded functions $u^k \in \mathbb{B}_w(\mathbf{Y})$. Now fix $y \in \mathbf{Y}$ and let $\{a^n\}$ be a sequence in $A(y)$ converging to $a \in A(y)$. Then, as $u^k \uparrow u_m$, Assumption 1.5.1(c) yields, for every k

$$\begin{aligned}\liminf_{n\to\infty} \int u_m(z) Q(dz|y, a^n, b) &\geq \liminf_{n\to\infty} \int u^k(z) Q(dz|y, a^n, b) \\ &= \int u^k(z) Q(dz|y, a, b).\end{aligned}$$

Hence, letting $k \to \infty$, by the Monotone Convergence Theorem we have that

$$\begin{aligned}\liminf_{n\to\infty} \int u_m(z) Q(dz|y, a^n, b) &\geq \liminf_{k\to\infty} \int u^k(z) Q(dz|y, a, b) \\ &= \int u_m(z) Q(dz|y, a, b)\end{aligned}$$

and, therefore, $\int u_m(z) Q(dz|y, \cdot, b)$ is l.s.c. on $A(y)$, which implies that $u'(y, \cdot, b)$ is l.s.c. on $A(y)$. In other words, $u'(y, \cdot, b)$ is l.s.c. on $A(y)$ for every function u in $\mathbb{B}_w(\mathbf{Y})$. Hence, if we now apply the latter fact to $-u$ in lieu of u, we see that $u'(y, \cdot, b)$ is also u.s.c. Thus $u'(y, \cdot, b)$ is continuous on $A(y)$. In a similar way, we can prove that $u'(y, a, \cdot)$ is continuous on $B(y)$. $\square$

Lemma 1.5.4. *Under the Assumption 1.5.1, for each* $u \in \mathbb{B}_w(\mathbf{Y})$*:*

(a) $Tu(y) = \min_{\varphi^2 \in \mathbb{B}(y)} \max_{\varphi^1 \in \mathbb{A}(y)} H(u; y, \varphi^1, \varphi^2)$*;*

(b) There exist $\varphi^1_*(y) \in \mathbb{A}(y)$ *and* $\varphi^2_*(y) \in \mathbb{B}(y)$ *such that,*

$$\begin{aligned} Tu(y) &= \max_{\varphi^1 \in \mathbb{A}(y)} H(u; y, \varphi^1, \varphi^2_*(y)) \\ &= \min_{\varphi^2 \in \mathbb{B}(y)} H(u; y, \varphi^1_*(y), \varphi^2) \\ &= H(u; y, \varphi^1_*(y), \varphi^1_*(y)) \quad \forall\ y \in \mathbf{Y}; \end{aligned}$$

(c) Tu *is an element of* $\mathbb{B}_w(\mathbf{Y})$*.*

Proof. Let u be an arbitrary function in $\mathbb{B}_w(\mathbf{Y})$.

(a) Lemma 1.5.3 yields the continuity on $a \in A(y)$ and $b \in B(y)$ of the integral in (1.18). This fact and the Assumption 1.5.1 (b) imply that, for each $(y, a, b) \in \mathbb{K}$, the function $H(u; y, \cdot, b)$ is u.s.c. on $A(y)$ and $H(u; y, a, \cdot)$ is l.s.c. on $B(y)$. Then, the function $H(u; y, \varphi^1, \varphi^2)$ is u.s.c. on $\varphi^1 \in \mathbb{A}(y)$ and l.s.c. on $\varphi^2 \in \mathbb{B}(y)$. Moreover, $H(u; y, \varphi^1, \varphi^2)$ is concave on φ^1 and convex on φ^2. Thus, by Fan's Minimax Theorem (see Theorem A.4.2) we prove part (a).

(b) Let us define

$$H_1(y, \varphi^1) := \min_{\varphi^2 \in \mathbb{B}(y)} H(u; y, \varphi^1, \varphi^2),$$

for every $y \in \mathbf{Y}$ and $\varphi^1 \in \mathbb{A}(y)$. From the proof of part (a), we can observe that $H_1(y, \cdot)$ is u.s.c. on $\mathbb{A}(y)$. Therefore, by Proposition B.2.5 and Theorem B.3.2, there exists $\varphi^1_*(y) \in \mathbb{A}(y)$ such that

$$H_1(y, \varphi^1_*(y)) = \max_{\varphi^1 \in \mathbb{A}(y)} H_1(y, \varphi^1).$$

As consequence, we obtain

$$H_1(y, \varphi^1_*(y)) = \max_{\varphi^1 \in \mathbb{A}(y)} \min_{\varphi^2 \in \mathbb{B}(y)} H(u; y, \varphi^1, \varphi^2). \tag{1.20}$$

By (1.20), we have

$$Tu(y) = \min_{\varphi^2 \in \mathbb{B}(y)} H(u; y, \varphi^1_*(y), \varphi^2).$$

Similarly, let

$$H_2(y, \varphi^2) := \max_{\varphi^1 \in \mathbb{A}(y)} H(u; y, \varphi^1, \varphi^2).$$

Then, there exists $\varphi^2_* \in \mathbb{B}(y)$ such that

$$Tu(y) = \max_{\varphi^1 \in \mathbb{A}(y)} H(u; y, \varphi^1, \varphi^2_*(y)).$$

(c) Since $|u(\cdot)| \leq \|u\|_w w(\cdot)$, from (1.14) and (1.16) we get, for any $(y, a, b) \in \mathbb{K}$,

$$\begin{aligned} |H(u; y, a, b)| &\leq Mw(y) + \|u\|_w e^{-\alpha} \int_Y w(z) Q(dz|y, a, b) \\ &\leq (M + \beta \|u\|_w e^{-\alpha^*}) w(y). \end{aligned}$$

Thus, by part (b) and the previous equation, Tu belongs to $\mathbb{B}_w(\mathbf{Y})$.

□

1.6 Existence of optimal strategies

Our objective in this section is to prove in each case, finite-horizon and infinite-horizon, the existence of a value of the game and an optimal pair of strategies.

1.6.1 The finite-horizon stochastic game

Let n be a positive integer. The n-stage stochastic game in which the players play up to time n is said to be a *finite-horizon game.* Let π^1 and π^2 be the strategies of players 1 and 2, respectively. Then the expected payoff $V_n(y, \pi^1, \pi^2)$ in such a game is given by (1.6).

Remark 1.6.1. *When we are working with the n-stage stochastic game, strategies $\pi^1 \in \Pi^1$ and $\pi^2 \in \Pi^2$ have n components and take the form $\pi^1 := (\pi^1_0, \pi^1_1, \dots, \pi^1_{n-1})$ and $\pi^2 := (\pi^2_0, \pi^2_1, \dots, \pi^2_{n-1})$.*

Theorem 1.6.2. *Suppose that Assumption 1.5.1 holds. Then the stochastic game with finite horizon has a value and both players have optimal Markov strategies. Moreover, if V^*_n is the value function for the n-stage game, then $V^*_n \in \mathbb{B}_w(\mathbf{Y})$ and $V^*_n(y) = TV^*_{n-1}(y)$ for each $n \geq 2$.*

Proof. Let us define the sequence of functions:

$$\begin{aligned}
\tilde{V}_0(y) &:= 0, \\
\tilde{V}_n(y) &:= TV_{n-1}(y) \\
&= \min_{\varphi^2 \in \mathbb{B}(y)} \max_{\varphi^1 \in \mathbb{A}(y)} \left\{ r(y, \varphi^1, \varphi^2) + e^{-\alpha} \int_{\mathbf{Y}} \tilde{V}_{n-1}(z) Q(dz|y, \varphi^1, \varphi^2) \right\}.
\end{aligned}$$

Since $\tilde{V}_0 \equiv 0 \in \mathbb{B}_w(\mathbf{Y})$, by Lemma 1.5.4 (c) we have that for any $n \in \mathbb{N}_0$ $\tilde{V}_n \in \mathbb{B}_w(\mathbf{Y})$. Then, for each $n \in \mathbb{N}_0$, by Lemma 1.5.4 (a),

$$\begin{aligned}
\tilde{V}_n(y) &= T\tilde{V}_{n-1}(y) \\
&= \min_{\varphi^2 \in \mathbb{B}(y)} \max_{\varphi^1 \in \mathbb{A}(y)} H(\tilde{V}_{n-1}; y, \varphi^1, \varphi^2) \\
&= \max_{\varphi^1 \in \mathbb{A}(y)} \min_{\varphi^2 \in \mathbb{B}(y)} H(\tilde{V}_{n-1}; y, \varphi^1, \varphi^2)
\end{aligned} \tag{1.21}$$

and by Lemma 1.5.4(b), there exist $\psi_{n-1} \in \Phi^1$ and $\rho_{n-1} \in \Phi^2$ such that

$$\begin{aligned}
\tilde{V}_n(y) &= T\tilde{V}_{n-1}(y) \\
&= \min_{\varphi^2 \in \mathbb{B}(y)} H(\tilde{V}_{n-1}; y, \psi_{n-1}, \varphi^2) \\
&= \max_{\varphi^1 \in \mathbb{A}(y)} H(\tilde{V}_{n-1}; y, \varphi^1, \rho_{n-1}) \\
&= H(\tilde{V}_{n-1}; y, \psi_{n-1}, \rho_{n-1}).
\end{aligned} \tag{1.22}$$

For $n \in \mathbb{N}_0$, we define

$$\pi^1_n := \{\psi_{n-1}, \psi_{n-2}, \dots, \psi_0\}, \tag{1.23}$$
$$\pi^2_n := \{\rho_{n-1}, \rho_{n-2}, \dots, \rho_0\}, \tag{1.24}$$

where $\psi_i \in \Phi^1$ and $\rho_i \in \Phi^2$ are the respective maximizer and minimizer of $\tilde{V}_{i+1}$ as in (1.22) for $i = 0, 1, \ldots, n-1$.

We will prove that, for every n, $\tilde{V}_n = L_n = U_n$, and that π_n^1 in (1.23) and π_n^2 in (1.24) are optimal Markov strategies for players 1 and 2, respectively.

We will proceed by mathematical induction over the game horizon n.

For $n = 1$, from the definition of V_1 in (1.7),

$$V_1(y, \pi^1, \pi^2) = E_y^{\pi^1, \pi^2}[r(y_0, a_0, b_0)] = r(y, \pi^1, \pi^2). \tag{1.25}$$

Notice that strategies in this case have only one component and take the form $\pi^1 := \varphi^1 \in \Phi^1$ and $\pi^2 := \varphi^2 \in \Phi^2$, that is, they are Markov strategies. By (1.25) and the definition (1.18) of H,

$$H(0; y, \pi^1, \pi^2) = V_1(y, \pi^1, \pi^2),$$

and by (1.21)

$$\begin{aligned}
\tilde{V}_1(y) &= \min_{\varphi^2 \in \mathbb{B}(y)} \max_{\varphi^1 \in \mathbb{A}(y)} V_1(y, \varphi^1, \varphi^2) = U_1(y) \\
&= \max_{\varphi^1 \in \mathbb{A}(y)} \min_{\varphi^2 \in \mathbb{B}(y)} V_1(y, \varphi^1, \varphi^2) = L_1(y) \\
&= V_1^*(y).
\end{aligned}$$

Thus, since $\pi_1^1 = \psi_0$ and $\pi_1^2 = \rho_0$, by (1.22)

$$V_1^*(y) = V_1(y, \psi_0, \rho_0),$$

which implies that (π_1^1, π_1^2) is a pair of optimal Markov strategies for the 1-stage game.

Suppose (the induction hypothesis) that for $n = k-1$,

$$\tilde{V}_{k-1}(y) = L_{k-1}(y) = U_{k-1}(y) = V_{k-1}(y, \pi_{k-1}^1, \pi_{k-1}^2).$$

We now prove that this fact holds for $n = k$. Indeed, let $\hat{\pi}_k^2 = (\hat{\rho}_{k-1}, \hat{\rho}_{k-2}, \ldots, \hat{\rho}_0)$ be an arbitrary strategy for player 2. Then,

$$\begin{aligned}
\tilde{V}_k(y) = T\tilde{V}_{k-1}(y) &= \min_{\varphi^2 \in \mathbb{B}(y)} \max_{\varphi^1 \in \mathbb{A}(y)} \left\{ r(y, \varphi^1, \varphi^2) + e^{-\alpha} \int_{\mathbf{Y}} \tilde{V}_{k-1}(z) Q(dz|y, \varphi^1, \varphi^2) \right\} \\
&= \min_{\varphi^2 \in \mathbb{B}(y)} \left\{ r(y, \psi_{k-1}, \varphi^2) + e^{-\alpha} \int_{\mathbf{Y}} \tilde{V}_{k-1}(z) Q(dz|y, \psi_{k-1}, \varphi^2) \right\} \\
&\leq r(y, \psi_{k-1}, \hat{\rho}_{k-1}) + e^{-\alpha} \int_{\mathbf{Y}} \tilde{V}_{k-1}(z) Q(dz|y, \psi_{k-1}, \hat{\rho}_{k-1}).
\end{aligned}$$

Iterating this inequality we obtain

$$\tilde{V}_k(y) \leq E_y^{\pi_k^1, \hat{\pi}_k^2} \left[\sum_{t=0}^{k-1} e^{-S_t} r(y_t, a_t, b_t) \right] = V_k(y, \pi_k^1, \hat{\pi}_k^2). \tag{1.26}$$

Similarly, for an arbitrary strategy for player 1 $\hat{\pi}_k^1 = (\hat{\psi}_{k-1}, \hat{\psi}_{k-2}, \ldots, \hat{\psi}_0)$,

$$\tilde{V}_k(y) \geq V_k(y, \hat{\pi}_k^1, \pi_k^2). \tag{1.27}$$

From (1.26) we have

$$\sup_{\pi^1\in\Pi^1}\inf_{\pi^2\in\Pi^2} V_k(y,\pi^1,\pi^2) \geq \inf_{\pi^2\in\Pi^2} V_k(y,\pi^1_k,\pi^2) \geq \tilde{V}_k(y), \tag{1.28}$$

and from (1.27),

$$\inf_{\pi^2\in\Pi^2}\sup_{\pi^1\in\Pi^1} V_k(y,\pi^1,\pi^2) \leq \sup_{\pi^1\in\Pi^1} V_k(y,\pi^1,\pi^2_k) \leq \tilde{V}_k(y). \tag{1.29}$$

From (1.28) and (1.29) we obtain

$$U_k(y) = \inf_{\pi^2\in\Pi^2}\sup_{\pi^1\in\Pi^1} V_k(y,\pi^1,\pi^2) \leq \tilde{V}_k(y) \leq \sup_{\pi^1\in\Pi^1}\inf_{\pi^2\in\Pi^2} V_k(y,\pi^1,\pi^2) = L_k(y),$$

which implies that the value exists and, by the induction hypothesis,

$$V_k^*(y) = \tilde{V}_k(y) = TV_{k-1}^*(y).$$

On the other hand, by (1.26) and (1.27),

$$V_k(y,\pi^1_k,\hat{\pi}^2_k) \geq V_k^*(y) \geq V_k(y,\hat{\pi}^1_k,\pi^2_k) \quad \forall\ (\hat{\pi}^1_k,\hat{\pi}^2_k)\in\Pi^1\times\Pi^2,$$

in particular for $(\pi^1_k,\pi^2_k)\in\Pi^1\times\Pi^2$ we have

$$V_k^*(y) = V_k(y,\pi^1_k,\pi^2_k),$$

that is, (π^1_k,π^2_k) is an optimal pair of Markov strategies. Therefore,

$$V_k^*(y) = TV_{k-1}^*(y) = V_k(y,\pi^1_k,\pi^2_k),$$

which complete the proof. □

1.6.2 The infinite-horizon stochastic game

In this section, we consider infinite-horizon stochastic games. We prove that the game value V^* is a fixed point of the operator T in (1.17), this is, $V^* = TV^*$, and that the sequence $\{V_n^*\}$ converges geometrically to V^* in the w-norm. The results are established under Assumption 1.5.1.

Let us consider the game model $\mathcal{GM}$ in (1.1) with the total expected discounted payoff function $V(y,a,b)$. The corresponding lower and upper values are

$$\begin{aligned} L(y) &:= \sup_{\pi^1\in\Pi^1}\inf_{\pi^2\in\Pi^2} V(y,\pi^1,\pi^2),\\ U(y) &:= \inf_{\pi^2\in\Pi^2}\sup_{\pi^1\in\Pi^1} V(y,\pi^1,\pi^2).\end{aligned}$$

The goal is to show that $L(\cdot) = U(\cdot)$, thus the value of the game $V^*(\cdot)$ exists. In order to prove this fact, we first introduce some preliminary results.

Lemma 1.6.3. *Under Assumption 1.5.1, the operator T defined in* (1.7) *is a contraction operator in $\mathbb{B}_w(\mathbf{Y})$ with modulus $\tau := \beta e^{-\alpha^*}$ (with β as in* (1.14)*).*

Proof. Let us first note that T is a monotone operator, that is, if u and $\tilde{u}$ are functions in $\mathbb{B}_w(\mathbf{Y})$, and $u \geq \tilde{u}$, then $Tu(y) \geq T\tilde{u}(y)$ for every $y \in \mathbf{Y}$. Indeed, if $u, \tilde{u} \in \mathbb{B}_w(\mathbf{Y})$ are functions such that $u \geq \tilde{u}$, then

$$\int_{\mathbf{Y}} u(z)Q(dz|y,a,b) \geq \int_{\mathbf{Y}} \tilde{u}(z)Q(dz|y,a,b), \quad \forall\ (y,a,b) \in \mathbb{K},$$

and this implies that $H(u;y,a,b) \geq H(\tilde{u};y,a,b)$ for every $(y,a,b) \in \mathbb{K}$, which in turn implies that $Tu(y) \geq T\tilde{u}(y)$ for every $y \in \mathbf{Y}$.

On the other hand, it holds, for any real number $k \geq 0$:

$$T(u+kw)(y) \leq Tu(y) + \beta e^{-\alpha} kw(y), \tag{1.30}$$

for every $y \in \mathbf{Y}$ and $u \in \mathbb{B}_w(\mathbf{Y})$.

Now, to verify that T is a contraction, let us choose u and $\tilde{u}$ in $\mathbb{B}_w(\mathbf{Y})$. Since $u \leq \tilde{u} + w\|u - \tilde{u}\|_w$, by applying that T is monotone and (1.30) with $k = \|u - \tilde{u}\|_w$ we have

$$Tu(y) \leq T(\tilde{u} + kw)(y) \leq T\tilde{u}(y) + \beta e^{-\alpha} kw(y),$$

that is,

$$Tu(y) - T\tilde{u}(y) \leq \beta e^{-\alpha}\|u - \tilde{u}\|_w w(y), \quad y \in \mathbf{Y}.$$

Interchanging u and $\tilde{u}$, we obtain

$$Tu(y) - T\tilde{u}(y) \geq -\beta e^{-\alpha}\|u - \tilde{u}\|_w w(y),$$

hence,

$$|Tu(y) - T\tilde{u}(y)| \leq \beta e^{-\alpha}\|u - \tilde{u}\|_w w(y), \quad y \in \mathbf{Y},$$

and since $\alpha^* \leq \alpha$ for all $\alpha \in \mathbf{\Gamma}$, we have

$$|Tu(y) - T\tilde{u}(y)| \leq \beta e^{-\alpha^*}\|u - \tilde{u}\|_w w(y), \quad y \in \mathbf{Y},$$

Therefore, taking $\tau := \beta e^{-\alpha^*}$, we obtain

$$\|Tu - T\tilde{u}\|_w \leq \tau\|u - \tilde{u}\|_w.$$

□

Lemma 1.6.4. *Let M, w, and β be as in Assumption 1.5.1, and moreover, let $\pi^1 \in \Pi^1$ and $\pi^2 \in \Pi^2$ be arbitrary strategies for players 1 and 2, respectively, and $y \in \mathbf{Y}$ the initial state. Then for each $t = 0, 1, \ldots,$*

(a) $E_y^{\pi^1,\pi^2} w(y_t) \leq \beta^t w(y)$;

(b) $|E_y^{\pi^1,\pi^2} r(y_t, a_t, b_t)| \leq M\beta^t w(y)$;

(c) $\lim_{t\to\infty} e^{-S_t} E_y^{\pi^1,\pi^2} u(y_t) = 0$ *for each* $u \in \mathbb{B}_w(\mathbf{Y})$.

Proof.

(a) Case $t = 0$ follows directly. Now, if $t \geq 1$, by Assumption 1.5.1(e) we have

$$E_y^{\pi^1,\pi^2}[w(y_t)|h_{t-1}, a_{t-1}, b_{t-1}] = \int_{\mathbf{Y}} w(z)Q(dz|y_{t-1}, a_{t-1}, b_{t-1})$$
$$\leq \beta w(y_{t-1}).$$

Therefore, we have the inequality

$$E_y^{\pi^1,\pi^2} w(y_t) \leq \beta E_y^{\pi^1,\pi^2} w(y_{t-1}).$$

Iteration of this inequality yields part (a).

(b) From Assumption 1.5.1(d) we have

$$|r(y_t, a_t, b_t)| \leq Mw(y_t), \quad \forall\, t = 0, 1, \ldots,$$

and by part (a),

$$|r(y_t, a_t, b_t)| \leq M\beta^t w(y).$$

(c) By Definition 1.5.2 of w-norm and part (a), we obtain

$$E_y^{\pi^1,\pi^2}|u(y_t)| \leq \|u\|_w E_y^{\pi^1,\pi^2} w(y_t) \leq \|u\|_w \beta^t w(y),$$

which implies (c).

□

Definition 1.6.5. *Let $\varphi^1 \in \Phi^1$, $\varphi^2 \in \Phi^2$, and H be as in* (1.18)*. We define the operator*

$$R_{\varphi^1\varphi^2} : \mathbb{B}_w(\mathbf{Y}) \to \mathbb{B}_w(\mathbf{Y}), \quad u \mapsto R_{\varphi^1\varphi^2} u$$

by

$$R_{\varphi^1\varphi^2} u(y) := H(u; y, \varphi^1(y), \varphi^2(y)) \quad \forall\, y \in \mathbf{Y}. \tag{1.31}$$

Lemma 1.6.6. *The operator $R_{\varphi^1\varphi^2}$ is a contraction operator with modulus $\tau := \beta e^{-\alpha^*}$ on $\mathbb{B}_w(\mathbf{Y})$, and $V(y, \varphi^1, \varphi^2)$ is its unique fixed point in $\mathbb{B}_w(\mathbf{Y})$.*

Proof. Using similar arguments as those in the proof of Lemma 1.6.3, it follows that $R_{\varphi^1\varphi^2}$ is a contraction operator on $\mathbb{B}_w(\mathbf{Y})$ with modulus $\tau := \beta e^{-\alpha^*}$. By Banach's fixed point Theorem A.3.2 (see Appendix A.3), $R_{\varphi^1\varphi^2}$ has a unique fixed point $u_{\varphi^1\varphi^2}$ in $\mathbb{B}_w(\mathbf{Y})$, this is,

$$u_{\varphi^1\varphi^2} = R_{\varphi^1\varphi^2} u_{\varphi^1\varphi^2}, \tag{1.32}$$

from which we have that $u_{\varphi^1\varphi^2}$ is the unique solution in $\mathbb{B}_w(\mathbf{Y})$ of the equation

$$u_{\varphi^1\varphi^2}(y) = r(y, \varphi^1(y), \varphi^2(y)) + e^{-\alpha} \int_{\mathbf{Y}} u_{\varphi^1\varphi^2}(z)Q(dz|y, \varphi^1(y), \varphi^2(y)), \quad \forall y \in \mathbf{Y}. \tag{1.33}$$

Moreover, iterating (1.32) and (1.33) we obtain

$$u_{\varphi^1\varphi^2}(y) = R^n_{\varphi^1\varphi^2} u_{\varphi^1\varphi^2}(y)$$
$$= E_y^{\varphi^1\varphi^2}\left[\sum_{t=0}^{n-1} e^{-S_t} r(y_t, \varphi^1(y_t), \varphi^2(y_t))\right] + e^{-S_n} E_y^{\varphi^1\varphi^2} u_{\varphi^1\varphi^2}(y_n)$$

for each $y \in \mathbf{Y}$ and $n \geq 1$, where $E_y^{\varphi^1\varphi^2} u(y_n) = \int_{\mathbf{Y}} Q^n(dz|y, \varphi^1, \varphi^2)$ and $Q^n(\cdot|y, \varphi^1, \varphi^2)$ is the n-th transition of the kernel of the Markov process $\{y_t\}$ when players use the strategies φ^1 and φ^2. Finally, by Lemma 1.6.4(c) and letting $t \to \infty$ we have, from the definition of V, that $u_{\varphi^1\varphi^2}(y) = V(y, \varphi^1, \varphi^2)$ for all $y \in \mathbf{Y}$. □

Now we present the main result in this section.

Theorem 1.6.7. *Suppose that Assumption 1.5.1 holds. Let β and M be the constants in Assumption 1.5.1(e), and $\tau := \beta e^{-\alpha^*}$. Then:*

(a) The value function V^* is the unique function in the space $\mathbb{B}_w(\mathbf{Y})$ that satisfies the equation $TV^* = V^*$, and

$$\|V_n^* - V^*\|_w \leq \frac{M\tau^n}{1-\tau} \quad \forall\, n = 1, 2, \ldots. \tag{1.34}$$

(b) There exists a pair of optimal strategies.

Proof. By Lemma 1.6.3 and Banach's Fixed Point Theorem (Proposition A.3.2), T has a unique fixed point $\tilde{V}$ in $\mathbb{B}_w(\mathbf{Y})$, i.e.,

$$T\tilde{V} = \tilde{V}, \tag{1.35}$$

and

$$\|T^n u - \tilde{V}\|_w \leq \tau^n \|u - \tilde{V}\|_w, \quad \forall\, u \in \mathbb{B}_w(\mathbf{Y}),\ n = 0, 1, \ldots \tag{1.36}$$

Hence, to prove part (a) we need to show that

(i) V^* is in $\mathbb{B}_w(\mathbf{Y})$, with norm $\|V^*\|_w \leq \frac{M}{1-\tau}$, and

(ii) $V^* = \tilde{V}$.

By Theorem 1.6.2,

$$V_n^* = TV_{n-1}^* = T^n V_0^* \quad \forall\, n = 0, 1, \ldots,\ V_0^* = 0, \tag{1.37}$$

thus, (1.34) will follow from (1.37) and (1.36) with $u \equiv 0$.

To prove (i), let $\pi^2 \in \Pi^2$ and $\pi^2 \in \Pi^2$ be arbitrary strategies for players 1 and 2, respectively, and let $y \in \mathbf{Y}$ be an arbitrary initial state, then (i) follows from Lemma 1.6.4 (b) since a direct calculation gives

$$|V(y, \pi^1, \pi^2)| \leq \sum_{t=0}^{\infty} e^{-t\alpha^*} E_y^{\pi^1, \pi^2} |r(y_t, a_t, b_t)| \leq \frac{Mw(y)}{1-\tau}$$

with $\tau := \beta e^{-\alpha^*}$. Thus, as $\pi^1 \in \Pi^1$, $\pi^2 \in \Pi^2$, and $y \in \mathbf{Y}$ are arbitrary,

$$|V^*(y)| \leq \frac{Mw(y)}{1-\tau}.$$

To prove (ii), let us note that by the equality $\tilde{V} = T\tilde{V}$ and Lemma 1.5.4, there exist $\varphi_*^1 \in \Phi^1$ and $\varphi_*^2 \in \Phi^2$ such that, for all $y \in \mathbf{Y}$

$$\begin{aligned} \tilde{V}(y) &= \sup_{\varphi^1(y) \in \mathbb{A}(y)} H(\tilde{V}; y, \varphi^1(y), \varphi_*^2(y)) \\ &= \inf_{\varphi^2(y) \in \mathbb{B}(y)} H(\tilde{V}; y, \varphi_*^1(y), \varphi^2(y)) \\ &= H(\tilde{V}; y, \varphi_*^1(y), \varphi_*^2(y)). \end{aligned} \tag{1.38}$$

Observe that (1.38) can be written as

$$\tilde{V}(y) = r(y, \varphi_*^1(y), \varphi_*^2(y)) + e^{-\alpha} \int_{\mathbf{Y}} \tilde{V}(z) Q(dz|y, \varphi_*^1(y), \varphi_*^2(y)).$$

Then it follows from Lemma 1.6.6 that $\tilde{V}(y) = V(y, \varphi_*^1, \varphi_*^2)$. Therefore, we have

$$V(y, \varphi_*^1, \varphi_*^2) = \sup_{\varphi^1 \in \mathbb{A}(y)} \left[r(y, \varphi^1, \varphi_*^2) + e^{-\alpha} \int_{\mathbf{Y}} V(z, \varphi_*^1, \varphi_*^2) Q(dz|y, \varphi^1, \varphi_*^2) \right]$$

for all $y \in \mathbf{Y}$. Then by standard dynamic programming results, it follows that

$$V(y, \varphi_*^1, \varphi_*^2) = \sup_{\pi^1 \in \Pi^1} V(y, \pi^1, \varphi_*^2).$$

Similarly, considering the infimum in (1.38) we get

$$V(y, \varphi_*^1, \varphi_*^2) = \inf_{\pi^2 \in \Pi^2} V(y, \varphi_*^1, \pi^2).$$

Consequently,

$$V(y, \varphi_*^1, \varphi_*^2) = \sup_{\pi^1 \in \Pi^1} V(y, \pi^1, \varphi_*^2) \geq \inf_{\pi^2 \in \Pi^2} \sup_{\pi^1 \in \Pi^1} V(y, \pi^1, \pi^2),$$

and, on the other hand,

$$V(y, \varphi_*^1, \varphi_*^2) = \inf_{\pi^2 \in \Pi^2} V(y, \varphi_*^1, \pi^2) \leq \sup_{\pi^1 \in \Pi^1} \inf_{\pi^2 \in \Pi^2} V(y, \pi^1, \pi^2).$$

Hence,

$$\inf_{\pi^2 \in \Pi^2} \sup_{\pi^1 \in \Pi^1} V(y, \pi^1, \pi^2) = V(y, \varphi_*^1, \varphi_*^2) = \sup_{\pi^1 \in \Pi^1} \inf_{\pi^2 \in \Pi^2} V(y, \pi^1, \pi^2).$$

This proves that the stochastic game has a value, that the value is $V^*(y) = V(y, \varphi_*^1, \varphi_*^2) = \tilde{V}(y)$ for all $y \in \mathbf{Y}$, and that φ_*^1 and φ_*^2 are optimal strategies for players 1 and 2, respectively. □

1.7 Example

Let us consider an infinite horizon game, with state space $\mathbf{X} = \mathbb{R}$, discount factor set $\mathbf{\Gamma} = [\alpha^*, \infty)$, and actions sets $\mathbf{A} = \mathbf{B} = \mathbb{R}$. The admissible actions sets $A(x, \alpha)$ and $B(x, \alpha)$ are compact for each $(x, \alpha) \in \mathbf{X} \times \mathbf{\Gamma}$.

The state process $\{x_t\}$ and the discount process $\{\alpha_t\}$ evolve according to the coupled difference equations

$$x_{n+1} = h(a_n, b_n) x_n + \epsilon_n \tag{1.39}$$
$$\alpha_{n+1} = \tilde{\gamma} \alpha_n + \eta_n \tag{1.40}$$

$t \in \mathbb{N}_0$, (x_0, α_0) given, where $\tilde{\gamma} < 1$ is a positive constant; $h : \mathbf{A} \times \mathbf{B} \to (0, \tilde{\gamma}]$ is a given continuous function; $\{\epsilon_n\}$ and $\{\eta_n\}$ are independent sequences of i.i.d. random variables, and independent of (x_0, α_0), that take values in $S_1 = [0, \bar{s}_1]$ and $S_2 = [0, \bar{s}_2]$, for some positive constants $\bar{s}_1$ and $\bar{s}_2$, and having continuous common density ρ_ϵ and ρ_η, respectively.

Notice that letting $\chi_n := (\epsilon_n, \eta_n)^T$, we have that χ_n is a random vector taking values in $\mathbf{S} := S_1 \times S_2$ with continuous joint density $\rho := \rho_\epsilon \rho_\eta$ that represents the state-discount random disturbance process. Denoting by $\{y_n\} := \left\{ \begin{pmatrix} x_n \\ \alpha_n \end{pmatrix} \right\}$ the joint state-discount process, we can write (1.39) and (1.40) as

$$\begin{aligned} y_{n+1} = \begin{pmatrix} x_{n+1} \\ \alpha_{n+1} \end{pmatrix} &= h(a_n, b_n) \begin{pmatrix} x_{n+1} \\ 0 \end{pmatrix} + \tilde{\gamma} \begin{pmatrix} 0 \\ \alpha_{n+1} \end{pmatrix} + \chi_n \\ &= H(y_n, a_n, b_n) + \chi_n, \end{aligned} \tag{1.41}$$

where y_n takes values in $\mathbf{Y} := \mathbf{X} \times \mathbf{\Gamma}$ and $H : \mathbf{X} \times \mathbf{\Gamma} \times \mathbf{A} \times \mathbf{B} \to \mathbf{X} \times \mathbf{\Gamma}$. Let $\bar{\chi} := E(\sqrt{\epsilon^2 + \eta^2})$, and let us suppose that $\alpha^* > \ln(1 + \bar{\chi})$.

The one-stage payoff function $r(y, a, b)$ is an arbitrary measurable function, which is u.s.c. in a and l.s.c. in b, and satisfies

$$|r(y, a, b)| \leq M(\|y\| + c), \quad y \in \mathbf{Y}, \tag{1.42}$$

for some $M > 0$ and $c > 1$, where $\|\cdot\|$ is the usual euclidian norm in $\mathbb{R}^2$. Let

$$w(y) := \|y\| + c.$$

Clearly, Assumption 1.5.1(a)-(b) are satisfied.

Observe that, in this example, the transition law of the state-discount process is defined by

$$Q(D|y, a, b) := \int_{\mathbf{S}} 1_D(H(y, a, b) + s)\rho(s)ds, \quad D \in \mathcal{B}(\mathbf{Y}). \tag{1.43}$$

To show that Assumption 1.5.1(d) is satisfied, we have to verify that

$$\int_{\mathbf{Y}} w(z)Q(dz|y, \cdot, b) \text{ and } \int_{\mathbf{Y}} w(z)Q(dz|y, a, \cdot) \tag{1.44}$$

are continuous on $A(y)$ and $B(y)$, respectively. Accordingly to (1.43), and from the definition of w,

$$\int_{\mathbf{Y}} w(z)Q(dz|y, \cdot, b) = \int_S (\|H(y, \cdot, b) + s\| + c)\rho(s)ds, \tag{1.45}$$

and,

$$\int_{\mathbf{Y}} w(z)Q(dz|y, a, \cdot) = \int_S (\|H(y, a, \cdot) + s\| + c)\rho(s)ds. \tag{1.46}$$

Then, we just have to show that

$$\int_S (\|H(y, \cdot, b) + s\| + c)\rho(s)ds \text{ and } \int_S (\|H(y, a, \cdot) + s\| + c)\rho(s)ds \tag{1.47}$$

are continuous on $A(y)$ and $B(y)$, respectively. Which follows from the continuity of the norm and H, and the Monotone Convergence Theorem A.2.2 (see Appendix A).

We will now proceed to verify Assumption 1.5.1(e). Let $\beta := (1 + \bar{\chi})$, thus

$$
\begin{aligned}
\int w(z)Q(dz|y,a,b) &= \int w(H(y,a,b)+s)\rho(s)ds \\
&\leq \int [\|\tilde{\gamma}y+s\|+c]\rho(s)ds \\
&\leq \|\tilde{\gamma}y\| + \int \|s\|\rho(s)ds + c \\
&\leq \|y\| + c + \bar{\chi} \\
&\leq (1+\bar{\chi})w(y) \\
&= \beta w(y).
\end{aligned}
$$

Finally, we have to verify that Assumption 1.5.1 (c) holds. Indeed, let v be an arbitrary bounded function, such that $|v(y)| \leq \bar{v}$ for all $y \in \mathbf{Y}$. Observe that for $y \in \mathbf{Y}$, $a \in A(y)$ and $b \in B(y)$,

$$
\int v(z)Q(dz|y,a,b) = \int v(H(y,a,b)+s)\rho(s)ds.
$$

By applying a change of variable, we get

$$
\int v(z)Q(dz|y,a,b) = \int v(z)\rho(z-H(y,a,b))dz.
$$

Let $\{(a_n,b_n)\}$ be a convergent sequence in $A(y)\times B(y)$ such that $(a_n,b_n) \to (a,b)$, then

$$
\begin{aligned}
\int |v(z)\rho(z-&H(y,a_n,b_n)) - v(z)\rho(z-H(y,a,b))|dz \\
&\leq \bar{v}\int |\rho(z-H(y,a_n,b_n)) - \rho(z-H(y,a,b))|dz.
\end{aligned} \tag{1.48}
$$

Since $|\rho(y)| \leq 1$ for all $y \in \mathbf{Y}$, then $|\rho(z-H(y,a_n,b_n)) - \rho(z-H(y,a,b))| \leq 2$, thus letting $n\to\infty$ in both sides of equation (1.48), by the Dominated Convergence Theorem, we obtain

$$
\begin{aligned}
\lim_{n\to\infty}\int |v(z)\rho(z-&H(y,a_n,b_n)) - v(z)\rho(z-H(y,a,b))|dz \\
&\leq \bar{v}\int \lim_{n\to\infty}|\rho(z-H(y,a_n,b_n)) - \rho(z-H(y,a,b))|dz = 0
\end{aligned}
$$

which is equal to zero since ρ and H are both continuous in $A(y)\times B(y)$. Therefore, by Proposition B.1.3 we conclude that Assumption 1.5.1 (e) holds.

Chapter 2

Markov games with state-actions-dependent discount factors

2.1 Introduction

In this chapter we deal with a discrete-time zero-sum Markov game under a discounted optimality criterion with state-actions-dependent discount factors of the form $\alpha(x_n, a_n, b_n)$, where x_n, a_n, and b_n represent the state and the actions of the players at time n, respectively. We begin presenting the corresponding game model on which we will define our problem. We also impose some assumptions implying the existence of a value as well as the existence of a stationary pair of optimal strategies. Finally in order to illustrate our results, we present an example of a game with random horizon.

For an easy reference, we again introduce some definitions and notation used in previous chapter, adapted to this class of games.

2.2 Game model

A zero-sum Markov game model with state-actions-dependent discount factors is defined by the following collection

$$\mathcal{GM} := (\mathbf{X}, \mathbf{A}, \mathbf{B}, \mathbb{K}_{\mathbf{A}}, \mathbb{K}_{\mathbf{B}}, Q, \alpha, r) \tag{2.1}$$

where:

- The state space $\mathbf{X}$ is a non-empty Borel space.
- The actions sets $\mathbf{A}$ and $\mathbf{B}$ for players 1 and 2, respectively, are both non-empty Borel spaces.
- The constraint sets $\mathbb{K}_{\mathbf{A}}$ and $\mathbb{K}_{\mathbf{B}}$ are non-empty Borel subsets of $\mathbf{X} \times \mathbf{A}$ and $\mathbf{X} \times \mathbf{B}$ respectively. For each $x \in \mathbf{X}$, the x-sections

$$A(x) := \{a \in \mathbf{A} \mid (x, a) \in \mathbb{K}_{\mathbf{A}}\}$$

and

$$B(x) := \{b \in \mathbf{B} \mid (x,b) \in \mathbb{K}_{\mathbf{B}}\}$$

are non-empty Borel subsets, and represent the admissible actions sets for players 1 and 2, respectively, when the system is in the state x.

The set

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

is Borel subset of $\mathbf{X} \times \mathbf{A} \times \mathbf{B}$.

- Q is a stochastic kernel (s.k.) on $\mathbf{X}$ given $\mathbb{K}$, and represents the transition law of the process.
- $\alpha : \mathbb{K} \to (0,1)$ is a measurable function that represents the discount factor.
- $r : \mathbb{K} \to \mathbb{R}$ is a measurable function which represents the one-stage payoff function.

Interpretation. The game model $\mathcal{GM}$ represents a controlled stochastic system and have the following interpretation. At the initial state $x_0 \in \mathbf{X}$, the players independently choose actions $a_0 \in A(x_0)$ and $b_0 \in B(x_0)$. Then the player 1 receives a payoff $r(x_0,a_0,b_0)$ from player 2, and the game jumps to a new state x_1 according to the transition law $Q(\cdot|x_0,a_0,b_0)$. Once the system is in state x_1 the players select actions $a_1 \in A(x_1)$ and $b_1 \in B(x_1)$ and player 1 receives a payoff $r(x_1,a_1,b_1)$ from player 2. Next the system moves to a state x_2 and the process is repeated over and over again. In general, at stage $n \in \mathbb{N}$, player 1 receives from player 2 $r(x_n,a_n,b_n)$ and the discounted payoff takes the form

$$\Gamma_n r(x_n,a_n,b_n), \tag{2.2}$$

where

$$\Gamma_n := \prod_{k=0}^{n-1} \alpha(x_k,a_k,b_k) \text{ if } n \in \mathbb{N}, \text{ and } \Gamma_0 = 1. \tag{2.3}$$

Thus the goal of player 1 is to maximize the total expected discounted payoff defined by the accumulation of the one-stage payoffs (2.2) over an infinite horizon, whereas the goal of player 2 is to minimize such payoff.

2.2.1 Strategies

Let $\mathbb{H}_0 := \mathbf{X}$ and $\mathbb{H}_n := \mathbb{H}_{n-1} \times \mathbf{X}$ for $n \in \mathbb{N}$. For each $n \in \mathbb{N}_0$ an element $h_n \in \mathbb{H}_n$ takes the form

$$h_n := (x_0,a_0,b_0,x_1,\ldots,x_{n-1},a_{n-1},b_{n-1},x_n),$$

which represents the history of the game up to n.

For each $x \in \mathbf{X}$, let $\mathbb{A}(x) := \mathbb{P}(A(x))$ and $\mathbb{B}(x) := \mathbb{P}(B(x))$. We denote the sets of stochastic kernels

$$\Phi^1 := \{\varphi^1 \in \mathbb{P}(\mathbf{A}|\mathbf{X}) \ : \ \varphi^1(\cdot|x) \in \mathbb{A}(x) \ \ \forall x \in \mathbf{X}\}$$
$$\Phi^2 := \{\varphi^2 \in \mathbb{P}(\mathbf{B}|\mathbf{X}) \ : \ \varphi^2(\cdot|x) \in \mathbb{B}(x) \ \ \forall x \in \mathbf{X}\}.$$

Definition 2.2.1. *A strategy for player 1 is a sequence $\pi^1 = \{\pi_n^1\}$ of stochastic kernels $\pi_n^1 \in \mathbb{P}(\mathbf{A}|\mathbb{H}_t)$ such that*

$$\pi_t^1(A(x_t)|h_t) = 1, \quad \forall\ h_t \in \mathbb{H}_t,\ t \in \mathbb{N}_0. \tag{2.4}$$

We denote by Π^1 the family of all strategies for player 1.

Definition 2.2.2. *A strategy $\pi^1 = \{\pi_t^1\}$ for player 1 is called:*

(a) a Markov strategy if $\pi_t^1 \in \Phi^1$ for all $t \in \mathbb{N}_0$, this is, each π_t depends only on the current state $x \in \mathbf{X}$ of the system. The set of all Markov strategies for player 1 is denoted by Π_M^1.

(b) a stationary (Markov) strategy if $\pi_t^1(\cdot|h_n) = \varphi^1(\cdot|x_n)$ for all $h_n \in \mathbb{H}_n$, $n \in \mathbb{N}_0$, for some stochastic kernel φ^1 in Φ^1, so that $\pi^1 = \{\varphi^1, \varphi^1, \ldots\} = \{\varphi^1\}$. The set of all stationary strategies for player 1 is denoted by Π_S^1.

We have the following relations

$$\Pi_S^1 \subset \Pi_M^1 \subset \Pi^1.$$

The sets of all strategies Π^2, Markov strategies Π_M^2 and stationary strategies Π_S^2 corresponding to player 2 are defined similarly.

2.2.2 The game process

Similar to Section 1.2.2, we define the game process as follows. Let $(\Omega', \mathcal{F}')$ be the measurable space consisting of the sample space $\Omega' := \mathbb{K}^\infty$ and its product σ-algebra $\mathcal{F}'$. Following standard arguments, for each pair of strategies $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$ and initial state $x_0 = x \in \mathbf{X}$, by Ionescu Tulcea's Theorem, there exists a unique probability measure $P_x^{\pi^1,\pi^2}$ and a stochastic process $\{(x_n, a_n, b_n)\}$, where x_n, a_n and b_n represent the state and the actions of player 1 and 2, respectively, at stage $n \in \mathbb{N}_0$, satisfying

$$\begin{aligned} P_x^{\pi^1,\pi^2}[x_0 \in X] &= \delta_x(X), \quad X \in \mathcal{B}(\mathbf{X}); && (2.5)\\ P_x^{\pi^1,\pi^2}[a_n \in A, b_n \in B|h_n] &= \pi_n^1(A|h_n)\pi_n^2(B|h_n), A \in \mathcal{B}(\mathbf{A}), B \in \mathcal{B}(\mathbf{B}); && (2.6)\\ P_x^{\pi^1,\pi^2}[x_{n+1} \in X|h_n, a_n, b_n] &= Q(X|x_n, a_n, b_n), \quad X \in \mathcal{B}(\mathbf{X}), && (2.7) \end{aligned}$$

where $\delta_x(\cdot)$ is the Dirac measure concentrated at x. We denote by $E_x^{\pi^1,\pi^2}$ the expectation operator with respect to $P_x^{\pi^1,\pi^2}$.

The stochastic process $\{x_n\}$ defined on $(\Omega, \mathcal{F}, P_x^{\pi^1,\pi^2})$ is called *game process*.

2.3 Optimality criterion

We suppose that the payoff is discounted by a multiplicative discount rate. That is, a payoff R at stage n is equivalent to a payoff $R\Gamma_n$ at stage 0, where Γ_n was defined in (2.3).

In this sense, when players choose the strategies $\pi^1 \in \Pi^1$ and $\pi^2 \in \Pi^2$, respectively, given the initial state $x_0 = x$ we define the total expected discounted payoff as

$$V(x, \pi^1, \pi^2) := E_x^{\pi^1,\pi^2}\left[\sum_{n=0}^{\infty} \Gamma_n r(x_n, a_n, b_n)\right]. \tag{2.8}$$

2.3.1 Game value

Definition 2.3.1. *The lower and the upper value of the game are defined as:*

$$L(x) := \sup_{\pi^1 \in \Pi^1} \inf_{\pi^2 \in \Pi^2} V(x, \pi^1, \pi^2), \tag{2.9}$$

and

$$U(x) := \inf_{\pi^2 \in \Pi^2} \sup_{\pi^1 \in \Pi^1} V(x, \pi^1, \pi^2), \tag{2.10}$$

respectively, for each initial state $x \in \mathbf{X}$.

If $U(\cdot) = L(\cdot)$ holds, then the common function is called *the game value*, and is denoted by $V^*(\cdot)$.

Definition 2.3.2. *Suppose that the game has a value V^*. A strategy $\pi^1_* \in \Pi^1$ is said to be* optimal for player 1 *if*

$$V^*(x) = \inf_{\pi^2 \in \Pi^2} V(x, \pi^1_*, \pi^2), \quad x \in \mathbf{X}.$$

Similarly, a strategy $\pi^2_* \in \Pi^2$ is said to be optimal for the player 2 *if*

$$V^*(x) = \sup_{\pi^1 \in \Pi^1} V(x, \pi^1, \pi^2_*), \quad x \in \mathbf{X}.$$

Hence, the pair (π^1_*, π^2_*) is called an optimal pair *of strategies.*

Observe that $(\pi^1_*, \pi^2_*) \in \Pi^1 \times \Pi^2$ is an optimal pair if and only if

$$V(x, \pi^1, \pi^2_*) \leq V(x, \pi^1_*, \pi^2_*) \leq V(x, \pi^1_*, \pi^2), \quad \forall\ (\pi^1, \pi^2) \in \Pi^1 \times \Pi^2,\ x \in \mathbf{X}. \tag{2.11}$$

2.4 Assumptions and preliminary results

For probability measures $\varphi^1(\cdot|x) \in \mathbb{A}(x)$ and $\varphi^2(\cdot|x) \in \mathbb{B}(x)$, $x \in \mathbf{X}$, we write $\varphi^i(x) = \varphi^i(\cdot|x)$, $i = 1, 2$. In addition, for a measurable function $u : \mathbb{K} \to \mathbb{R}$,

$$u(x, \varphi^1, \varphi^2) = u(x, \varphi^1(x), \varphi^2(x)) := \int_{B(x)} \int_{A(x)} u(x, a, b) \varphi^1(da|x) \varphi^2(db|x). \tag{2.12}$$

In this way, for the functions r and Q in the game model (2.1), and for each $x \in \mathbf{X}$, we have

$$r(x, \varphi^1, \varphi^2) := \int_{A(x)} \int_{B(x)} r(x, a, b) \varphi^1(da|x) \varphi^2(db|x),$$

and

$$Q(X|x,\varphi^1,\varphi^2) := \int_{A(x)}\int_{B(x)} Q(X|x,a,b)\varphi^1(da|x)\varphi^2(db|x), \quad X \in \mathcal{B}(\mathbf{X}).$$

The existence of the game value as well as a pair of optimal strategies is analyzed under the following conditions.

Assumption 2.4.1. *The game model* (2.1) *satisfies the following:*

(a) For each $x \in \mathbf{X}$, the sets $A(x)$ and $B(x)$ are compact.

(b) For each $(x,a,b) \in \mathbb{K}$, $r(x,\cdot,b)$ is upper semicontinuous on $A(x)$, and $r(x,a,\cdot)$ is lower semicontinuous on $B(x)$. Moreover, there exists a constant $r_0 > 0$ and a function $w : \mathbf{X} \to [1,\infty)$ such that

$$|r(x,a,b)| \le r_0 w(x), \tag{2.13}$$

and the functions

$$\int_{\mathbf{X}} w(y)Q(dy|x,\cdot,b) \quad \text{and} \quad \int_{\mathbf{X}} w(y)Q(dy|x,a,\cdot) \tag{2.14}$$

are continuous on $A(x)$ and $B(x)$, respectively.

(c) For each $(x,a,b) \in \mathbb{K}$ and any bounded measurable function u on $\mathbf{X}$, the functions

$$\int_{\mathbf{X}} u(y)Q(dy|x,\cdot,b) \quad \text{and} \quad \int_{\mathbf{X}} u(y)Q(dy|x,a,\cdot)$$

are continuous on $A(x)$ and $B(x)$, respectively.

(d) The function α is continuous on $\mathbb{K}$, and

$$\alpha^* := \sup_{(x,a,b)\in\mathbb{K}} \alpha(x,a,b) < 1. \tag{2.15}$$

(e) There exists a positive constant β such that $1 \le \beta < (\alpha^*)^{-1}$ and, for all $(x,a,b) \in \mathbb{K}$,

$$\int_{\mathbf{X}} w(y)Q(dy|x,a,b) \le \beta w(x). \tag{2.16}$$

Definition 2.4.2. *For each measurable function $u : \mathbf{X} \to \mathbb{R}$, we define the w-norm as*

$$\|u\|_w := \sup_{x\in\mathbf{X}} \frac{|u(x)|}{w(x)},$$

and let $\mathbb{B}_w$ be the Banach space of all real-valued measurable functions defined on $\mathbf{X}$ with finite w-norm.

We define the Shapley operator T as

$$Tu(x) := \inf_{\varphi^2\in\mathbb{B}(x)} \sup_{\varphi^1\in\mathbb{A}(x)} H(u;x,\varphi^1,\varphi^2), \quad x \in \mathbf{X}, \tag{2.17}$$

where, for each $u \in \mathbb{B}_w(\mathbf{X})$ and $(x, a, b) \in \mathbb{K}$, H is defined as

$$H(u; x, a, b) := r(x, a, b) + \alpha(x, a, b) \int_{\mathbf{X}} u(y) Q(dy|x, a, b). \tag{2.18}$$

The Shapley operator maps $\mathbb{B}_w$ into itself. Moreover, as we will see later, the interchange of inf and sup in (2.17) holds.

Lemma 2.4.3. *Suppose that Assumptions 2.4.1(b) and 2.4.1 (c) hold. Then for each $u \in \mathbb{B}_w(\mathbf{X})$, the function $u'(x, a, b) := \int u(y) Q(dy|x, a, b)$ is continuous in $a \in A(x)$ and $b \in B(x)$ for every $x \in \mathbf{X}$.*

The proof of Lemma 2.4.3 is similar to the proof of Lemma 1.5.3

Lemma 2.4.4. *Under Assumption 2.4.1, for each $u \in \mathbb{B}_w(\mathbf{X})$:*

(a) $Tu(x) = \max_{\varphi^1 \in \mathbb{A}(x)} \min_{\varphi^2 \in \mathbb{B}(x)} H(u; x, \varphi^1(x), \varphi^2(x))$, $x \in \mathbf{X}$;

(b) There exist $\varphi^1_*(x) \in \mathbb{A}(x)$ and $\varphi^2_* \in \mathbb{B}(x)$ such that,

$$\begin{aligned} Tu(x) &= H(u; x, \varphi^1_*(x), \varphi^2_*(x)) \\ &= \max_{\varphi^1 \in \mathbb{A}(x)} H(u; x, \varphi^1, \varphi^2_*) \\ &= \min_{\varphi^2 \in \mathbb{B}(x)} H(u; x, \varphi^1_*, \varphi^2), \quad x \in \mathbf{X}; \end{aligned}$$

(c) Tu is an element of $\mathbb{B}_w(\mathbf{X})$.

Proof.

(a) Since Assumption 2.4.1 holds, $A(x)$ and $B(x)$ are both compact, and by Proposition B.2.5 (see Appendix B.2), $\mathbb{A}(x)$ and $\mathbb{B}(x)$ are also compact for every $x \in \mathbf{X}$, which implies that the maximum and minimum are attained on $\mathbb{A}(x)$ and $\mathbb{B}(x)$, respectively. That is, we can use max and min instead of sup and inf in (2.17). On the other hand, by Lemma 2.4.3 and Assumption 2.4.1(d), for $u \in \mathbb{B}_w(\mathbf{X})$ and $(x, a, b) \in \mathbb{K}$, we have that $H(u; x, \cdot, b)$ is u.s.c. on $A(x)$ and $H(u; x, a, \cdot)$ is l.s.c. on $B(x)$. Furthermore, since $A(x)$ and $B(x)$ are both compact, by Proposition A.1.4, $H(u; x, \cdot, b)$ is bounded above on $A(x)$ and $H(u; x, a, \cdot)$ is bounded below on $B(x)$. Then, applying Proposition B.2.6 we can prove that the function $H(u; x, \cdot, \varphi^2)$ is u.s.c. on $\mathbb{A}(x)$ while $H(u; x, \varphi^1, \cdot)$ is l.s.c. on $\mathbb{B}(x)$. In addition, since $H(u; x, \varphi^1, \varphi^2)$ is concave in φ^1 and convex in φ^2, the Fan's Minimax Theorem (see Theorem A.4.2 in Appendix A.4) implies that we can interchange min and max in (2.17), i.e.,

$$Tu(x) = \max_{\varphi^1 \in \mathbb{A}(x)} \min_{\varphi^2 \in \mathbb{B}(x)} H(u; x, \varphi^1(x), \varphi^2(x)), \quad x \in \mathbf{X}.$$

(b) Theorem B.3.2 (see Appendix B.3) yields the existence of $\varphi^1_* \in \mathbb{A}(x)$ and $\varphi^2_* \in \mathbb{B}(x)$ such that

$$\begin{aligned} Tu(x) &= H(u; x, \varphi^1_*(x), \varphi^2_*(x)) \\ &= \max_{\varphi^1 \in \mathbb{A}(x)} H(u; x, \varphi^1, \varphi^2_*) \\ &= \min_{\varphi^2 \in \mathbb{B}(x)} H(u; x, \varphi^1_*, \varphi^2), \quad x \in \mathbf{X}, \end{aligned}$$

which proves part (b).

(c) Since $|u(\cdot)| \leq \|u\|_w w(\cdot)$, from (2.13) and (2.16) we obtain, for arbitrary $(x, a, b) \in \mathbb{K}$, that

$$\begin{aligned} |H(u; x, a, b)| &\leq r_0 w(y) + \|u\|_w \int_{\mathbf{X}} w(y) Q(dy|x, a, b) \\ &\leq (r_0 + \beta \|u\|_w) w(y), \end{aligned}$$

which implies that $Tu \in \mathbb{B}_w(\mathbf{X})$.

$\square$

Lemma 2.4.5. *Under Assumption 2.4.1, the operator T defined in* (2.17) *is a contraction operator in $\mathbb{B}_w(\mathbf{X})$ with modulus $\alpha^*\beta < 1$.*

Proof. Let us first notice that T is a monotone operator, that is, for $u, \tilde{u} \in \mathbb{B}_w(\mathbf{X})$ such that $u \geq \tilde{u}$, we have

$$Tu(x) \geq T\tilde{u}(x), \quad \forall\, x \in \mathbf{X}.$$

On the other hand, since Assumption 2.4.1(e) holds and $\alpha(x, a, b) \leq \alpha^*$ $\forall (x, a, b) \in \mathbb{K}$, we have for any real number $k \geq 0$

$$T(u + kw)(x) \leq Tu(x) + \alpha^* \beta k w(x), \tag{2.19}$$

for all $x \in \mathbf{X}$ and $u \in \mathbb{B}_w(\mathbf{X})$.

Let $u, v \in \mathbb{B}_w(x)$ be arbitrary. Since $u \leq v + w\|u - v\|_w$, by the monotonicity of T and (2.19), with $k = \|u - v\|_w$, we have

$$Tu(x) \leq T(v + kw)(x) \leq Tv(x) + \alpha^* \beta k w(x),$$

that is,

$$Tu(x) - Tv(x) \leq \alpha^* \beta \|u - v\|_w w(x), \quad x \in \mathbf{X}. \tag{2.20}$$

Interchanging u and v, we obtain

$$Tu(x) - Tv(x) \geq -\alpha^* \beta \|u - v\|_w w(x), \quad x \in \mathbf{X}. \tag{2.21}$$

Thus, from (2.20) and (2.21) we obtain

$$|Tu(x) - Tv(x)| \leq \alpha^* \beta \|u - v\|_w w(x), \quad x \in \mathbf{X}.$$

Therefore,

$$\|Tu - Tv\|_w \leq \alpha^* \beta \|u - v\|_w,$$

that is, T is a contraction operator with modulus $\alpha^* \beta < 1$. $\square$

Lemma 2.4.6. *Let w and β be as in Assumption 2.4.1, and let $\pi^1 \in \Pi^1$ and $\pi^2 \in \Pi^2$ be arbitrary strategies for players 1 and 2, respectively, and $x \in \mathbf{X}$ the initial state. Then for each $u \in \mathbb{B}_w(\mathbf{X})$ and $n \in \mathbb{N}_0$,*

(a) $E_x^{\pi^1, \pi^2}[w(x_{n+1})] \leq \beta^{n+1} w(x)$;

(b) $\lim_{n \to \infty} E_x^{\pi^1, \pi^2} \Gamma_n u(x_n) = 0.$

Proof.

(a) From Assumption 2.4.1(e), for each $x \in \mathbf{X}$, $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$, and $n \in \mathbb{N}_0$,

$$E_x^{\pi^1,\pi^2}[w(x_{n+1})] \leq \beta E_x^{\pi^1,\pi^2}[w(x_n)].$$

The iteration of this inequality yields

$$E_x^{\pi^1,\pi^2}[w(x_{n+1})] \leq \beta^{n+1} w(x), \quad x \in \mathbf{X}, \; n \in \mathbb{N}_0. \tag{2.22}$$

(b) From (2.22) and (2.3), for each $u \in \mathbb{B}_w(\mathbf{X})$, $x \in \mathbf{X}$, $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$, and $n \in \mathbb{N}_0$,

$$\begin{aligned}\left| E_x^{\pi^1,\pi^2} \Gamma_n u(x_n) \right| &\leq (\alpha^*)^n \|u\|_w E_x^{\pi^1,\pi^2}[w(x_n)] \\ &\leq (\beta\alpha^*)^n \|u\|_w w(x).\end{aligned}$$

Therefore,

$$\lim_{n\to\infty} E_x^{\pi^1,\pi^2} \Gamma_n u(x_n) = 0, \quad x \in \mathbf{X}, \; (\pi^1, \pi^2) \in \Pi^1 \times \Pi^2. \tag{2.23}$$

□

Definition 2.4.7. *Let $\varphi^1 \in \Phi^1$, $\varphi^2 \in \Phi^2$, and H be as in (2.18). We define the operator $T_{\varphi^1\varphi^2} : \mathbb{B}_w(\mathbf{X}) \to \mathbb{B}_w(\mathbf{X})$ as*

$$T_{\varphi^1\varphi^2} u(x) := H(u; x, \varphi^1(x), \varphi^2(x)), \quad x \in \mathbf{X}. \tag{2.24}$$

Lemma 2.4.8. *The operator $T_{\varphi^1\varphi^2}$ is a contraction operator with modulus $\alpha^*\beta$ on $\mathbb{B}_w(\mathbf{X})$, and $V(x, \varphi^1, \varphi^2)$ is its unique fixed point in $\mathbb{B}_w(\mathbf{X})$.*

Proof. Using similar arguments as those in the proof of Lemma 2.4.5, it follows that $T_{\varphi^1\varphi^2}$ is a contraction operator on $\mathbb{B}_w(\mathbf{X})$ with modulus $\tau := \beta\alpha^*$. By Banach's fixed point Theorem A.3.2 (see Appendix A.3), $T_{\varphi^1\varphi^2}$ has a unique fixed point $u_{\varphi^1\varphi^2}$ in $\mathbb{B}_w(\mathbf{X})$, that is,

$$u_{\varphi^1\varphi^2} = T_{\varphi^1\varphi^2} u_{\varphi^1\varphi^2}, \tag{2.25}$$

from which we have that $u_{\varphi^1\varphi^2}$ is the only solution in $\mathbb{B}_w(\mathbf{X})$ to the equation

$$u_{\varphi^1,\varphi^2}(x) = r(x, \varphi^1(x), \varphi^2(x)) + \alpha(x, \varphi^1(x), \varphi^2(x)) \int_{\mathbf{X}} u_{\varphi^1,\varphi^2}(y) Q(dy|x, \varphi^1(x), \varphi^1(x)) \quad x \in \mathbf{X}. \tag{2.26}$$

Thus,

$$u_{\varphi^1,\varphi^2}(x) = \int_{\mathbf{B}} \int_{\mathbf{A}} \left[r(x,a,b) + \alpha(x,a,b) \int_{\mathbf{X}} v_{\varphi^1\varphi^2}(y) Q(dy|x,a,b) \right] \varphi^1(da|x)\varphi^2(db|x), \;\; x \in \mathbf{X}.$$

Iterating this equation we obtain

$$u_{\varphi^1,\varphi^2}(x) = E_x^{\pi^1,\pi^2} \sum_{n=0}^{m-1} \Gamma_n r(x_n, a_n, b_n) + E_x^{\pi^1,\pi^2} \Gamma_m v_{\varphi^1,\varphi^2}(x_m).$$

Now, letting $m \to \infty$, from (2.23) and (2.8) we obtain

$$u_{\varphi^1,\varphi^2}(x) = V(x, \varphi^1, \varphi^2), \quad \forall\, x \in \mathbf{X}.$$

Therefore, $V(x, \varphi^1, \varphi^2)$ is the unique fixed point in $\mathbb{B}_w(\mathbf{X})$ of the operator T_{φ^1,φ^2}.

□

2.5 Existence of optimal strategies

We now present our main result as follows.

Theorem 2.5.1. *Suppose that Assumption 2.4.1 holds. Then:*

(a) The game $\mathcal{GM}$ has a value $V^* \in \mathbb{B}_w$.

(b) The game value V^* is the unique function in $\mathbb{B}_w$ such that $TV^* = V^*$.

(c) There exist $\varphi_*^1(x) \in \mathbb{A}(x)$ and $\varphi_*^2(x) \in \mathbb{B}(x)$ such that

$$\begin{aligned} V^*(x) &= H(V^*; x, \varphi_*^1, \varphi_*^2) && (2.27)\\ &= \max_{\varphi^1 \in \mathbb{A}(x)} H(V^*; x, \varphi^1, \varphi_*^2) && (2.28)\\ &= \min_{\varphi^2 \in \mathbb{B}(x)} H(V^*; x, \varphi_*^1, \varphi^2), \quad \forall\, x \in \mathbf{X}. && (2.29) \end{aligned}$$

In addition, the stationary strategies $\varphi_*^1 = \{\varphi_*^1\} \in \Pi_S^1$ and $\varphi_*^2 = \{\varphi_*^2\} \in \Pi_S^2$ form an optimal pair of strategies for the game $\mathcal{GM}$.

Proof. From Lemma 2.4.5, T is a contraction on $\mathbb{B}_w(\mathbf{X})$, thus, by Banach's Fixed Point Theorem (see Appendix A.3), there exists a unique fixed point $v \in \mathbb{B}_w(\mathbf{X})$ of T. From (2.17) and Lemma 2.4.4(a), v satisfies

$$\begin{aligned} v(x) = Tv(x) &= \max_{\varphi^1 \in \mathbb{A}(x)} \min_{\varphi^2 \in \mathbb{B}(x)} H(v; x, \varphi^1(x), \varphi^2(x))\\ &= \min_{\varphi^2 \in \mathbb{B}(x)} \max_{\varphi^1 \in \mathbb{A}(x)} H(v; x, \varphi^1(x), \varphi^2(x)), \quad x \in \mathbf{X}. \end{aligned}$$

In addition, from Lemma 2.4.4 (b), there exists a pair of stationary strategies $(\varphi_*^1, \varphi_*^2) \in \Pi_S^1 \times \Pi_S^2$ such that

$$\begin{aligned} v(x) &= H(v; x, \varphi_*^1(x), \varphi_*^2(x)) = T_{\varphi_*^1 \varphi_*^2} v(x) && (2.30)\\ &= \max_{\varphi^1 \in \mathbb{A}(x)} H(v; x, \varphi^1(x), \varphi_*^2(x)) && (2.31)\\ &= \min_{\varphi^2 \in \mathbb{B}(x)} H(v; x, \varphi_*^1(x), \varphi^2(x)), \quad x \in \mathbf{X}. && (2.32) \end{aligned}$$

Moreover, from Lemma 2.4.8, $V(\cdot, \varphi_*^1, \varphi_*^2)$ is the unique fixed point of $T_{\varphi_*^1 \varphi_*^2}$, so (2.30) implies $v(x) = V(x, \varphi_*^1, \varphi_*^2)$, $x \in \mathbf{X}$. Therefore, considering (2.31) and (2.32), the theorem will be proved if we show that

$$V(x, \pi^1, \varphi_*^2) \le V(x, \varphi_*^1, \varphi_*^2) \le V(x, \varphi_*^1, \pi^2), \quad \forall (\pi^1, \pi^2) \in \Pi^1 \times \Pi^2,\ x \in \mathbf{X}. \tag{2.33}$$

To prove the first inequality in (2.33), let $\pi^1 \in \Pi^1$ be an arbitrary strategy for player 1. Then, for all $n \in \mathbb{N}$,

$$\begin{aligned}
E_x^{\pi^1,\varphi_*^2}[\Gamma_{n+1}V(x_{n+1},\varphi_*^1,\varphi_*^2)|h_n] &= \Gamma_n E_x^{\pi^1,\varphi_*^2}[\alpha(x_n,a_n,b_n)V(x_{n+1},\varphi_*^1,\varphi_*^2)|h_n] \\
&= \Gamma_n\left\{\alpha(x_n,\pi_n^1(h_n),\varphi_*^2)\int_{\mathbf{X}} V(y,\varphi_*^1,\varphi_*^2)Q(dy|x_n,\pi_n^1(h_n),\varphi_*^2)\right\} \\
&= \Gamma_n\left\{\alpha(x_n,\pi_n^1(h_n),\varphi_*^2)\int_{\mathbf{X}} V(y,\varphi_*^1,\varphi_*^2)Q(dy|x_n,\pi_n^1(h_n),\varphi_*^2)\right. \\
&\quad \left. + r(x_n,\pi_n^1(h_n),\varphi_*^2) - r(x_n,\pi_n^1(h_n),\varphi_*^2)\right\} \\
&\leq \Gamma_n\left\{\sup_{\varphi^1\in\mathbb{A}(x_n)} H(v;x_n,\varphi^1(x_n),\varphi_*^2(x_n)) - r(x_n,\pi_n^1(h_n),\varphi_*^2)\right\} \\
&= \Gamma_n\{v(x_n) - r(x_n,\pi_n^1(h_n),\varphi_*^2)\} \\
&= \Gamma_n\{V(x_n,\varphi_*^1,\varphi_*^2) - r(x_n,\pi_n^1(h_n),\varphi_*^2)\}
\end{aligned} \tag{2.34}$$

where the last two equalities come from (2.30) and (2.31). Now, from (2.34), for all $n \in \mathbb{N}$,

$$\Gamma_n V(x_n,\varphi_*^1,\varphi_*^2) - E_x^{\pi^1,\varphi_*^2}[\Gamma_{n+1}V(x_{n+1},\varphi_*^1,\varphi_*^2)|h_n] \geq \Gamma_n r(x_n,\pi_n^1(h_n),\varphi_*^2),$$

thus, taking expectation $E_x^{\pi^1,\varphi_*^2}$ and adding over $n = 0,1,\ldots,m-1, m > 0$, we obtain

$$V(x,\varphi_*^1,\varphi_*^2) - E_x^{\pi^1,\varphi_*^2}[\Gamma_{m+1}V(x_{m+1},\varphi_*^1,\varphi_*^2)] \geq E_x^{\pi^1,\varphi_*^2}\sum_{n=0}^{m-1}\Gamma_n r(x_n,a_n,b_n).$$

Letting $m \to \infty$, from (2.8), Theorem A.2.3 and Lemma 2.4.6(b) we obtain

$$V(x,\varphi_*^1,\varphi_*^2) \geq V(x,\pi^1,\varphi_*^2), \quad x \in \mathbf{X},$$

that is, the first inequality in (2.33) holds. The second inequality is proved similarly. Hence, the proof of Theorem 2.5.1 is completed. □

2.6 Example: A game with random horizon

Let $(\mathbf{X},\mathbf{A},\mathbf{B},\mathbb{K}_A,\mathbb{K}_B,Q,r)$ be a standard game model which is played as follows. At time n, when the game is in state x_n, players choose actions (a_n,b_n) and player 1 receives a payoff $r(x_n,a_n,b_n)$ from player 2. There is a positive probability $1-\alpha(x_n,a_n,b_n)$ that the game stops; otherwise, the game process jumps to a new state x_{n+1} according to a stochastic kernel $Q(\cdot|x_n,a_n,b_n)$, and the process is repeated. Hence, the game has a random horizon.

Our objective is to show that this class of games can be modeled by means of models of the form (2.1), which includes to prove that the corresponding random horizon total payoff can be written as a performance index with state-actions-dependent discount factors as (2.8).

We impose the following condition on the stopping probability.

Assumption 2.6.1. *There is* $\gamma \in (0,1)$ *such that*

$$\alpha^* := \sup_{(x,a,b)\in\mathbb{K}} \alpha(x,a,b) \leq 1-\gamma < 1.$$

Let x^* and (a^*, b^*) be artificial state and actions. We define the game model

$$\mathcal{GM}^* = (\mathbf{X}^*, \mathbf{A}^*, \mathbf{B}^*, \mathbb{K}_{\mathbf{A}^*}, \mathbb{K}_{\mathbf{B}^*}, Q^*, \alpha, r^*)$$

where $\mathbf{X}^* = \mathbf{X} \cup \{x^*\}$, $\mathbf{A}^* = \mathbf{A} \cup \{a^*\}$, $\mathbf{B}^* = \mathbf{B} \cup \{b^*\}$ and the corresponding x-sections are the sets

$$A^*(x) := \begin{cases} \{a^*\} & \text{if } x = x^*, \\ A(x) & \text{if } x \in \mathbf{X}; \end{cases}$$

$$B^*(x) := \begin{cases} \{b^*\} & \text{if } x = x^*, \\ B(x) & \text{if } x \in \mathbf{X}. \end{cases}$$

We denote

$$\mathbb{K}^* := \{(x,a,b) : x \in \mathbf{X}^*, a \in \mathbf{A}^*, b \in \mathbf{B}^*\}.$$

The transition law Q^* among the states in $\mathbf{X}^*$ is a stochastic kernel on $\mathbf{X}^*$ given $\mathbb{K}^*$ defined as follows. For $(x,a,b) \in \mathbb{K}$

$$\begin{aligned} Q^*(X|x,a,b) &:= \alpha(x,a,b)Q(X|x,a,b), \quad X \in \mathcal{B}(\mathbf{X}) \\ Q^*(\{x^*\}|x,a,b) &:= 1 - \alpha(x,a,b), \\ Q^*(\{x^*\}|x^*,a^*,b^*) &:= 1. \end{aligned}$$

The payoff function $r^* : \mathbb{K}^* \to \mathbb{R}$ is given by

$$r^*(x) := \begin{cases} r(x,a,b) & \text{if } (x,a,b) \in \mathbb{K}, \\ 0 & \text{if } (x,a,b) = (x^*,a^*,b^*). \end{cases}$$

Let $(\Omega', \mathcal{F}')$ be the measurable space associated to the game model $\mathcal{GM}^*$ (see Section 2.2.2) and define the first passage time $\tau : \Omega' \to \mathbb{N}_0 \cup \{+\infty\}$ as

$$\tau := \tau(x_0, a_0, b_0, \ldots) := \inf\{n \in \mathbb{N}_0 : x_n = x^*\},$$

where, as usual, $\inf \emptyset = +\infty$. In other words, τ is the first entrance time of the game process into the set $\{x^*\}$, which will never be left once it is reached, and where the players incur no payoff.

In next chapter it is proved that it is sufficient to consider the family of Markov strategies, in order to analyze the game. Then, we restrict our analysis to the sets Π^i_M, $i = 1, 2$.

For each pair of strategies $(\varphi^1, \varphi^2) \in \Pi^1_M \times \Pi^2_M$ and initial state $x \in \mathbf{X}$, the total expected payoff with random horizon τ takes the form

$$V_\tau(x, \varphi^1, \varphi^2) := E_x^{*,\varphi^1,\varphi^2} \sum_{n=0}^{\tau} r(x_n, a_n, b_n),$$

where E^* is the expectation operator corresponding to the game model $\mathcal{GM}^*$.

Theorem 2.6.2. *For each $x \in \mathbf{X}$ and $(\varphi^1, \varphi^2) \in \Pi_M^1 \times \Pi_M^2$*

$$V_\tau(x, \varphi^1, \varphi^2) = V(x, \varphi^1, \varphi^2),$$

where V is the performance index defined in (2.8)

Proof. For each $x \in \mathbf{X}$ and $(\varphi^1, \varphi^2) \in \Pi_M^1 \times \Pi_M^2$, observe that

$$E_x^{*,\varphi^1,\varphi^2} \sum_{n=0}^{\tau} r^*(x_n, a_n, b_n) = E_x^{*,\varphi^1,\varphi^2} \sum_{n=0}^{\infty} r(x_n, a_n, b_n) I_{[\tau > n]}.$$

Now we have

$$E_x^{*,\varphi^1,\varphi^2} r(x_0, a_0, b_0) = E_x^{\varphi^1,\varphi^2} r(x_0, a_0, b_0).$$

In addition, on the set $[\tau > n]$, we have

$$\begin{aligned}
E_x^{*,\varphi^1,\varphi^2} r^*(x_1, a_1, b_1) &= \int_{\mathbf{A}^*}\int_{\mathbf{B}^*}\int_{\mathbf{X}^*}\int_{\mathbf{A}^*}\int_{\mathbf{B}^*} r^*(x_1, a_1, b_1)\ \varphi_1^2(db_1|x_1)\varphi_1^1(da_1|x_1)Q^*(dx_1|x, a_0, b_0) \\
&\qquad \varphi_0^2(db_0|x)\varphi_0^1(da_0|x) \\
&= \int_{\mathbf{A}}\int_{\mathbf{B}}\int_{\mathbf{X}}\int_{\mathbf{A}}\int_{\mathbf{B}} r(x_1, a_1, b_1)\ \varphi_1^2(db_1|x_1)\varphi_1^1(da_1|x_1)\alpha(x, a_0, b_0)Q(dx_1|x, a_0, b_0) \\
&\qquad \varphi_0^2(db_0|x)\varphi_0^1(da_0|x) \\
&= \int_{\mathbf{A}}\int_{\mathbf{B}}\int_{\mathbf{X}}\int_{\mathbf{A}}\int_{\mathbf{B}} \alpha(x, a_0, b_0) r(x_1, a_1, b_1)\ \varphi_1^2(db_1|x_1)\varphi_1^1(da_1|x_1)Q(dx_1|x, a_0, b_0) \\
&\qquad \varphi_0^2(db_0|x)\varphi_0^1(da_0|x) \\
&= E_x^{\varphi^1,\varphi^2} \alpha(x_0, a_0, b_0) r(x_1, a_1, b_1).
\end{aligned}$$

By applying similar arguments we can prove

$$E_x^{*,\varphi^1,\varphi^2} [r^*(x_n, a_n, b_n)] = E_x^{\varphi^1,\varphi^2} \left[\prod_{n=0}^{n-1} \alpha(x_n, a_n, b_n) r(x_n, a_n, b_n) \right].$$

This fact proves the result. □

Similar game models with random horizon have been studied in previous works under several settings. For instance in [16] it is considered a finite stochastic game, and in [14] it is assumed a bounded payoff. Our example is based on the paper [6].

Chapter 3

Markov games with random state-actions-dependent discount factors

3.1 Introduction

In this chapter we study a class of games which can be considered as a mixture of the games analyzed in previous chapters. Indeed, we now consider games with random state-actions-dependent discount factors of the form $\alpha(x_n, a_n, b_n, \xi_{n+1})$, where x_n, a_n, b_n and ξ_{n+1} are the state, the actions chosen by the players and the discount factor's random disturbance at time n. The key point to analyze this kind of games is to prove their equivalence to games with state-actions-dependent discount factors, i.e., non-randomized. Once stated such an equivalence we can use Theorem 2.5.1 above to prove the existence of a value of the game and a pair of optimal strategies. However this approach holds over the class of Markov strategies for which we first need to prove its sufficiency in the sense of Theorem 3.5.5 below.

As in the previous chapters, we rewrite all elements needed to formulate the game problem, adapted to this kind of games.

3.2 Game model

A zero-sum Markov game model with random state-actions-dependent discount factors is defined by the following collection

$$\tilde{\mathcal{GM}} := (\mathbf{X}, \mathbf{A}, \mathbf{B}, \mathbb{K}_{\mathbf{A}}, \mathbb{K}_{\mathbf{B}}, \mathbf{S}, Q, \tilde{\alpha}, r) \tag{3.1}$$

satisfying the following conditions. The state space $\mathbf{X}$ and the actions sets $\mathbf{A}$ and $\mathbf{B}$ for players 1 and 2, respectively, as well as the discount factors disturbance spaces $\mathbf{S}$, are assumed to be Borel spaces. The constraint sets $\mathbb{K}_{\mathbf{A}}$ and $\mathbb{K}_{\mathbf{B}}$ are Borel subsets of $\mathbf{X} \times \mathbf{A}$ and $\mathbf{X} \times \mathbf{B}$, respectively. For each $x \in \mathbf{X}$, the x-sections

$$A(x) := \{a \in \mathbf{A} \mid (x, a) \in \mathbb{K}_{\mathbf{A}}\}$$

and

$$B(x) := \{b \in \mathbf{B} \mid (x,b) \in \mathbb{K}_{\mathbf{B}}\}$$

are non-empty Borel subsets, and represent the admissible actions sets for players 1 and 2, respectively, when the system is in the state x.

The set

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

of admissible state-actions triplets is a Borel subset of $\mathbf{X}\times\mathbf{A}\times\mathbf{B}$. The transition law $Q(\cdot|x,a,b)$ is a stochastic kernel on $\mathbf{X}$ given $\mathbb{K}$, and $\tilde{\alpha} : \mathbb{K}\times\mathbf{S} \to (0,1)$ is a measurable function which gives the discount factor $\tilde{\alpha}(x_n, a_n, b_n, \xi_{n+1})$ at stage $n \in \mathbb{N}$, where $\{\xi_n\}$ is a sequence of independent and identically distributed (i.i.d.) random variables defined on a probability space taking values in $\mathbf{S}$ with common distribution $\theta \in \mathbb{P}(S)$. That is

$$\theta(S) = P(\xi_n \in S), \quad S \in \mathcal{B}(\mathbf{S}),\ n \in \mathbb{N}.$$

Finally, $r(\cdot,\cdot,\cdot)$ is a real-valued measurable function on $\mathbb{K}$ that represents the one-stage payoff function.

Interpretation.

The game is played as follows. At the initial state $x_0 \in \mathbf{X}$, the players independently choose actions $a_0 \in A(x_0)$ and $b_0 \in B(x_0)$. Then player 1 receives a payoff $r(x_0,a_0,b_0)$ from player 2, and the game moves to a new state x_1 according to the transition law $Q(\cdot|x_0,a_0,b_0)$, and the random disturbance ξ_1 appears. Once the system is in the state x_1 the players select actions $a_1 \in A(x_1)$ and $b_1 \in B(x_1)$ and player 1 receives a payoff $r(x_1,a_1,b_1)$ from player 2. Next the system moves to a new state x_2, the random disturbance ξ_2 appears, and the process is repeated over and over again. In general, at stage $n \in \mathbb{N}$, on the record of the state-actions and random disturbances, player 1 receives $r(x_n,a_n,b_n)$ from player 2 and the corresponding discounted payoff takes the form

$$\tilde{\Gamma}_n r(x_n,a_n,b_n) \tag{3.2}$$

where

$$\tilde{\Gamma}_n := \prod_{k=0}^{n-1} \tilde{\alpha}(x_k, a_x, b_k, \xi_{k+1}) \text{ if } n \in \mathbb{N}, \text{ and } \tilde{\Gamma}_0 = 1. \tag{3.3}$$

Thus, the goal of player 1 is to maximize, while player 2 wants to minimize, the total expected discounted payoff defined by the accumulation of the one-stage payoffs (3.2) over an infinite horizon.

3.2.1 Strategies

The players' strategies are defined exactly in the same way as in Chapter 2 but changing the definition of the set $\mathbb{H}_n$ of histories up to time n. In this case, we define $\mathbb{H}_0 := \mathbf{X}$ and $\mathbb{H}_n := \mathbb{K}\times\mathbf{S}\times\mathbb{H}_{n-1}$ for $n \in \mathbb{N}$, where, for each $n \in \mathbb{N}_0$, an element $h_n \in \mathbb{H}_n$ takes the form

$$h_n = (x_0, a_0, b_0, \xi_1, \ldots, x_{n-1}, a_{n-1}, b_{n-1}, \xi_n, x_n),$$

which represents the history of the game up to time n.

We keep the notation and definitions introduced in Section 2.2.1 of Chapter 2, for the sets $\mathbb{A}(x)$, $\mathbb{B}(x)$, Φ^1 and Φ^2.

Thus, strategies for player 1 are defined in Definition 2.2.1, and Π^1 represents the family of all strategies for player 1. In a likewise manner, Markov and stationary strategies are defined in Definition 2.2.2, and Π^1_M and Π^1_S represent the set of all Markov strategies and the set of all stationary strategies, respectively.

The sets of strategies Π^2, Markov strategies Π^2_M and stationary strategies Π^2_S for player 2 are defined similarly.

3.2.2 The game process

Let $(\Omega', \mathcal{F}')$ be the measurable space consisting of the sample space $\Omega' = (\mathbb{K} \times \mathbf{S})^\infty$ and its product σ-algebra $\mathcal{F}'$. As in the previous chapters, for each pair of strategies $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$ and initial state $x_0 = x \in \mathbf{X}$, there exists a unique probability measure $P_x^{\pi^1,\pi^2}$ and a stochastic process $\{(x_n, a_n, b_n, \xi_{n+1})\}$, where x_n, a_n, b_n and ξ_{n+1} represent the state, the actions of players, and the discount factor random disturbance, respectively, at stage $n \in \mathbb{N}_0$, satisfying

$$
\begin{aligned}
P_x^{\pi^1,\pi^2}[x_0 \in X] &= \delta_x(X), \quad X \in \mathcal{B}(\mathbf{X}); &(3.4)\\
P_x^{\pi^1,\pi^2}[a_n \in A, b_n \in B|h_n] &= \pi^1_n(A|h_n)\pi^2_n(B|h_n), A \in \mathcal{B}(\mathbf{A}), B \in \mathcal{B}(\mathbf{B}); &(3.5)\\
P_x^{\pi^1,\pi^2}[x_{n+1} \in X|h_n, a_n, b_n, \xi_{n+1}] &= Q(X|x_n, a_n, b_n), \quad X \in \mathcal{B}(\mathbf{X}); &(3.6)\\
P_x^{\pi^1,\pi^2}[\xi_{n+1} \in S|h_n, a_n, b_n] &= \theta(S), \quad S \in \mathcal{B}(\mathbf{S}), &(3.7)
\end{aligned}
$$

where $\delta_x(\cdot)$ is the Dirac measure concentrated at x. We denote by $E_x^{\pi^1,\pi^2}$ the expectation operator with respect to $P_x^{\pi^1,\pi^2}$.

The stochastic process $\{x_n\}$ defined on $(\Omega, \mathcal{F}, P_x^{\pi^1,\pi^2})$ is called *game process*.

3.3 Optimality criterion

According to (3.2) and (3.3), the total expected discounted payoff — with random state-actions-dependent discount factors — for a pair of strategies $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$, given the initial state $x_0 = x \in \mathbf{X}$, is defined as

$$\tilde{V}(x, \pi^1, \pi^2) := E_x^{\pi^1,\pi^2}\left[\sum_{n=0}^{\infty} \tilde{\Gamma}_n r(x_n, a_n, b_n)\right]. \tag{3.8}$$

where $\tilde{\Gamma}_n$ is given by (3.3).

3.3.1 Game value

Definition 3.3.1. *The lower and the upper value of the game are defined as:*

$$L(x) := \sup_{\pi^1 \in \Pi^1} \inf_{\pi^2 \in \Pi^2} \tilde{V}(x, \pi^1, \pi^2) \quad \textit{and} \quad U(x) := \inf_{\pi^2 \in \Pi^2}, \sup_{\pi^1 \in \Pi^1} \tilde{V}(x, \pi^1, \pi^2)$$

respectively, for each initial state $x \in \mathbf{X}$.

If $U(\cdot) = L(\cdot)$ holds, then the common function is called the value of the game and is denoted by V^*.

Definition 3.3.2. *Suppose that the game has a value V^*. A strategy $\pi^1_* \in \Pi^1$ is said to be* optimal for player 1 *if*

$$V^*(x) = \inf_{\pi^2 \in \Pi^2} \tilde{V}(x, \pi^1_*, \pi^2), \quad x \in \mathbf{X}.$$

Similarly, a strategy $\pi^2_* \in \Pi^2$ is said to be optimal for player 2 *if*

$$V^*(x) = \sup_{\pi^1 \in \Pi^1} \tilde{V}(x, \pi^1, \pi^2_*), \quad x \in \mathbf{X}.$$

Hence, the pair (π^1_*, π^2_*) is called an optimal pair *of strategies.*

Notice that $(\pi^1_*, \pi^2_*) \in \Pi^1 \times \Pi^2$ is an optimal pair if and only if

$$\tilde{V}(x, \pi^1, \pi^2_*) \leq \tilde{V}(x, \pi^1_*, \pi^2_*) \leq \tilde{V}(x, \pi^1_*, \pi^2), \ \forall (\pi^1, \pi^2) \in \Pi^1 \times \Pi^2, \ x \in \mathbf{X}, \quad (3.9)$$

which is known as the *saddle point property.*

Similar to previous chapters, for probability measures $\varphi^1(\cdot|x) \in \mathbb{A}(x)$ and $\varphi^2(\cdot|x) \in \mathbb{B}(x)$, $x \in \mathbf{X}$, we write $\varphi^i(x) = \varphi^i(\cdot|x)$, $i = 1, 2$. In addition, for a measurable function $u : \mathbb{K} \to \mathbb{R}$,

$$u(x, \varphi^1, \varphi^2) = u(x, \varphi^1(x), \varphi^2(x)) := \int_{B(x)} \int_{A(x)} u(x, a, b) \varphi^1(da|x) \varphi^2(db|x). \quad (3.10)$$

3.4 Preliminary results

In this section we will present some lemmas that we use in the proof of our main results, which we will present in the next section. In order to prove such lemmas, we introduce the following notation.

Let us fix $\varphi^2 \in \Pi^2_S$. Define the stochastic kernel Q_{φ^2} on $\mathbf{X}$ given $\mathbb{K}_A$ as

$$Q_{\varphi^2}(X|x, a) := \int_{\mathbf{B}} Q(X|x, a, b) \varphi^2(db|x), \quad X \in \mathcal{B}(\mathbf{X}), \quad (3.11)$$

likewise, $r_{\varphi^2} : \mathbb{K}_A \to \mathbb{R}$ and $\tilde{\alpha}_{\varphi^2} : \mathbb{K}_A \times \mathbf{S} \to (0, 1)$ are the measurable functions defined as

$$r_{\varphi^2}(x, a) := \int_{\mathbf{B}} r(x, a, b) \varphi^2(db|x), \quad (3.12)$$

$$\tilde{\alpha}_{\varphi^2}(x, a, s) := \int_{\mathbf{B}} \tilde{\alpha}(x, a, b, s) \varphi^2(db|x). \quad (3.13)$$

In addition, let $\pi^1 \in \Pi^1$ be an arbitrary strategy and $x \in \mathbf{X}$, we define the performance index

$$\tilde{V}_{\varphi^2}(x, \pi^1) := E^{\pi^1}_x \left[\sum_{n=0}^{\infty} \tilde{\Gamma}^{\varphi^2}_n r_{\varphi^2}(x_n, a_n) \right], \quad (3.14)$$

where,

$$\tilde{\Gamma}_n^{\varphi^2} = \prod_{k=0}^{n-1} \tilde{\alpha}_{\varphi^2}(x_k, a_k, \xi_{k+1}) \quad \text{for } n \in \mathbb{N}, \text{ and } \tilde{\Gamma}_0^{\varphi^2} = 1,$$

and $E_x^{\pi^1}$ is the expectation operator with respect to the probability measure $P_x^{\pi^1} \equiv P_x^{\pi^1,\varphi^2}$ induced by $(\pi^1, \varphi^2) \in \Pi^1 \times \Pi_S^2$ and $x_0 = x$. Then, from (3.4) – (3.7), $P_x^{\pi^1}$ satisfies the following properties:

$$P_x^{\pi^1}[x_0 \in X] = \delta_x(X), \quad X \in \mathcal{B}(\mathbf{X}); \tag{3.15}$$

$$\begin{aligned} P_x^{\pi^1}[a_n \in A|h_n] &= P_x^{\pi^1}[a_n \in A, b_n \in \mathbf{B}|h_n] \\ &= \pi_n^1(A|h_n)\varphi_n^2(\mathbf{B}|x_n) \\ &= \pi_n^1(A|h_n), \quad A \in \mathcal{B}(\mathbf{A}); \end{aligned} \tag{3.16}$$

$$P_x^{\pi^1}[x_{n+1} \in X|h_n, a_n, b_n, \xi_{n+1}] = Q_{\varphi^2}(X|x_n, a_n), \quad X \in \mathcal{B}(\mathbf{X}); \tag{3.17}$$

$$P_x^{\pi^1}[\xi_{n+1} \in S|h_n, a_n, b_n] = \theta(S), \quad S \in \mathcal{B}(\mathbf{S}). \tag{3.18}$$

Similarly, for a fixed $\varphi^1 \in \Pi_S^1$, we define Q_{φ^1}, r_{φ^1}, $\tilde{\alpha}_{\varphi^1}$ and the performance index

$$\tilde{V}_{\varphi^1}(x, \pi^2) := E_x^{\pi^2}\left[\sum_{n=0}^{\infty} \tilde{\Gamma}_n^{\varphi^1} r_{\varphi^1}(x_n, b_n)\right], \quad \pi^2 \in \Pi^2, \ x \in \mathbf{X}, \tag{3.19}$$

where,

$$\tilde{\Gamma}_n^{\varphi^1} = \prod_{k=0}^{n-1} \tilde{\alpha}_{\varphi^1}(x_k, b_k, \xi_{k+1}) \quad \text{for } n \in \mathbb{N}, \text{ and } \tilde{\Gamma}_0^{\varphi^1} = 1.$$

Lemma 3.4.1. *For each $x \in \mathbf{X}$, $\varphi^2 \in \Pi_S^2$, and $\pi^1 \in \Pi^1$ there exists $\varphi^1 \in \Pi_M^1$ such that*

$$\tilde{V}_{\varphi^2}(x, \pi^1) = \tilde{V}_{\varphi^2}(x, \varphi^1). \tag{3.20}$$

Proof. Let $x \in \mathbf{X}$, $\varphi^2 \in \Pi_S^2$, and $\pi^1 \in \Pi^1$ be arbitrary. Let us consider the finite measures

$$M_{x,n}^{\pi^1}(K) := E_x^{\pi^1} \tilde{\Gamma}_n^{\varphi^2} I_K(x_n, a_n), \quad K \in \mathcal{B}(\mathbf{X} \times \mathbf{A}), \ n \in \mathbb{N}_0, \tag{3.21}$$

and

$$m_{x,n}^{\pi^1}(X) := E_x^{\pi^1} \tilde{\Gamma}_n^{\varphi^2} I_X(x_n), \quad X \in \mathcal{B}(\mathbf{X}), \ n \in \mathbb{N}_0. \tag{3.22}$$

Notice that $m_{x,n}^{\pi^1}$ is the marginal of $M_{x,n}^{\pi^1}$ on $\mathbf{X}$ (see Definition B.2.1, in Appendix B). Indeed, for $X \in \mathcal{B}(\mathbf{X})$,

$$\begin{aligned} M_{x,n}^{\pi^1}(X \times \mathbf{A}) &= E_x^{\pi^1} \tilde{\Gamma}_n^{\varphi^2} I_{X \times \mathbf{A}}(x_n, a_n) = E_x^{\pi^1} \tilde{\Gamma}_n^{\varphi^2} I_X(x_n) I_{\mathbf{A}}(a_n) \\ &= E_x^{\pi^1} \tilde{\Gamma}_n^{\varphi^2} I_X(x_n) = m_{x,n}^{\pi^1}(X). \end{aligned}$$

Then, by Proposition B.2.2, there exists a stochastic kernel φ_n^1 on $\mathbf{A}$ given $\mathbf{X}$ such that, for $X \in \mathcal{B}(\mathbf{X})$ and $A \in \mathcal{B}(\mathbf{A})$,

$$M_{x,n}^{\pi^1}(X \times A) = \int_X \varphi_n^1(A|y) m_{x,n}^{\pi^1}(dy). \tag{3.23}$$

Furthermore, since $M_{x,n}^{\pi^1}$ is concentrated on $\mathbb{K}_{\mathbf{A}}$, we can assume that $\varphi_n^1(A(y)|y) = 1$, $y \in \mathbf{X}$. Thus, φ_n^1 determines the Markov strategy $\varphi^1 := \{\varphi_n^1\} \in \Pi_M^1$.

For φ^1, we also define $M_{x,n}^{\varphi^1}$ and $m_{x,n}^{\varphi^1}$ in a similar way as in (3.21) and (3.22).

On the other hand, observe that for any measurable simple function f on $\mathbf{X} \times \mathbf{A}$ we have

$$E_x^{\pi^1} \tilde{\Gamma}_n^{\varphi^2} f(x_n, a_n) = \int_{\mathbf{X}\times\mathbf{A}} f(y,a) M_{x,n}^{\pi^1}(d(y,a)). \tag{3.24}$$

And from Proposition B.2.3 and the Monotone convergence theorem A.2.2 we obtain that (3.24) holds for any measurable function $f : \mathbf{X} \times \mathbf{A} \to \mathbb{R}$. Moreover, (3.24) also holds for any measurable function $f : \mathbf{X} \times \mathbf{A} \to \mathbb{R}$, with φ^1 instead of π^1, that is,

$$E_x^{\varphi^1} \tilde{\Gamma}_n^{\varphi^2} f(x_n, a_n) = \int_{\mathbf{X}\times\mathbf{A}} f(y,a) M_{x,n}^{\varphi^1}(d(y,a)). \tag{3.25}$$

Therefore, the key point in getting (3.20) is to show that

$$M_{x,n}^{\pi^1} = M_{x,n}^{\varphi^1}, \quad x \in \mathbf{X},\ n \in \mathbb{N}_0. \tag{3.26}$$

We will proceed by induction over n as follows. From (3.15), for $X \in \mathcal{B}(\mathbf{X})$ we have

$$m_{x,0}^{\pi^1}(X) := E_x^{\pi^1} I_X(x_0) = \delta_x(X) = m_{x,0}^{\varphi^1}(X).$$

Then, from (3.21) and (3.23) we obtain, for $X \times A \in \mathcal{B}(\mathbf{X} \times \mathbf{A})$,

$$\begin{aligned} M_{x,0}^{\pi^1}(X \times A) &:= \int_X \varphi_0^1(A|y) m_{x,0}^{\pi^1}(dy) \\ &= \int_X \varphi_0^1(A|y) m_{x,0}^{\varphi^1}(dy) = M_{x,0}^{\varphi^1}(X \times A). \end{aligned}$$

Hence (3.26) holds for $n = 0$. Now, let us suppose that (3.26) holds for some $n \in \mathbb{N}$. Then, by (3.17), (3.18), (3.22), and (3.24) with $f(y,a) = Q(X|y,a)\int_{\mathbf{S}} \tilde{\alpha}_{\varphi^2}(y,a,s)\theta(ds)$ we have the following:

$$\begin{aligned} m_{x,n+1}^{\pi^1}(X) &= E_x^{\pi^1} \tilde{\Gamma}_{n+1}^{\varphi^2} I_X(x_{n+1}) = E_x^{\pi^1}\left[E_x^{\pi^1}\left[\tilde{\Gamma}_{n+1}^{\varphi^2} I_X(x_{n+1})|h_n, a_n\right]\right] \\ &= E_x^{\pi^1}\left[E_x^{\pi^1}\left[\tilde{\Gamma}_n^{\varphi^2} \tilde{\alpha}_{\varphi^2}(x_n, a_n, \xi_{n+1}) I_X(x_{n+1})|h_n, a_n\right]\right] \\ &= E_x^{\pi^1}\left[\tilde{\Gamma}_n^{\varphi^2} E_x^{\pi^1}\left[\tilde{\alpha}_{\varphi^2}(x_n, a_n, \xi_{n+1}) I_X(x_{n+1})|h_n, a_n\right]\right] \\ &= E_x^{\pi^1}\left[\tilde{\Gamma}_n^{\varphi^2} Q_{\varphi^2}(X|x_n, a_n) \int_{\mathbf{S}} \tilde{\alpha}_{\varphi^2}(x_n, a_n, s)\theta(ds)\right] \\ &= E_x^{\pi^1}\left[\tilde{\Gamma}_n^{\varphi^2} f(x_n, a_n)\right] = \int_{\mathbf{X}\times\mathbf{A}} f(y,a) M_{x,n}^{\pi^1}(d(y,a)) \\ &= \int_{\mathbf{X}\times\mathbf{A}} f(y,a) M_{x,n}^{\varphi^1}(d(y,a)). \end{aligned} \tag{3.27}$$

Similarly, as in (3.27), with φ^1 instead of π^1, we obtain

$$m_{x,n+1}^{\varphi^1}(X) = \int_{\mathbf{X}\times\mathbf{A}} f(y,a) M_{x,n}^{\varphi^1}(d(y,a)). \tag{3.28}$$

From (3.27) and (3.28) we get

$$m_{x,n+1}^{\pi^1}(X) = m_{x,n+1}^{\varphi^1}(X). \tag{3.29}$$

Now, let us prove $M_{x,n+1}^{\pi^1} = M_{x,n+1}^{\varphi^1}$. On the one hand, from (3.23) and (3.29),

$$\begin{aligned} M_{x,n+1}^{\pi^1}(X\times A) &= \int_X \varphi_{n+1}^1(A|y) m_{x,n+1}^{\pi^1}(dy) \\ &= \int_X \int_A \varphi_{n+1}^1(da|y) m_{x,n+1}^{\pi^1}(dy) \\ &= \int_X \int_A \varphi_{n+1}^1(da|y) m_{x,n+1}^{\varphi^1}(dy). \end{aligned} \tag{3.30}$$

Further, observe that following similar arguments as in (3.24), it can be shown that for each $\pi^1 \in \Pi^1$, $n \in \mathbb{N}_0$, and any measurable function $g : \mathbf{X} \to \mathbb{R}$,

$$E_x^{\pi^1} \tilde{\Gamma}_n^{\varphi^2} g(x_n) = \int_{\mathbf{X}} g(y) m_{x,n}^{\pi^1}(dy). \tag{3.31}$$

On the other hand, from (3.21), (3.16), and (3.31) with $g(y) = I_X(y)\int_A \varphi_{n+1}^1(da|y)$ we obtain

$$\begin{aligned} M_{x,n+1}^{\varphi^1}(X\times A) &= E_x^{\varphi^1} \tilde{\Gamma}_{n+1}^{\varphi^2} I_{X\times A}(x_{n+1}, a_{n+1}) \\ &= E_x^{\varphi^1}\left[E_x^{\varphi^1}\left[\tilde{\Gamma}_{n+1}^{\varphi^2} I_{X\times A}(x_{n+1}, a_{n+1}) | h_{n+1}\right]\right] \\ &= E_x^{\varphi^1}\left[\tilde{\Gamma}_{n+1}^{\varphi^2} I_X(x_{n+1}) E_x^{\varphi^1}\left[I_A(a_{n+1}) | h_{n+1}\right]\right] \\ &= E_x^{\varphi^1}\left[\tilde{\Gamma}_{n+1}^{\varphi^2} I_X(x_{n+1}) \int_A \varphi_{n+1}^1(da_{n+1}|x_{n+1})\right] \\ &= E_x^{\varphi^1} \tilde{\Gamma}_{n+1}^{\varphi^2} g(x_{n+1}) = \int_{\mathbf{X}} g(y) m_{x,n+1}^{\varphi^1}(dy) \\ &= \int_X \int_A \varphi_{n+1}^1(da|y) m_{x,n+1}^{\varphi^1}(dy). \end{aligned} \tag{3.32}$$

Hence, (3.30) and (3.32) yield (3.26).

Finally, by (3.26) and (3.24) with $f(y,a) = r_{\varphi^2}(y,a)$,

$$\begin{aligned} E_x^{\pi^1}\left[\tilde{\Gamma}_n^{\varphi^2} r_{\varphi^2}(x_n, a_n)\right] &= \int_{\mathbf{X}\times\mathbf{A}} r_{\varphi^2}(y,a) M_{x,n}^{\pi^1}(d(y,a)) \\ &= \int_{\mathbf{X}\times\mathbf{A}} r_{\varphi^2}(y,a) M_{x,n}^{\varphi^1}(d(y,a)) \\ &= E_x^{\varphi^1}\left[\tilde{\Gamma}_n^{\varphi^2} r_{\varphi^2}(x_n, a_n)\right]. \end{aligned}$$

Therefore, from (3.14) we obtain (3.20), providing that the interchange of expectation and sum holds, which follows from Assumption 3.5.2 below.

□

Remark 3.4.2. *Let us fix $\varphi^1 \in \Pi_S^1$. Then, once the necessary changes have been made in the proof of Lemma 3.4.1, we can conclude that for each $\pi^2 \in \Pi^2$ there exists $\varphi^2 \in \Pi_M^2$ such that*

$$\tilde{V}_{\varphi^1}(x, \pi^2) = \tilde{V}_{\varphi^1}(x, \varphi^2), \quad x \in \mathbf{X}, \tag{3.33}$$

where $\tilde{V}_{\varphi^1}$ is the performance index defined in (3.19).

Lemma 3.4.3.

(a) For each $\pi^1 \in \Pi^1$ and $\varphi^2 \in \Pi_S^2$ there exists $\varphi^1 \in \Pi_M^1$ such that

$$\tilde{V}(x, \pi^1, \varphi^2) = \tilde{V}(x, \varphi^1, \varphi^2), \quad x \in \mathbf{X}. \tag{3.34}$$

(b) For each $\pi^2 \in \Pi^2$ and $\varphi^1 \in \Pi_S^2$ there exists $\varphi^2 \in \Pi_M^2$ such that

$$\tilde{V}(x, \varphi^1, \pi^2) = \tilde{V}(x, \varphi^1, \varphi^2), \quad x \in \mathbf{X}. \tag{3.35}$$

Proof. Let $\pi^1 \in \Pi^1$ and $\varphi^2 \in \Pi_S^2$ be arbitrary strategies. Let us consider the corresponding performance index $\tilde{V}_{\varphi^2}(x, \pi^1)$, $x \in \mathbf{X}$. Lemma 3.4.1 yields the existence of $\varphi^1 \in \Pi_M^1$ such that $\tilde{V}_{\varphi^2}(x, \pi^1) = \tilde{V}_{\varphi^2}(x, \varphi^1)$, $x \in \mathbf{X}$. Thus, to obtain (3.34) it is enough to prove

$$\tilde{V}_{\varphi^2}(x, \pi^1) = \tilde{V}(x, \pi^1, \varphi^2), \quad x \in \mathbf{X}. \tag{3.36}$$

In order to prove so, we will compare the corresponding terms in the sums (3.14) and (3.8).

Indeed, for the first term, from (3.12),

$$\begin{aligned} E_x^{\pi^1} r_{\varphi^2}(x_0, a_0) &= \int_{\mathbf{A}} r_{\varphi^2}(x, a_0) \pi_0^1(da_0|x) \\ &= \int_{\mathbf{A}} \int_{\mathbf{B}} r(x, a_0, b_0) \varphi_0^2(db_0|x) \pi_0^1(da_0|x) \\ &= E_x^{\pi^1, \varphi^2} r(x_0, a_0, b_0). \end{aligned}$$

Moreover, from (3.13)

$$\begin{aligned} &E_x^{\pi^1} \tilde{\Gamma}_1^{\varphi^2} r_{\varphi^2}(x_1, a_1) = E_x^{\pi^1} \tilde{\alpha}_{\varphi^2}(x_0, a_0, \xi_1) r_{\varphi^2}(x_1, a_1) \\ &= \int_{\mathbf{A}\times\mathbf{S}\times\mathbf{X}\times\mathbf{A}} \tilde{\alpha}_{\varphi^2}(x, a_0, \xi_1) r_{\varphi^2}(x_1, a_1) \pi_1^1(da_1|h_1) Q_{\varphi^2}(dx_1|x_0, a_0) \theta(d\xi_1) \pi_0^1(da_0|x) \\ &= \int_{\mathbf{A}\times\mathbf{B}\times\mathbf{S}\times\mathbf{X}\times\mathbf{A}\times\mathbf{B}} \tilde{\alpha}(x, a_0, b_0, \xi_1) r(x_1, a_1, b_1) \\ &\qquad \varphi_1^2(db_1|x) \pi_1^1(da_1|h_1) Q(dx_1|x_0, a_0, b_0) \theta(d\xi_1) \varphi_0^2(db_0|x) \pi_0^1(da_0|x) \\ &= E_x^{\pi^1, \varphi^2} \tilde{\alpha}(x_0, a_0, b_0, \xi_1) r(x_1, a_1, b_1) \\ &= E_x^{\pi^1, \varphi^2} \tilde{\Gamma}_1 r(x_1, a_1, b_1). \end{aligned}$$

By an induction argument it is shown that

$$E_x^{\pi^1} \tilde{\Gamma}_n^{\varphi^2} r_{\varphi^2}(x_n, a_n) = E_x^{\pi^1, \varphi^2} \tilde{\Gamma}_n r(x_n, a_n, b_n), \quad \forall n \in \mathbb{N}_0.$$

Therefore, from (3.14) and (3.8) we obtain (3.36), providing that the interchange of expectation and sum holds, which follows from Assumption 3.5.2 below.

The proof of part (b) is similar.

□

3.5 Existence of optimal strategies

Our approach to show the existence of optimal strategies is to prove the equivalence between the performance index (3.8) and one of those studied in Chapter 2, and therefore to apply Theorem 2.5.1. To this end, let us define the *mean* discount factor function $\alpha_\theta : \mathbb{K} \to (0,1)$ as

$$\alpha_\theta(x,a,b) := \int_{\mathbf{S}} \tilde{\alpha}(x,a,b,s)\theta(ds), \quad (x,a,b) \in \mathbb{K}, \tag{3.37}$$

and denote

$$\Gamma_n = \prod_{k=0}^{n-1} \alpha_\theta(x_k,a_k,b_k) \text{ if } n \in \mathbb{N}, \quad \text{and } \Gamma_0 = 1. \tag{3.38}$$

For each pair of strategies $(\pi^1,\pi^2) \in \Pi^1 \times \Pi^2$ and initial state $x \in \mathbf{X}$, we define

$$V(x,\pi^1,\pi^2) := E_x^{\pi^1,\pi^2}\left[\sum_{n=0}^{\infty} \Gamma_n r(x_n,a_n,b_n)\right]. \tag{3.39}$$

We suppose, for the moment, that the class of Markov strategies are sufficient in the sense of Theorem 3.5.5. Then we can establish our first result as follows.

Theorem 3.5.1. *For every initial state $x \in \mathbf{X}$ and every pair of strategies $(\varphi^1,\varphi^2) \in \Pi_M^1 \times \Pi_M^2$,*

$$V(x,\varphi^1,\varphi^2) = \tilde{V}(x,\varphi^1,\varphi^2) \tag{3.40}$$

Proof. For each $x \in \mathbf{X}$ and $(\varphi^1,\varphi^2) \in \Pi_M^1 \times \Pi_M^2$, from (3.37)

$$\begin{aligned}
&E_x^{\varphi^1,\varphi^2}\tilde{\Gamma}_1 r(x_1,a_1,b_1) = E_x^{\varphi^1,\varphi^2}\tilde{\alpha}(x_0,a_0,b_0,\xi_1)r(x_1,a_1,b_1)\\
&= \int_{\mathbf{A}\times\mathbf{B}\times\mathbf{S}\times\mathbf{X}\times\mathbf{A}\times\mathbf{B}} \tilde{\alpha}(x_0,a_0,b_0,\xi_1)r(x_1,a_1,b_1)\\
&\qquad \varphi_1^2(db_1|x_1)\varphi_1^1(da_1|x_1)Q(dx_1|x_0,a_0,b_0)\theta(d\xi_1)\varphi_0^2(db_0|x)\varphi_0^1(da_0|x)\\
&= \int_{\mathbf{A}\times\mathbf{B}}\int_{\mathbf{S}} \tilde{\alpha}(x_0,a_0,b_0,\xi_1)\theta(\xi_1)\int_{\mathbf{X}}\int_{\mathbf{A}\times\mathbf{B}} r(x_1,a_1,b_1)\\
&\qquad \varphi_1^2(db_1|x_1)\varphi_1^1(da_1|x_1)Q(dx_1|x_0,a_0,b_0)\varphi_0^2(db_0|x)\varphi_0^1(da_0|x)\\
&= E_x^{\varphi^1,\varphi^2}\alpha_\theta(x_0,a_0,b_0)r(x_1,a_1,b_1)\\
&= E_x^{\varphi^1,\varphi^2}\Gamma_1 r(x_1,a_1,b_1).
\end{aligned}$$

Following an induction argument it is shown that

$$E_x^{\varphi^1,\varphi^2}\tilde{\Gamma}_n r(x_n,a_n,b_n) = E_x^{\varphi^1,\varphi^2}\Gamma_n r(x_n,a_n,b_n), \quad \forall\, n \in \mathbb{N}_0.$$

Therefore, from (3.8) and (3.39) we get (3.40). □

The existence of a value of the game as well as a pair of optimal strategies is analyzed under the following conditions. (See Assumption 2.4.1)

Assumption 3.5.2. *The game model* (3.1) *satisfies the following:*

(a) For each $x \in \mathbf{X}$, the sets $A(x)$ and $B(x)$ are compact.

(b) For each $(x, a, b) \in \mathbb{K}$, $r(x, \cdot, b)$ is upper semicontinuous (u.s.c.) on $A(x)$, and $r(x, a, \cdot)$ is lower semicontinuous (l.s.c.) on $B(x)$. Moreover, there exist a constant $r_0 > 0$ and a function $w : \mathbf{X} \to [1, \infty)$ such that

$$|r(x, a, b)| \le r_0 w(x), \tag{3.41}$$

and the functions

$$\int_{\mathbf{X}} w(y) Q(dy|x, \cdot, b) \quad \text{and} \quad \int_{\mathbf{X}} w(y) Q(dy|x, a, \cdot) \tag{3.42}$$

are continuous on $A(x)$ and $B(x)$, respectively.

(c) For each $(x, a, b) \in \mathbb{K}$ and each bounded measurable function u on $\mathbf{X}$, the functions

$$\int_{\mathbf{X}} u(y) Q(dy|x, \cdot, b) \quad \text{and} \quad \int_{\mathbf{X}} u(y) Q(dy|x, a, \cdot)$$

are continuous on $A(x)$ and $B(x)$, respectively.

(d) The function $\tilde{\alpha}(x, a, b, s)$ is continuous on $\mathbb{K} \times \mathbf{S}$, and

$$\alpha^* := \sup_{(x,a,b) \in \mathbb{K}} \alpha_\theta(x, a, b) < 1. \tag{3.43}$$

(e) There exists a positive constant β such that $1 \le \beta < (\alpha^*)^{-1}$, and for all $(x, a, b) \in \mathbb{K}$

$$\int_{\mathbf{X}} w(y) Q(dy|x, a, b) \le \beta w(x). \tag{3.44}$$

Remark 3.5.3. *Notice that Assumption 3.5.2 (d) implies that Assumption 2.4.1 holds for α_θ. Indeed, Theorem A.2.3 (see Appendix A) yields the continuity of α_θ.*

We now present our main result which is consequence of Theorem 2.5.1.

Theorem 3.5.4. *Suppose that Assumption 3.5.2 holds. Then:*

(a) The game $\tilde{\mathcal{GM}}$ has a value $V^* \in \mathbb{B}_w$.

(b) The value V^* is the unique function in $\mathbb{B}_w$ such that $TV^* = V^*$.

(c) There exists a pair of strategies $(\varphi^1_*, \varphi^2_*) \in \Pi^1_S \times \Pi^2_S$ which is optimal respect to the Markov strategies

Hence, from Theorem 3.5.5 below, $(\varphi^1_*, \varphi^2_*)$ is an optimal pair of strategies for the game $\tilde{\mathcal{GM}}$.

Finally, we conclude presenting the proof of the sufficiency of Markov strategies.

Theorem 3.5.5. *Let* $(\varphi_*^1, \varphi_*^2) \in \Pi_S^1 \times \Pi_S^2$ *be an optimal pair with respect to the Markov strategies, i.e., for every* $x \in \mathbf{X}$*,*

$$\tilde{V}(x, \varphi^1, \varphi_*^2) \leq \tilde{V}(x, \varphi_*^1, \varphi_*^2) \leq \tilde{V}(x, \varphi_*^1, \varphi^2), \quad \forall (\varphi^1, \varphi^2) \in \Pi_M^1 \times \Pi_M^2. \tag{3.45}$$

Then $(\varphi_*^1, \varphi_*^2)$ *is an optimal pair with respect to all strategies, i.e.,* (3.9) *holds.*

Proof. Let $(\varphi_*^1, \varphi_*^2) \in \Pi_S^1 \times \Pi_S^2$ be an arbitrary optimal pair with respect to the Markov strategies. From Lemma 3.4.3, we have, for each $\varphi^2 \in \Pi_S^2$,

$$\max_{\pi^1 \in \Pi^1} \tilde{V}(x, \pi^1, \varphi^2) = \max_{\varphi^1 \in \Pi_M^1} \tilde{V}(x, \varphi^1, \varphi^2), \quad x \in \mathbf{X}, \tag{3.46}$$

and for each $\varphi^1 \in \Pi_S^1$,

$$\min_{\pi^2 \in \Pi^2} \tilde{V}(x, \varphi^1, \pi^2) = \min_{\varphi^2 \in \Pi_M^2} \tilde{V}(x, \varphi^1, \varphi^2), \quad x \in \mathbf{X}. \tag{3.47}$$

Then, from (3.45) and (3.46)

$$\begin{aligned} \tilde{V}(x, \varphi_*^1, \varphi_*^2) &\geq \max_{\varphi^1 \in \Pi_M^1} \tilde{V}(x, \varphi^1, \varphi_*^2) \\ &= \max_{\pi^1 \in \Pi^1} \tilde{V}(x, \pi^1, \varphi_*^2) \\ &\geq \tilde{V}(x, \pi^1, \varphi_*^2), \quad \forall\, \pi^1 \in \Pi^1, \; x \in \mathbf{X}. \end{aligned} \tag{3.48}$$

Similarly, from (3.45) and (3.47)

$$\begin{aligned} \tilde{V}(x, \varphi_*^1, \varphi_*^2) &\leq \min_{\varphi^2 \in \Pi_M^2} \tilde{V}(x, \varphi_*^1, \varphi^2) \\ &= \min_{\pi^2 \in \Pi^2} \tilde{V}(x, \varphi_*^1, \pi^2) \\ &\leq \tilde{V}(x, \varphi_*^1, \pi^2), \quad \forall\, \pi^2 \in \Pi^2, \; x \in \mathbf{X}. \end{aligned} \tag{3.49}$$

Therefore, combining (3.48) and (3.49) we obtain (3.9). □

3.6 Example: Semi-Markov games

A standard two-person semi-Markov game can be formulated as follows: If at the nth decision epoch the game is in state $x_n = x$, then the players independently choose actions $a_n = a$ and $b_n = b$ an the following happens:

(1) the game remains in the state x during a nonnegative random time ξ_{n+1} with distribution $H(\cdot|x, a, b)$;

(2) a payoff $r(x, a, b)$ is generated which represents a reward for player 1 and a cost for player 2;

(3) the game jumps to a new state $x_{n+1} = y$ according to a transition law $Q(\cdot|x, a, b)$.

Once the transition to the state y occurs, the process is repeated. (See [8], [12], [10]).

Observe that the decision epochs are $T_n := T_{n-1} + \xi_n$, $n \in \mathbb{N}$ and $T_0 = 0$, i.e., $T_n = \sum_{k=0}^{n-1} \xi_{k+1}$. The random variable $\xi_{n+1} = T_{n+1} - T_n$ is called the sojourn or holding time at state x_n.

The standard performance index is defined as follows. For each pair of strategies $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$ and initial state $x_0 = x \in \mathbf{X}$, we define

$$\begin{aligned} V(x, \pi^1, \pi^2) &= E_x^{\pi^1,\pi^2}\left[\sum_{n=0}^{\infty} e^{-\alpha T_n} r(x_n, a_n, b_n)\right] \\ &= E_x^{\pi^1,\pi^2}\left[r(x_0, a_0, b_0) + \sum_{n=1}^{\infty} e^{-\alpha \sum_{k=0}^{n-1} \xi_{k+1}} r(x_n, a_n, b_n)\right] \\ &= E_x^{\pi^1,\pi^2}\left[r(x_0, a_0, b_0) + \sum_{n=1}^{\infty} \prod_{k=0}^{n-1} e^{-\alpha \xi_{k+1}} r(x_n, a_n, b_n)\right], \end{aligned} \tag{3.50}$$

where $\alpha > 0$ is the so-called discount factor.

In our case, we assume that $\{\xi_n\}$ is a sequence of i.i.d. random variables with common exponential distribution with parameter $\lambda > 0$. Moreover, we suppose that the discount factor is a continuous function $\gamma : \mathbb{K} \to [d, \infty)$ for $d > 0$. Under this context, the index (3.50) takes the form

$$V(x, \pi^1, \pi^2) = E_x^{\pi^1,\pi^2}\left[r(x_0, a_0, b_0) + \sum_{n=1}^{\infty} \prod_{k=0}^{n-1} e^{-\gamma(x_n,a_n,b_n)\xi_{k+1}} r(x_n, a_n, b_n)\right]. \tag{3.51}$$

Hence, defining $\tilde{\alpha} : \mathbb{K} \times \mathbf{S} \to (0, 1)$ as

$$\tilde{\alpha}(x, a, b, \xi) = e^{-\gamma(x,a,b)\xi},$$

where $S = (0, \infty)$, then the performance index (3.51) takes the form (3.8).

Observe that the function $\tilde{\alpha}$ satisfies the Assumption 3.5.2 (d). Indeed, $\tilde{\alpha}$ is continuous and

$$\alpha_\theta(x, a, b) = \lambda \int_0^{\infty} e^{-\gamma(x,a,b)s} e^{-\lambda s} ds = \frac{\lambda}{\lambda + \gamma(x, a, b)} \leq \frac{\lambda}{\lambda + d}.$$

Thus

$$\alpha^* \leq \frac{\lambda}{\lambda + d} < 1.$$

Appendix A

Miscellaneous results

A.1 Lower semicontinuous functions

Definition A.1.1. *Let* $\mathbf{X}$ *be a metric space and* v *a function from* $\mathbf{X}$ *to* $\mathbb{R} \cup \{\infty\}$ *such that* $v(x) < \infty$ *for at least one point* $x \in \mathbf{X}$. v *is said to be*

- Lower semicontinuous (l.s.c.) at $x \in \mathbf{X}$ *if*

$$\liminf_{n \to \infty} v(x_n) \geq v(x)$$

for any sequence $\{x_n\}$ *in* $\mathbf{X}$ *that converges to* x. *The function is called* lower semicontinuous *(l.s.c.) if it is l.s.c. at every point of* X.

- Upper semicontinuous (u.s.c.) at $x \in \mathbf{X}$ *if*

$$\limsup_{n \to \infty} v(x_n) \leq v(x)$$

for any sequence $\{x_n\}$ *in* $\mathbf{X}$ *that converges to* x. *Similarly, the function is called* upper semicontinuous *(u.s.c.) if it is u.s.c. at every point of* X.

The following result is immediate.

Proposition A.1.2. *Let* $\mathbf{X}$ *be a metric space. A function* $v : \mathbf{X} \to \mathbb{R}$ *is u.s.c. at* $x \in \mathbf{X}$, *if and only if, the function* $-v$ *is l.s.c. at* x. *Moreover,* v *is continuous if and only if* v *is both l.s.c. and u.s.c.*

Let $L(\mathbf{X})$ be the family of all the functions on $\mathbf{X}$ that are l.s.c. and bounded below.

Proposition A.1.3. v *is in* $L(\mathbf{X})$ *if and only if there exists a sequence of continuous and bounded functions* v_n *on* $\mathbf{X}$ *such that* $v_n \uparrow v$.

Proposition A.1.4. *Let* $\mathbf{X}$ *be a compact metric space and* $f : \mathbf{X} \to \mathbb{R} \cup \{\infty\}$ *a l.s.c. function. Then the function* f *attains its minimum value at some* $x_0 \in \mathbf{X}$, *that is,* $f(x_0) \leq f(x)$ *for all* $x \in \mathbf{X}$. *Further, the set of points where* f *attains its minimum value is compact.*

Similarly, for an u.s.c. real-valued function f *defined on a compact space* $\mathbf{X}$, *the set of points for which the maximum is attained is nonempty and compact.*

A.2 Basic integration theorems

For proofs of the following three theorems, A.2.1– A.2.3, see theorems 1.5.4 – 1.4.6 in [3], pp. 25-26.

Theorem A.2.1 (Fatou's lemma). *If $f_n \geq 0$ then*

$$\liminf_{n\to\infty} \int f_n d\mu \geq \int (\liminf_{n\to\infty} f_n) d\mu.$$

Theorem A.2.2 (Monotone convergence theorem). *If $f_n \geq 0$ and $f_n \uparrow f$ then*

$$\int f_n d\mu \uparrow \int f d\mu.$$

Theorem A.2.3 (Dominated convergence theorem). *If $f_n \to f$ a.e., $|f_n| \leq g$ for all n, and g is integrable, then*

$$\int f_n d\mu \to \int f d\mu.$$

A.3 Banach's fixed point theorem

Definition A.3.1. *Let (S, d) be a metric space . A map $T : S \to S$ is called a* contraction *if there is a number $0 \leq \tau < 1$ such that*

$$d(Ts_1, Ts_2) \leq \tau d(s_1, s_2)$$

for all $s_1, s_2 \in S$. In this case τ is called the modulus of T.

For a proof of the following Proposition see Theorem 5.1-2 in [9], pp. 300-302.

Proposition A.3.2 (**Banach's Fixed Point Theorem**). *A contraction map T on a complete metric space (S, d) has a unique fixed point s^*. Moreover, $d(T^n s, s^*) \leq \tau^n d(s, s^*)$ for all $s \in S$, $n = 0, 1, \ldots$, where τ is the modulus of T, and $T^n := T(T^{n-1})$ for $n = 1, 2, \ldots$, with $T^0 := I$ (the identity).*

A.4 Fan's minimax theorem

Definition A.4.1. *Let f be a real-valued function defined on the product set $\mathbf{X} \times \mathbf{Y}$ of two arbitrary sets $\mathbf{X}$, $\mathbf{Y}$ (not necessarily topologized). f is said to be*

(a) convex *on $\mathbf{X}$ if for any two elements $x_1, x_2 \in \mathbf{X}$ and number $\alpha \in [0, 1]$, there exists an element $x_0 \in \mathbf{X}$ such that*

$$f(x_0, y) \leq \alpha f(x_1, y) + (1 - \alpha) f(x_2, y), \quad y \in \mathbf{Y};$$

(b) concave *on $\mathbf{Y}$ if for any two elements $y_1, y_2 \in \mathbf{Y}$ and number $\alpha \in [0, 1]$, there exists an element $y_0 \in \mathbf{Y}$ such that*

$$f(x, y_0) \geq \alpha f(x, y_1) + (1 - \alpha) f(x, y_2), \quad x \in \mathbf{X}.$$

Theorem A.4.2 (Ky Fan's Minimax Theorem)**.** *Let $\mathbf{X}$ and $\mathbf{Y}$ be two compact Hausdorff spaces, and f a real valued-function defined on $\mathbf{X} \times \mathbf{Y}$. Suppose that, for every $y \in \mathbf{Y}$, $f(\cdot, y)$ is l.s.c. on $\mathbf{X}$; and for every $x \in \mathbf{X}$, $f(x, \cdot)$ is u.s.c. on $\mathbf{Y}$. Then:*

(i) The equality

$$\min_{x \in \mathbf{X}} \max_{y \in \mathbf{Y}} f(x, y) = \max_{y \in \mathbf{Y}} \min_{x \in \mathbf{X}} f(x, y) \tag{A.1}$$

holds, if and only if the following condition holds: For any two finite sets $\{x_1, x_2, \ldots, x_n\} \subset \mathbf{X}$ and $\{y_1, y_2, \ldots, y_m\} \subset \mathbf{Y}$, there exist $x_0 \in \mathbf{X}$ and $y_0 \in \mathbf{Y}$ such that

$$f(x_0, y_k) \leq f(x_i, y_0) \quad (1 \leq i \leq n, 1 \leq k \leq m).$$

(ii) In particular, if f is convex on $\mathbf{X}$ and concave on $\mathbf{Y}$, then (A.1) *holds.*

Appendix B

Borel spaces, stochastic kernels and multifunctions

B.1 Borel spaces and stochastic kernels

A topological space will always be endowed with the Borel σ-algebra $\mathcal{B}(\mathbf{X})$, that is, the smallest σ-algebra of subsets of $\mathbf{X}$ that contains all of the open sets in $\mathbf{X}$. Thus, for sets or functions, "mesurable" means "Borel-measurable".

A Borel subset $\mathbf{X}$ of a complete and separable metric space is called a Borel Space, and its σ-algebra is denoted by $\mathcal{B}(\mathbf{X})$. A Borel subset of a Borel space is itself a Borel space.

Definition B.1.1. *Let $\mathbf{X}$ and $\mathbf{Y}$ be Borel spaces.A stochastic kernel on $\mathbf{X}$ given $\mathbf{Y}$ is a function $Q(\cdot|\cdot)$ such that*

(a) $Q(\cdot|y)$ is a probability measure on $\mathbf{X}$ for each fixed $y \in \mathbf{Y}$, and

(b) $Q(X|\cdot)$ is a measurable function on $\mathbf{Y}$ for each fixed $X \in \mathcal{B}(\mathbf{X})$.

The set of all stochastic kernels on $\mathbf{X}$ given $\mathbf{Y}$ is denoted by $\mathbb{P}(\mathbf{X}|\mathbf{Y})$.

Definition B.1.2. *The stochastic kernel $Q \in \mathbb{P}(\mathbf{X}|\mathbf{Y})$ is said to be*

(a) weakly continuous (or that it satisfies the Feller property) if the function

$$y \to \int v(x)Q(dx|y) \tag{B.1}$$

is continuous and bounded on $\mathbf{Y}$ for each continuous and bounded function v on $\mathbf{X}$;

(b) strongly continuous (or that it satisfies the strong Feller property) if the function in (B.1) *is continuous and bounded on $\mathbf{Y}$ for each bounded function v on $\mathbf{X}$.*

The following proposition is a consequence of Theorem 16 in [15, p. 89]

Proposition B.1.3. *Let g and g_n, $n = 1, 2, \ldots$, be integrable functions such that $g_n \to g$ almost everywhere. Then*

$$\int |g_n - g| \to 0$$

if and only if $\int |g_n| \to \int |g|$. In particular, if g and g_n are probability density functions, then the stated result is known as Sheffé's Theorem.

Proposition B.1.4 (Theorem of C. Ionescu Tulcea)**.** *Let $\mathbf{X}_0, \mathbf{X}_1, \ldots$ be a sequence of Borel spaces and, for $n \in \mathbb{N}_0$, define $\mathbf{Y}_n := \mathbf{X}_0 \times \cdots \times \mathbf{X}_n$ and $\mathbf{Y} := \prod_{n=0}^{\infty} \mathbf{X}_n$. Let ν be an arbitrary probability measure on $\mathbf{X}_0$ and, for every $n \in \mathbb{N}_0$, let $P_n \in \mathbb{P}(\mathbf{X}_{n+1}|\mathbf{Y}_n)$. Then there exists a unique probability measure P_ν on $\mathbf{Y}$ such that, for every measurable rectangle $B_0 \times \cdots \times B_n$ in $\mathbf{Y}_n$,*

$$\begin{aligned} P_\nu(B_0 \times \cdots \times B_n) = \int_{B_0} \nu(dx_0) \int_{B_1} P_0(dx_1|x_0) \int_{B_2} P_1(dx_2|x_0, x_1) \\ \cdots \int_{B_n} P_{n-1}(dx_n|x_0, \ldots, x_{n-1}). \end{aligned} \tag{B.2}$$

Moreover, for any nonnegative measurable function u on $\mathbf{Y}$, the function

$$x \mapsto \int u(y) P_x(dy)$$

is measurable on X_0, where P_x stands for P_ν when ν is the probability concentrated at $x \in \mathbf{X}_0$.

B.2 Probability Measures on Borel Spaces

Let $\mathbf{X}$ and $\mathbf{Y}$ be Borel spaces with Borel σ-algebras $\mathcal{B}(\mathbf{X})$ and $\mathcal{B}(\mathbf{Y})$, respectively. We denote the family of all probability measures on $\mathbf{X}$ by $\mathbb{P}(\mathbf{X})$.

Definition B.2.1. *Let μ be a probability measure on $\mathbf{X} \times \mathbf{Y}$, we denote by μ_1 the* marginal *(or* projection*) of μ on $\mathbf{X}$, i.e.,*

$$\mu_1(X) := \mu(X \times \mathbf{Y}), \quad \forall\ X \in \mathcal{B}(\mathbf{X}).$$

Proposition B.2.2 (Corollary 7.27.2 in [1, p. 139])**.** *Let $\mathbf{X}$ and $\mathbf{Y}$ be Borel spaces and let $\mu \in \mathbb{P}(\mathbf{X} \times \mathbf{Y})$. Then there exists a Borel-measurable stochastic kernel $\varphi(dy|x)$ on $\mathbf{Y}$ given $\mathbf{X}$ such that*

$$\mu(X \times Y) = \int_X \varphi(Y|x) \mu_1(dx) \quad \forall\ X \in \mathcal{B}(\mathbf{X}),\ Y \in \mathcal{B}(\mathbf{Y}),$$

where μ_1 is the marginal of μ on $\mathbf{X}$.

Proposition B.2.3. *Let f be a non-negative $\mathcal{F}$-measurable function. Then there exists a sequence of simple $\mathcal{F}$-measurable functions $\{s_n\}$ such that $0 \leq s_1 \leq \ldots \leq s_n \leq s_{n+1} \leq \ldots$ and $\lim_{n\to\infty} = f$.*

B.2.1 Convergence of Probability Measures

We assume that $\mathbb{P}(\mathbf{X})$ is endowed with the weak topology, which is induced by the weak convergence of measures defined as follows.

Definition B.2.4. *Let μ and μ_n, $n \geq 1$, be probability measures on $\mathbf{X}$. μ_n converges weakly to μ (which we write as $\mu_n \stackrel{w}{\to} \mu$) if*

$$\int v d\mu_n \to \int v d\mu \quad as \quad n \to \infty \tag{B.3}$$

for every continuous and bounded function v on $\mathbf{X}$.

Proposition B.2.5. *If $\mathbf{X}$ is a Borel space, then so is$\mathbb{P}(\mathbf{X})$. Moreover, if $\mathbf{X}$ is compact, then $\mathbb{P}(\mathbf{X})$ is also compact.*

Using Proposition A.1.3, one can easily prove the next result:

Proposition B.2.6. *If $\mu_n \stackrel{w}{\to} \mu$ and $v : \mathbf{X} \to \mathbb{R}$ is l.s.c. (see Definition A.1.1) and bounded below, then*

$$\liminf_{n\to\infty} \int v d\mu_n \geq \int v d\mu.$$

B.3 Multifunctions and Selectors

Let $\mathbf{X}$ and $\mathbf{A}$ be (nonempty) Borel spaces.

A **multifunction** (also known as a **correspondence** or **set-valued mapping**) ψ from $\mathbf{X}$ to $\mathbf{A}$ is a function such that $\psi(x)$ is a nonempty subset of $\mathbf{A}$ for all $x \in \mathbf{X}$. A single-valued mapping $\psi : \mathbf{X} \to \mathbf{A}$ is an example of a multifunction. The graph of the multifunction ψ is the subset of $\mathbf{X} \times \mathbf{A}$ defined as

$$Gr(\psi) := \{(x, a) | x \in \mathbf{X},\ a \in \psi(x)\}. \tag{B.4}$$

In this work, we write $\psi(x)$ as $A(x)$ $(B(x))$.

For every subset A of $\mathbf{A}$, let

$$\psi^{-1}(A) := \{x \in \mathbf{X} | \psi(x) \cap A \neq \emptyset\}$$

Definition B.3.1. *A multifunction ψ from $\mathbf{X}$ to $\mathbf{A}$ is said to be*

(a) Borel-measurable *if $\psi^{-1}(G)$ is a Borel subset of $\mathbf{X}$ for every open set $G \subset \mathbf{A}$;*

(b) upper semicontinuous *(u.s.c.) if $\psi^{-1}(F)$ is closed in $\mathbf{X}$ for every closed set $F \subset \mathbf{A}$;*

(c) lower semicontinuous *(l.s.c.) if $\psi^{-1}(G)$ is open in $\mathbf{X}$ for every open set $G \subset \mathbf{A}$;*

(d) continuous *if it is both u.s.c. and l.s.c.*

Let ψ be a Borel-measurable multifunction from $\mathbf{X}$ to $\mathbf{A}$, we denote by $\mathbb{F}$ the set of measurable functions $f : \mathbf{X} \to \mathbf{A}$ with $f(x) \in \psi(x)$ for all $x \in \mathbf{X}$. A function $f \in \mathbb{F}$ is called a measurable selector of the multifunction ψ.

Let $M : \mathbf{X} \to \mathbf{Y}$ be a correspondence, we define the correspondence $\Phi : \mathbf{X} \to \mathbb{P}(\mathbf{Y})$ by

$$\Phi(x) := \mathbb{P}(M(x)), \quad x \in \mathbf{X}.$$

Theorem B.3.2 (Measurable Selection Theorem)**.** *Let $\mathbf{X}$ and $\mathbf{Y}$ be Borel spaces, $\psi : \mathbf{X} \to \mathbf{Y}$ be a correspondence with nonempty compact values and suppose that the function $u : Gr(\psi) \to \mathbb{R}$ is Borel-measurable such that $u(x, \cdot)$ is u.s.c. on $\psi(x)$ for each $x \in \mathbf{X}$. Then, there exists a Borel-measurable selector $f : \mathbf{X} \to \mathbf{Y}$ for each ψ such that*

$$u(x, f(x)) = \max_{y \in \psi(x)} u(x, y) \quad \textit{para cada} \quad x \in \mathbf{X}.$$

Moreover, if u is l.s.c. on $\psi(x)$ for each $x \in \mathbf{X}$, then there exists a measurable selector $g : \mathbf{X} \to \mathbf{Y}$ for φ such that

$$u(x, g(x)) = \min_{y \in \psi(x)} u(x, y) \quad \textit{for each} \quad x \in \mathbf{X},$$

and the function v defined by $v(x) = \min_{y \in \psi(x)} u(x, y)$ is Borel-measurable.

Appendix C

Conditional expectation

Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{G}$ a sub-σ-algebra of $\mathcal{F}$, and η a $\mathcal{F}$-measurable random variable. If η is P-integrable, then the conditional expectation of η given $\mathcal{G}$, denoted by $E(\eta|\mathcal{G})$, is any function u on Ω such that

(i) u is $\mathcal{G}$-measurable, and

(ii) $\int_G u dP = \int_G \eta dP$ for every $G \in \mathcal{G}$.

If C is an event in $\mathcal{F}$, the conditional probability of C given $\mathcal{G}$ is defined as $P(C|\mathcal{G}) := E(I_C|\mathcal{G})$, where I_C is the indicator function of C.

If $\mathcal{G}$ is a σ-algebra generated by a collection $\{g_t, t \in T\}$ of measurable functions, that is, $\mathcal{G} = \sigma\{g_t, t \in T\}$, we usually write $E(\eta|g_t, t \in T)$ instead of $E(\eta|\mathcal{G})$.

Of course, the conditions (i) and (ii) above determine u only up to a P-null set in $\mathcal{G}$. However, in relationships involving conditional expectations, we will usually omit the qualifying "P-almost surely", since by $u = E(\eta|\mathcal{G})$ we simply mean that u satisfies (i) and (ii).

Proposition C.0.1. *Let η and η' be P-integrable random variables on $(\Omega, \mathcal{F}, P)$ and $\mathcal{G}$ and $\mathcal{G}'$ sub-σ-algebras of $\mathcal{F}$.*

(a) If η is a constant k, then $E(\eta|\mathcal{G}) = k$;

(b) $E(\eta + \eta'|\mathcal{G}) = E(\eta|\mathcal{G}) + E(\eta'|\mathcal{G})$;

(c) $E[E(\eta|\mathcal{G})] = E(\eta)$;

(d) If η is $\mathcal{G}$-measurable then $E(\eta\eta'|\mathcal{G}) = \eta E(\eta'|\mathcal{G})$; In particular, $E(\eta|\mathcal{G}) = \eta$;

(e) If $\mathcal{G} \subset \mathcal{G}'$, then

$$E[E(\eta|\mathcal{G})|\mathcal{G}'] = E[E(\eta|\mathcal{G}')|\mathcal{G}] = E(\eta|\mathcal{G});$$

(f) If $\eta_n \geq 0$ and $\eta_n \uparrow \eta$, then $E(\eta_n|\mathcal{G}) \uparrow E(\eta|\mathcal{G})$

(g) If $\eta_n \geq 0$, then $E(\sum_{n=1}^{\infty} \eta_n|\mathcal{G}) = \sum_{n=1}^{\infty} E(\eta_n|\mathcal{G})$.

Bibliography

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

[2] Cruz-Suárez H., Ilhuicatzi-Roldán R. and Montes-de-Oca R. (2014) J. Optim. Theory Appl. **162**: 329-346.

[3] Durret R. (2010) Probability: Theory and Examples, 4th Ed.. Cambridge University Press, New York.

[4] González-Hernández J., López-Martínez R.R. and Minjárez-Sosa J.A. (2008) Adaptive policies for stochastic systems under a randomized discounted cost criterion. Bol. Soc. Mat. Mexicana **14**(3): 149-163.

[5] González-Hernández J., López-Martínez R.R. and Minjárez-Sosa J.A. (2009) Approximation, estimation and control of stochastic systems under a randomized discounted cost criterion. Kybernetika **45**(5): 737-754.

[6] González-Sánchez D., Luque-Vásquez F. and Minjárez-Sosa J.A. (2017) Markov games with random state-actions-dependent discount factors: existence of optimal strategies. (Submitted for publication).

[7] Hernández-Lerma O. and Laserre J.B. (1996) Discrete-Time Markov Control Processes: Basic Optimality Criteria. Springer-Verlag, New York.

[8] Jáskiewicz, A. and Nowak, A. (2006) Approximation of noncooperative semi-Markov games. J. Optim. Theory Appl. **131**(1): 115-134.

[9] Kreyszig E. (1978) Introductory Functional Analysis With Applications. John Wiley & Sons, New York.

[10] Luque-Vásquez, F. (2002) Zero-sum semi-Markov games in Borel spaces: discounted and average payoff. Bol. Soc. Mat. Mexicana **8**: 227-241.

[11] Luque-Vásquez F. and Minjárez-Sosa J.A. (2017) Iteration algorithms in Markov decision processes with state-action-dependent discount factors and unbounded costs. In Operations Research: The Art of Making Good Decisions, K. Jian Ed., Ch. 4, Intech, Rijeka, Croatia.

[12] Luque-Vásquez, F. and Minjárez-Sosa, J.A. (2008) Two person zero-sum semi-Markov games with unknown holding times distribution on one side: A discounted payoff criterion. Appl. Math. Optim. **57**(3): 289-305.

[13] Minjárez-Sosa J.A. (2015) Markov control models with unknown random state-actions-dependent discount factors. TOP **23**(3): 743-772.

[14] Rieder, U. (1991) Non-cooperative dynamic games with general utility functions. In: T.E.S. Raghavan, T.S. Ferguson, T. Parthasarathy, O.J. Vrieze (eds.) Stochastic Games and Related Topics, Theory and Decision Library. Springer **7**: 161-174.

[15] Royden H.L. (1968) Real Analysis, 2nd ed. Macmillan, New York.

[16] Shapley, L.S. (1953) Stochastic games. Proc. Natl. Acad. Sci. USA **39**(10): 1095-1100.