Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional

Unidad Zacatenco

Departamento de Matemáticas

"Juegos markovianos a tiempo continuo: Optimalidad del caso descontado al caso promedio."

Tesis que presenta

Max Emmanuel Mitre Báez

Para obtener el Grado de

Maestro en Ciencias

en la Especialidad de

Matemáticas

Director de Tesis: Dr. Hector Jasso Fuentes

Ciudad de México Octubre, 2016

Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional

Campus Zacatenco

Department of Mathematics

"Continuous-time Markov games: From discounted to average optimality."

A Dissertation presented by

Max Emmanuel Mitre Báez

to obtain the Degree of

Master in Science

in the Speciality of

Mathematics

Thesis Advisor: Dr. Hector Jasso Fuentes

Mexico City October, 2016

Agradecimientos

Agradezco a mis padres, Mirna Báez y Alvaro Reyes, mis hermanos y mi pequeña sobrina, muchas gracias por haberme apoyado e impulsado a crecer profesionalmente, sin ustedes no hubiera logrado llegar a donde estoy, espero lograr que se sientan orgullosos de mi.

Gracias al Dr. Hector Jasso Fuentes, mi asesor, por regalarme su tiempo y esfuerzo, aclarar mis dudas, ayudarme a finalizar este trabajo y aceptarme como su futuro estudiante de doctorado.

Gracias al Dr. Onésimo Hernández Lerma y al Dr. José Daniel López Barrientos por aceptar ser mis sinodales y ayudarme con la revisión de ésta tesis y por sus observaciones.

También gracias a los profesores que ayudaron con mi formación de maestría, por sus cursos impartidos así como la dedicación a estos, me han hecho crecer mucho académicamente.

Gracias mi Anel Galaviz, mi suricata, por apoyarme en esos momentos que tanto te necesité y por compartir tu tiempo conmigo, sigamos creciendo juntos.

Por último, gracias al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico proporcionado que me permitió realizar este trabajo.

Resumen

En ésta tesis se estudian juegos markovianos no cooperativos a tiempo continuo. Nuestro análisis incluye el caso descontado con horizonte finito e infinito, y el caso de costo promedio (o ergódico). La principal herramienta es el método de programación dinámica, el cual nos permite ver los valores óptimos de un juego markoviano no cooperativo como la solución de un sistema de ecuaciones funcionales (no lineales). Nuestro principal resultado es relacionado a los llamados teoremas de verificación, los cuales vinculan la optimalidad de un juego con las ecuaciones funcionales antes mencionadas. Como un caso especial, estudiamos una clase de juegos cuya dinamica evoluciona como un proceso de difusión markoviano y mostramos que la teoría general de los capítulos anteriores aplica a este tipo especial de juegos.

Es importante notar que este trabajo extiende, al caso de juegos, la monografía [10], y también va un poco mas lejos que [11] (ver también [12]); precisamente, en referencia a [10] el autor presenta un estudio acerca de procesos de control markoviano en el contexto de una dinámica general; en otras palabras, el estudio en [10] se convierte en un caso particular al nuestro en el caso de un solo jugador. Por otro lado, en [11] o [12], el estudio de juegos markovianos no cooperativos fue analizado para el criterio de costo descontado. Por lo tanto, este trabajo rellena las referencias mencionadas anteriormente y trata de cubrir los criterios mas utilizados en la literatura; precisamente, costos descontados y promedio.

Abstract

In this thesis we study non-cooperative continuous-time Makov games. Our analysis includes the discounted case for finite-horizon and infinite-horizon, and the average (or ergodic) case. The main tool is the dynamic programming method, which allows to regard optimal values of a noncooperative Markov game as the solution of a system of functional (non-linear) equations. Our main results have to do with the so-named verification theorems that link the optimality of the original game with the solution of the aforementioned functional equations. As a special case, we study a class of games whose dynamic evolves as a Markov diffusion process and show that the general theory in previous chapters apply to this type of games.

It is worth noting that this work extends to the case of games the monograph [10], and also goes a bit further than [11], (see also [12]); namely, in reference [10] the author presents a nice study on continuous-time Markov control processes in a general dynamic framework; in other words, the study in [10] becomes a special case than ours for the case of one player. On the other hand, in [11] or [12], the study of noncooperative Markov games was also analysed for the discounted payoff criterion. Thus, this work fills-out the aforementioned references and tries to cover the most used payoff criteria in the literature in the dynamic framework; namely, discounted and average payoffs.

Contents

Preface

A *game* is a mathematical model of strategic interactions between independent agents, also known as players. Each player has the possibility of selecting an action over a fixed set, with the purpose of optimizing a certain performance index (payoff function). In contrast with a simple optimization problem, in a game model players' payoffs are linked with the strategies of the others. As a consequence, an individual optimization action is not enough to get the best revenue or cost, since a simple movement of strategy of other players may change the payoff values. This problematic was understood in the last past decades yielding some alternative definitions of optimality such as the concepts of equilibria.

Some of the common criteria are the so-named non-cooperative equilibria, also known as Nash equilibria. In this scenario, players act independently taking care only on their own benefit and they do not allow alliances nor coalitions.

This is the type of games we are concerned with in this thesis. Indeed, the aim of this work is to study non-cooperative continuous-time Makov games. Our analysis includes the discounted case for finite and infinite horizon, and the average (or ergodic) case. The main tool is the dynamic programming method, which allows us to regard optimal values of a noncooperative Markov game as the solution of a system of functional (non-linear) equations. Our main results have to do with the so-named verification theorems that link the optimality of the original game with the solution of the aforementioned functional equations.

This proposal extends to the case of games the monograph [10], and also goes further than [11] (see also [12]); namely, in reference [10] the author presents a study on continuous-time Markov control processes; in other words, the analysis in [10] becomes a special case of ours, for the case of one player. On the other hand, in [11] or [12] the study of non-cooperative Markov games was also analyzed for the discounted payoff criterion. Thus, this work fills-out the aforementioned references and tries to cover the most used payoff criteria in the literature in the dynamic framework; namely, discounted and average payoffs.

Related literature

Non-cooperative continuous-time Markov games has been studied separately for specific dynamical systems. For instance, [3] studies the case of (deterministic) differential games, whereas recent works such as [9, 14, 15] study stochastic differential games . On the other hand, [17] comprises the study of games for Markov chains, and [8] generalize games for the case of continuous-time Markov jump processes in Polish spaces. As we mentioned earlier, our work is based on references [10, 11, 12] that in fact were our departure point to the development of our results.

Outline

The rest of this thesis is organized as follows: In chapter 2 we introduce the game model, the general concept of strategies and some important special cases of strategies. Next we define the concepts of non-cooperative equilibria and saddle points. This last concept becomes a special case of the former when we are in the zero-sum game case. We conclude this section by showing some preliminary results about the existence the value of a Markov game as well as the existence of saddle points. Chapter 3 can divided in three parts. The first part studies the discounted case. A verification theorem is provided for both the zero and nonzero-sum scenarios. In the second part, we define the average payoff criterion and propose a set of sufficient conditions to get optimality results, in particular, as in the discounted case, we present its corresponding verification theorem. Finally, the third part is about the passing from the discounted to the average case. This is possible thanks to the well-known vanishing discount factor technique. We prove that under suitable conditions, it is possible to regard the average dynamic programming equation as a limit of discounted dynamic programming equations when the discount factor approaches zero. Finally, in Chapter 4 we apply our previous results to the case when the Markov process evolves as a Markov diffusion process. The aim is to illustrate that the general theory introduced in previous sections apply to this case; in particular, we will show the existence of non-cooperative equilibrium for the nonzero-sum case. This case of games is under the assumption that the drift of the dynamic as well as the cost rate have an additive structure.

Chapter 1

Markov Processes

1.1 Introduction

In this section we introduce the concepts of Markov processes, semigroups, infinitesimal generators, and the relation among them. Furthermore, we present useful results about ergodicity properties of Markov processes; in particular, we will show conditions to ensure the existence of solutions to the *Poisson equation* by means of the exponential (or geometric) ergodicity property of a Markov process, and state the so-named Abelian theorems that relate the concepts of resolvents with a certain type of average operators.

1.2 Elements of Markov processes

Let S be a metric space and $\{\mathtt{y}(\cdot), t \geqslant 0\}$ an S-valued stochastic process defined on a probability space $(\Omega, \mathcal{F}, P)$. Denote by $\mathcal{F}_t^{\mathtt{y}} = \sigma(\mathtt{y}(s) : s \leqslant t)$ the natural filtration of the process $\mathtt{y}(\cdot)$. We say that $\mathtt{y}(\cdot)$ is a *Markov process* if

$$P(\mathtt{y}(t) \in C | \mathcal{F}_s^{\mathtt{y}}) = P(\mathtt{y}(t) \in C | \mathtt{y}(s)) \quad \forall t \geqslant s \geqslant 0, C \in \mathcal{B}(S) \tag{1.2.1}$$

In general, if $\{\mathcal{G}_t\}$ is a filtration such that $\mathcal{F}_t^{\mathtt{y}} \subset \mathcal{G}_t$, $t \geqslant 0$, then $\mathtt{y}(\cdot)$ is a *Markov process with respect to* $\{\mathcal{G}_t\}$ if (1.2.1) holds with $\mathcal{G}_t$ instead of $\mathcal{F}_t^{\mathtt{y}}$.

Let $\mathcal{P}(S)$ be the space of probability measures on S. A function $P(s, y, t, C)$ defined for all $t \geqslant s \geqslant 0$, $y \in S$ and $C \in \mathcal{B}(S)$ is said to be a transition function if

$$\text{i)} P(s, y, t, \cdot) \in \mathcal{P}(S) \quad \forall (s, y, t) \in [0, \infty) \times S \times [s, \infty), \tag{1.2.2}$$

$$\text{ii)} P(s, y, s, \cdot) = \delta_y(\cdot) \text{ (the Dirac measure)} \quad \forall s \in [0, \infty), y \in S, \tag{1.2.3}$$

$$\text{iii)} P(s, \cdot, \cdot, C) \text{ is a measurable function on } S \times [s, \infty) \quad \forall (s, C) \in [0, \infty) \times \mathcal{B}(S), \tag{1.2.4}$$

$$\text{iv)} P(s, y, r, C) = \int_S P(s, y, t, dz) P(t, z, r, C), \quad r \geqslant t \geqslant s,\ y \in S,\ C \in \mathcal{B}(S), \tag{1.2.5}$$

Equation (1.2.5) is known as the Chapman-Kolmogorov equation.

A transition function is said to be *time-homogeneous* if

$$P(s,y,t,C) = P(0,y,t-s,C) =: P(t-s,y,C),$$

for all $t \geqslant s \geqslant 0$, $y \in S$, $C \in \mathcal{B}(S)$.

From (1.2.1), the function

$$P(s,y,t,C) := P(\mathtt{y}(t) \in C | \mathtt{y}(s) = y), \tag{1.2.6}$$

becomes a transition function for all $t \geqslant s \geqslant 0$, $y \in S$ and $C \in \mathcal{B}(S)$ (see [5], p. 77).

Defining the probability measure $\mu \in \mathcal{P}(S)$ by $\mu(C) := P(\mathtt{y}(0) \in C)$, for all $C \in \mathcal{B}(S)$ (the initial distribution of $\mathtt{y}(\cdot)$), the transition function (1.2.6) and the initial distribution μ, both determine the finite dimensional distributions of $\mathtt{y}(\cdot)$ by

$$\begin{aligned} &P(\mathtt{y}(0) \in C_0, \mathtt{y}(t_1) \in C_1, \ldots, \mathtt{y}(t_n) \in C_n) \\ &= \int_{C_0}\int_{C_1}\cdots\int_{C_{n-1}} P(t_{n-1}, y_{n-1}, t_n, C_n) \cdots P(t_1, y_1, t_2, dy_2) P(0, y_0, t_1, dy_1) \mu(dy_0), \end{aligned} \tag{1.2.7}$$

for every finite set $0 < t_1 < \cdots < t_n$ and $C_i \in \mathcal{B}(S)$ for $i = 0, \ldots, n$, and $n \geqslant 1$.

For the converse, we have the following result (see [6], Theorem 1.1, p. 157).

***Proposition* 1.1.** *Let S be a Polish space, $P(s,y,t,C)$ a transition function and μ a probability measure on S. Then, there exists a Markov Process $\mathtt{y}(\cdot)$ with values on S, with initial distribution μ, and whose finite dimensional distributions are uniquely determined by (1.2.7).*

1.3 Semigroups and infinitesimal generators

***Definition* 1.2.** *A one-parameter family $\{T(t), t \geqslant 0\}$ of bounded lineal operators on a Banach space B is called a semigroup if*

$$\text{(i)} \quad T(0) = I, \text{ the identity operator}, \tag{1.3.1}$$

$$\text{(ii)} \quad T(s+t) = T(s)T(t) \qquad \forall s, t, \geqslant 0. \tag{1.3.2}$$

If the semigroup satisfies that

$$\text{(iii)} \quad \lim_{t \to 0^+} T(t)f = f \qquad \forall f \in B, \tag{1.3.3}$$

it is said to be strongly continuous, whereas if the semigroup has the property

$$\text{(iv)} \quad \|T(t)\| \leqslant 1 \qquad \forall t \geqslant 0, \tag{1.3.4}$$

then it is referred to as a contraction semigroup.

***Definition* 1.3.** *Let $\{T(t)\}_{t\geqslant 0}$ be a semigroup satisfying the conditions (1.3.1)-(1.3.3). The infinitesimal generator of $\{T(t)\}_{t\geqslant 0}$ is the linear operator $\mathcal{L}$ (usually unbounded) on the Banach space B defined by*

$$\mathcal{L}f := \lim_{h\downarrow 0} h^{-1}[T(h)f - f],$$

with domain

$$D_{\mathcal{L}} = \{f \in B : \lim_{h\downarrow 0} h^{-1}[T(h)f - f] \text{ exists}\}.$$

The next result is the well-known Hille-Yosida theorem. It gives a characterization of a semigroup that verifies conditions (1.3.1) to (1.3.4).

***Theorem* 1.4** ([18], p. 129)**.** *A linear operator $\mathcal{L}$ on a Banach Space B is the generator of a semigroup $\{T(t)\}_{t\geqslant 0}$ satisfying (1.3.1)-(1.3.4) if and only if*

(i) $D(\mathcal{L})$ *is dense in* B*,*

(ii) $\mathcal{L}$ *is a closed operator,*

(iii) $(\mathcal{L} - \lambda I)$ *is invertible* $\quad \forall \lambda > 0$*, and*

(iv) $\|(\mathcal{L} - \lambda I)^{-1}\| \leqslant \lambda^{-1} \quad \forall \lambda > 0$.

1.3.1 Markov process semigroup

Let us now apply the previous concepts to the specific case in which $P(s,y,t,B)$ is the transition function of a Markov process with values in a Polish space S.

First, define $\hat{S} := [0,\infty) \times S$ and let $M(\hat{S})$ be the linear space of all real-valued measurable functions v on $\hat{S}$ such that

$$\int_S P(s,y,t,dz)|v(t,z)| < \infty \quad \text{for all} \quad t \geqslant s \geqslant 0,\ y \in S.$$

Now, for each $t \geqslant 0$, we define a function $T_t : M(\hat{S}) \to M(\hat{S})$ such that

$$T_t v(s,y) := \int_S P(s,y,s+t,dz)v(s+t,z). \tag{1.3.5}$$

***Proposition* 1.5.** *The family of operators $\{T_t\}_{t\geqslant 0}$ defined by (1.3.5) form a semigroup of operators on $M(\hat{S})$.*

Proof. Take $v \in M(\hat{S})$, then:

(i) Using (1.2.3) we get

$$T_0 v(s,y) = \int_S P(s,y,s,dz)v(s,z) = \int_S \delta_y(dz)v(s,z) = v(s,y).$$

(ii) Using (1.3.5), the Chapman-Kolmogorov equation and then interchanging integration orders we obtain that for every $v \in M(\hat{S})$,

$$\begin{aligned} T_{t+r}v(s,y) &= \int_S P(s,y,s+t+r,dz)v(s+t+r,z) \\ &= \int_S P(s,y,s+t,dw)\left[\int_S P(s+t,w,s+t+r,dz)v(s+t+r,z)\right] \\ &= \int_S P(s,y,s+t,dw)T_r v(s+t,w) \\ &= T_t T_r v(s,y) \qquad \forall t,r \geqslant 0. \end{aligned}$$

■

***Definition* 1.6.** *Let $M_0(\hat{S}) \subset M(\hat{S})$ be the set of functions $v \in M(\hat{S})$ under which:*

(a) The semigroup $\{T_t\}_{t\geqslant 0}$ in (1.3.5) is strongly continuous, i.e.,

$$\lim_{t\downarrow 0} T_t v(s,y) = T_0 v(s,y) = v(s,y) \qquad \forall (s,y) \in \hat{S},$$

(b) there exist $t_0 > 0$ and $u \in M(\hat{S})$ such that

$$T_t|v|(s,y) \leqslant u(s,y) \qquad \forall (s,y) \in \hat{S}, t_0 \geqslant t \geqslant 0.$$

Moreover, let $\mathcal{D}_{\mathcal{L}}(\hat{S}) \subset M_0(\hat{S})$ be the set of functions $v \in M_0(\hat{S})$ for which

(c) $v \in D_{\mathcal{L}}(\hat{S})$ (see Definition 1.3), where

$$\mathcal{L}v(s,y) := \lim_{t\downarrow 0} t^{-1}[T_t v(s,y) - v(s,y)] \ \forall\, (s,y) \in \hat{S}, \tag{1.3.6}$$

(d) $\mathcal{L}v$ is in $M_0(\hat{S})$,

(e) there exists $t_0 > 0$ and $u \in M(\hat{S})$ such that

$$t^{-1}|T_t v(s,y) - v(s,y)| \leqslant u(s,y),$$

for all $(s,y) \in \hat{S}$ and $t_0 \geqslant t \geqslant 0$.

Some properties of the operator $\mathcal{L}$ are listed below.

***Lemma* 1.7.** *For each $v \in \mathcal{D}_{\mathcal{L}}(\hat{S})$, the operator $\mathcal{L}$ is such that:*

(i) $\dfrac{d^+}{dt}T_t v := \lim_{h\downarrow 0} h^{-1}[T_{t+h}v - T_t v] = T_t\mathcal{L}v$,

(ii) $T_t v(s,y) - v(s,y) = \int_0^t T_r(\mathcal{L}v)(s,y)dr$,

(iii) if $\rho > 0$ and $v_\rho(s,y) := e^{-\rho s}v(s,y)$, then $v_\rho \in \mathcal{D}_{\mathcal{L}}(\hat{S})$ and

$$\mathcal{L}v_\rho(s,y) = e^{-\rho s}[\mathcal{L}v(s,y) - \rho v(s,y)].$$

Proof.

(i) Let $v \in \mathcal{D}_{\mathcal{L}}(\hat{S})$. By Proposition 1.5,

$$\lim_{h \downarrow 0} \frac{T_{t+h}v - T_t v}{h} = T_t \lim_{h \downarrow 0} \left[\frac{T_h v - v}{h} \right] = T_t \mathcal{L} v,$$

where the interchange of the limit $h \downarrow 0$ is due to the boundedness of the operator T_t.

(ii) Using part (i) we have that

$$\int_0^t T_r(\mathcal{L}v)(s,y)dr = \int_0^t \frac{d^+}{dr} T_t v(s,y) dr = T_t v(s,y) - T_0 v(s,y). \tag{1.3.7}$$

(iii) From Taylor series we have

$$e^{-\rho t} = 1 - \rho t + o(t) \quad \text{as } t \downarrow 0.$$

With this,

$$\begin{aligned} T_t v_\rho(s,y) - v_\rho(s,y) = &\int_S \left[P(s,y,s+t,dz) e^{-\rho(s+t)} v(s+t,z) - e^{-\rho s} v(s,y) \right] \\ =& e^{-\rho s} \left[\int_S P(s,y,s+t,dz) v(s+t,z) - v(s,y) \right. \\ &\left. +(-\rho t + o(t)) \int_S P(s,y,s+t,dz) v(s+t,z) \right] \\ =& e^{-\rho s} \left[T_t v(s,y) - v(s,y) \right] - e^{-\rho s} \left[\rho t + o(t) \right] T_t v(s,y). \end{aligned}$$

Finally, we multiply both sides by $1/t$ and let $t \downarrow 0$ to prove the result. ■

1.3.2 Dynkin's formula

Consider a Markov process $\{\mathtt{y}(t)\}_{t \geqslant 0}$ with values in a Polish space S and with transition function $P(s,y,t,C)$ for all $t \geqslant s \geqslant 0$, $y \in S$ and $C \in \mathcal{B}(S)$. The semigroup defined in (1.3.5) can be rewritten as

$$T_t v(s,y) = E_{s,y}[v(s+t, \mathtt{y}(s+t))],$$

where $E_{s,y}[\cdot] = E[\cdot | \mathtt{y}(s) = y]$ denotes the conditional expectation given $\mathtt{y}(s) = y$.

Thus, Lemma 1.7(ii) can also be interpreted as

$$E_{s,y}[v(s+t, \mathtt{y}(s+t))] - v(s,y) = E_{s,y}\left[\int_0^t \mathcal{L} v(s+w, \mathtt{y}(s+w)) dw \right], \tag{1.3.8}$$

for each $v \in \mathcal{D}_{\mathcal{L}}(\hat{S})$.

Equation (1.3.8) turns out to be a version of *Dynkin's formula* for the special case $t \equiv t(\omega)$, for all $\omega \in \Omega$. Besides, the infinitesimal generator of the semigroup $\{T_t\}_{t \geqslant 0}$ will be referred to as the *infinitesimal generator of the Markov process* $\mathtt{y}(\cdot)$.

1.4 Some ergodicity results

The following result is a special version of the so-named Abelian theorems (see [10, pp. 180-183] or [19, pp. 7-8]).

***Theorem* 1.8.** *Let $\alpha : [0, \infty) \to \mathbb{R}$ be a nondecreasing function with $\alpha(0) = 0$ and such that*

$$\limsup_{t\to\infty} \alpha(t)/t < \infty. \tag{1.4.1}$$

Then, for every $\rho > 0$,

(i) $$\int_0^\infty e^{-\rho t} d\alpha(t) = \rho \int_0^\infty e^{-\rho t} \alpha(t) dt. \tag{1.4.2}$$

(ii) $$\begin{aligned} \liminf_{t\to\infty} \alpha(t)/t &\leqslant \liminf_{\rho\to 0^+} \rho \int_0^\infty e^{-\rho t} d\alpha(t) \\ &\leqslant \limsup_{\rho\to 0^+} \rho \int_0^\infty e^{-\rho t} d\alpha(t) \leqslant \limsup_{t\to\infty} \alpha(t)/t. \end{aligned}$$

(iii) If the limit $j := \lim_{t\to\infty} \alpha(t)/t < \infty$ exists, then

$$\lim_{\rho\to 0^+} \rho \int_0^\infty e^{-\rho t} d\alpha(t) = j. \tag{1.4.3}$$

The proof of this result essentially follows from [10], p. 8. However, for the readers' convenience, we provide a proof based on our present context.

Proof.

(i) From (1.4.1) we get

$$\limsup_{t\to\infty} e^{-\rho t} \alpha(t) = 0, \tag{1.4.4}$$

using this fact and the integration-by-parts formula, we easily deduce (1.4.2)

(ii) Let $K := \liminf_{t\to\infty} \alpha(t)/t$. To prove the first inequality, let $\epsilon > 0$ and $\tau = \tau(\epsilon)$ be such that

$$\inf_{r\geqslant t} \alpha(r)/r \geqslant K - \epsilon \quad \forall t \geqslant \tau, \tag{1.4.5}$$

Then,

$$\begin{aligned} \rho^2 \int_\tau^t e^{-\rho r} \alpha(r) dr &= \rho^2 \int_\tau^t r e^{-\rho r} [\alpha(r)/r] dr \\ &\geqslant [K - \epsilon] \rho^2 \int_\tau^t r e^{-\rho r} dr. \end{aligned}$$

Then, for $t \geqslant \tau$, a simple use of integration by parts yields

$$\begin{aligned} \rho \int_0^t e^{-\rho r} d\alpha(r) &= \rho e^{-\rho t} \alpha(t) + \rho^2 \int_0^t e^{-\rho r} \alpha(r) dr \\ &\geqslant \rho e^{-\rho t} \alpha(t) + \rho^2 \int_0^\tau e^{-\rho r} \alpha(r) dr + (K - \epsilon) \rho^2 \int_\tau^t r e^{-\rho r} dr. \end{aligned}$$

Letting $t \to \infty$ and using (1.4.4), we get

$$\rho \int_0^\infty e^{-\rho r} d\alpha(r) \geqslant \rho^2 \int_0^\tau e^{-\rho r} \alpha(r) dr + (K - \epsilon)[e^{-\rho\tau} + \rho\tau e^{-\rho\tau}].$$

Finally, letting $\rho \downarrow 0$, we obtain

$$\liminf_{\rho \to 0^+} \rho \int_0^\infty e^{-\rho r} d\alpha(r) \geqslant K - \epsilon.$$

Since ϵ was arbitrary, the first inequality is proved. The second inequality is obvious and the third inequality is similar to the first one.

(*iii*) Since, in this case,

$$\limsup_{t\to\infty} \alpha(t)/t = \liminf_{t\to\infty} \alpha(t)/t,$$

this part follows immediately from the statement (*ii*). ■

A function $v \in M(\hat{S})$ (or in $M_0(\hat{S})$ or $\mathcal{D}_{\mathcal{L}}(\hat{S})$) is said to be *time-invariant* if $v(s, y) = v(t, y)$ for every $s, t \geqslant 0$. To simplify notation, a time-invariant function will be written as $v(y)$. We shall denote by $M(S)$ the set of all time-invariant functions. Similarly, we shall write $M_0(S)$ and $\mathcal{D}_{\mathcal{L}}(S)$ the spaces of invariant functions in $M_0(\hat{S})$ and $\mathcal{D}_{\mathcal{L}}(\hat{S})$, respectively.

Define $\mathbb{B}(S)$ as the space of all measurable functions $v : S \to \mathbb{R}$ with finite supremum norm, i.e.,

$$\|v\| = \sup_{y\in S} |v(y)| < \infty.$$

Furthermore, $\mathbb{M}(S)$ will denote the space of signed measures μ on S such that its total variation norm, denoted by $\|\mu\|_{TV}$, is finite.

For the case when we are dealing with time-homogeneous transition function and time-invariant functions, the following proposition gives sufficient conditions for the existence of a solution to the so-called Poisson equation introduced in (1.4.8) below.

***Proposition* 1.9.** *Let $P(t, y, B)$ be a time-homogeneous transition function which is **uniformly ergodic**; that is, there exist positive constants κ, γ, and a probability measure $\mu \in \mathbb{M}(S)$ such that*

$$\|P(t, y, \cdot) - \mu(\cdot)\|_{TV} \leqslant \kappa e^{-\gamma t} \qquad \forall t \geqslant 0, y \in S.$$

Let $r \in M_0(S) \bigcap L^1(\mu)$ and define

$$j^* := \int_S r(y)\mu(dy), \tag{1.4.6}$$

and

$$h(y) := \int_0^\infty (T_t r(y) - j^*) dt, \quad y \in S. \tag{1.4.7}$$

Then h belongs to $\mathbb{B}(S) \bigcap \mathcal{D}_{\mathcal{L}}(S)$ and the pair (j^*, h) satisfies the so-named Poisson equation

$$j^* = r(y) + \mathcal{L}h(y) \qquad \forall y \in S. \tag{1.4.8}$$

Proof. From (1.4.6) and the definition of j^*,

$$\begin{aligned} |T_t r(y) - j^*| &= \left| \int_S P(t, y, dz) r(z) - \int_S r(z)\mu(dz) \right| \\ &\leqslant \|r\| \|P(t, y, \cdot) - \mu(\cdot)\|_{TV} \\ &\leqslant \|r\| \kappa e^{-\gamma t} \qquad \forall t \geqslant 0, y \in S. \end{aligned} \tag{1.4.9}$$

The previous analysis implies that $h \in \mathbb{B}(S)$.

Now, since (1.4.9) allows us to interchange integrals, we have

$$\begin{aligned} T_s h(y) &= \int_0^\infty T_s[(T_t r(y) - j^*)]dt \\ &= \int_s^\infty [T_t r(y) - j^*]dt \\ &= h(y) - \int_0^s [T_t r(y) - j^*]dt, \end{aligned} \tag{1.4.10}$$

rearranging and multiplying by $1/s$ we get

$$j^* = \frac{1}{s}[T_s h(y) - h(y)] + \int_0^s T_t r(y) dr. \tag{1.4.11}$$

Finally, letting $s \downarrow 0$ in the later expression we get (1.4.8) and that $h \in \mathcal{D}_{\mathcal{L}}(S)$. ■

Chapter 2

Markov Games

2.1 Introduction

Markov games belong to the family of dynamic games which, under suitable conditions, evolve as Markov processes. As was pointed out in Chapter 1, here we only study the class of non-cooperative games.

For notational ease, we shall restrict ourselves to the case of two players, but the extension to any finite number of players ≥ 2 is completely analogous.

The rest of this chapter is organized as follows: First, we introduce the game model, the general concept of strategies, and some important special cases of strategies that will be used in this work. Next we define the concept of a non-cooperative equilibrium. We conclude this chapter by showing some preliminary results about the existence of the value of a Markov game and the existence of saddle points.

2.2 The game model and strategies

Strictly speaking, a (two player) continuous-time Markov Game can be expressed in a compact form as

$$\Gamma(S, A_1, A_2, \mathcal{L}^{(a_1,a_2)}, r_1, r_2), \tag{2.2.1}$$

whose elements are described as follows:

- S is the state space, which will be assumed to be a Polish space. An element y of S will be called a state of the game.

- For each player $i = 1, 2$, we have the pair

  $$(A_i, r_i),$$

  where A_i is the action space (or control set) for player i, which is also assumed to be a Polish space.

  The function r_i is real-valued and measurable on

  $$[0, \infty) \times S \times A_1 \times A_2,$$

  referred to as the *reward rate* function por the player i.

- For each pair $(a_1, a_2) \in A_1 \times A_2$, $\mathcal{L}^{a_1,a_2}$ is the infinitesimal generator of a S-valued Markov Process with transition function $P^{a_1,a_2}(s,y,t,B)$, with domain $\mathcal{D}_{\mathcal{L}^{a_1,a_2}}(\hat{S})$.

We say that a game Γ (in the sense of (2.2.1)) is *time-homogeneous* if the transition functions are time-homogeneous and the reward rates are time-invariant; that is,

$$P^{a_1,a_2}(s,y,t,B) = P^{a_1,a_2}(t-s,y,B) \quad \text{and} \quad r_i^{a_1,a_2}(s,y) = r_i^{a_1,a_2}(y) \;\; \text{for} \;\; i = 1,2,$$

where we have denoted $r_i^{a_1,a_2}(s,y) := r_i(s,y,a_1,a_2)$.

In this case $\mathcal{D}_{\mathcal{L}^{a_1,a_2}}(\hat{S})$ reduces to $\mathcal{D}_{\mathcal{L}^{a_1,a_2}}(S)$, where $\hat{S} := [0,\infty) \times S$.

Strategies.

For $i = 1,2$, we denote the Borel σ-algebra of A_i as $\mathcal{B}(A_i)$. Besides, let $\mathcal{P}(A_i)$ be the family of all probability measures on A_i.

***Definition* 2.1.** *A Markov (randomized) strategy for player i ($i = 1,2$) is defined as a family $\pi_i := \{\pi_i(\cdot|t,\cdot) : t \geqslant 0\}$ of stochastic kernels each of them defined on $\mathcal{B}(A_i) \times S$, satisfying:*

(a) for each $(t,y) \in \hat{S}$, $\pi_i(\cdot|t,y)$ is a probability measure on A_i, so that $\pi_i(A_i|t,y) = 1$;

(b) for each $D \in \mathcal{B}(A_i)$ and $t \geqslant 0$, $\pi_i(D|t,\cdot)$ is a Borel function on S;

(c) for each $D \in \mathcal{B}(A_i)$ and $y \in S$, $\pi_i(D|\cdot,y)$ is Borel on $[0,\infty)$.

We now introduce important subclasses of the above family of strategies:

Definition* 2.2.** *(a) A Markov (randomized) strategy $\pi_i := \{\pi_i(\cdot|t,\cdot) : t \geqslant 0\}$ is said to be* ***stationary *if there is a stochastic kernel π_i on $\mathcal{B}(A_i) \times S$ such that $\pi_i(\cdot|t,\cdot) = \pi_i(\cdot|\cdot)$ $\forall t \geqslant 0$.*

(b) Let $\mathbb{F}_i$ ($i = 1,2$) be the family of measurable functions $f_i : \hat{S} \to A_i$. A strategy $\pi_i := \{\pi_i(\cdot|t,\cdot) : t \geqslant 0\}$ is said to be ***deterministic*** *if $\pi_i(\cdot|t,y) = \delta_{f_i(t,y)}(\cdot)$, where $\delta_a(\cdot)$ denotes the Dirac measure at a.*

We denote by Π_i ($i = 1,2$) the family of Markov strategies for player i.

We will restrict ourselves to the above family of Markov strategies that satisfy the next assumptions.

***Assumption* 2.3.** *For each $(\pi_1,\pi_2) \in \Pi_1 \times \Pi_2$, there exists a strong Markov process $\mathrm{y}^{\pi_1,\pi_2}(\cdot) = \{\mathrm{y}(t), t \geqslant 0\}$ such that:*

(a) Almost all the sample paths of $\mathrm{y}(\cdot)$ are right-continuous with left-hand limits, and have finitely many discontinuities in any bounded time interval.

(b) *The infinitesimal generator $\mathcal{L}^{\pi_1,\pi_2}$ of $\{y(t)\}_{t\geqslant 0}$ is such that*

$$\mathcal{L}^{\pi_1,\pi_2}h(s,y) = \int_{A_1}\int_{A_2} \mathcal{L}^{a_1,a_2}h(s,y)\pi_2(da_2|s,y)\pi_1(da_1|s,y). \tag{2.2.2}$$

The set $\Pi_1 \times \Pi_2$ satisfying Assumption 2.3 is called the family of pairs of admissible Markov strategies.

The transition probabilities of the Markov process $y^{\pi_1,\pi_2}(\cdot)$ will be denoted as $P^{\pi_1,\pi_2}(s,y,t,B)$, for each pair $(\pi_1,\pi_2) \in \Pi_1 \times \Pi_2$.

In section 1.3 we have introduced the spaces of functions $M(\hat{S})$, $M_0(\hat{S})$ and $\mathcal{D}_{\mathcal{L}}(\hat{S})$, with $\mathcal{D}_{\mathcal{L}}(\hat{S}) \subset M_0(\hat{S}) \subset M(\hat{S})$. However, when dealing with Markov processes coming from Assumption 2.3, these spaces depend on the choice of each pair of strategies (π_1,π_2), so rigorously speaking we may write such spaces by $M^{\pi_1,\pi_2}(\hat{S})$, $M_0^{\pi_1,\pi_2}(\hat{S})$ and $\mathcal{D}_{\mathcal{L}^{\pi_1,\pi_2}}(\hat{S})$, and they will be supossed to fulfill the following conditions.

***Assumption* 2.4.** (a) *There exist nonempty spaces $\mathcal{M}(\hat{S}) \supset \mathcal{M}_0(\hat{S}) \supset \boldsymbol{D}(\hat{S})$ such that, for all $(\pi_1,\pi_2) \in \Pi_1 \times \Pi_2$,*

$$\mathcal{M}(\hat{S}) \subset M^{\pi_1,\pi_2}(\hat{S}), \qquad \mathcal{M}_0(\hat{S}) \subset M_0^{\pi_1,\pi_2}(\hat{S}), \qquad \boldsymbol{D}(\hat{S}) \subset \mathcal{D}_{\mathcal{L}^{\pi_1,\pi_2}}(\hat{S}).$$

In this case, the operator $\mathcal{L}^{\pi_1,\pi_2}$ is the closure of its restriction to $\boldsymbol{D}(\hat{S})$.

(b) *For $(\pi_1,\pi_2) \in \Pi_1 \times \Pi_2$ and $i \in \{1,2\}$, the reward rate $r_i^{\pi_1,\pi_2}$ defined by*

$$r_i^{\pi_1,\pi_2}(s,y) := \int_{A_1}\int_{A_2} r_i^{a_1,a_2}(s,y)\pi_2(da_2|s,y)\pi_1(da_1|s,y), \tag{2.2.3}$$

belongs to $\mathcal{M}_0(\hat{S})$.

If the game Γ is time-homogeneous, then we may write (2.2.3) as

$$r_i^{\pi_1,\pi_2}(s,y) = \int_{A_1}\int_{A_2} r_i^{a_1,a_2}(y)\pi_2(da_2|s,y)\pi_1(da_1|s,y). \tag{2.2.4}$$

If, moreover, (π_1,π_2) is a pair of stationary strategies, then (2.2.3) turns out to be

$$r_i^{\pi_1,\pi_2}(y) = \int_{A_1}\int_{A_2} r_i^{a_1,a_2}(y)\pi_2(da_2|y)\pi_1(da_1|y). \tag{2.2.5}$$

***Remark* 2.5.** *Observe that for the time-homogeneous case the statement of Assumption 2.4 is essentially the same, the only change is the replacement of the space $\hat{S}$ to S, as well as the special cases of the reward rate r^{π_1,π_2} such as those given in (2.2.4)-(2.2.5).*

Hereafter, we consider games satisfying Assumptions 2.3 and 2.4.

2.3 Noncooperative equilibria

In the noncooperative framework, players will try to perform the best they can in order to get the highest payoff along the game's life. To formalize that idea we present the following definition.

***Definition* 2.6.** *For each i=1,2, let $F_i(s,y,\pi_1,\pi_2)$ be the payoff function of the respective player. Then a pair $(\pi_1^*,\pi_2^*)\in\Pi_1\times\Pi_2$ is a noncooperative (a.k.a Nash) equilibrium if for all $(s,y)\in\hat{S}$,*

$$F_1(s,y,\pi_1^*,\pi_2^*) \geqslant F_1(s,y,\pi_1,\pi_2^*) \quad \forall\pi_1\in\Pi_1, \tag{2.3.1}$$

and

$$F_2(s,y,\pi_1^*,\pi_2^*) \geqslant F_2(s,y,\pi_1^*,\pi_2) \quad \forall\pi_2\in\Pi_2. \tag{2.3.2}$$

The above payoff function will depend on the reward rate, which at the same time depends on (s,y,π_1,π_2), as we will see later with more detail. Also, sometimes we will use either $S_\tau := [0,\tau]\times S$ or S instead of $\hat{S}$ in some of the subsequent results.

2.4 Zero-sum games

In this section we present a special type of games in which the revenue of one of the players is the lose for the other. This is formalized as follows.

***Definition* 2.7.** *Let $F_i(s,y,\pi_1,\pi_2)$ be the payoff function for the player i $(i=1,2)$. The game is said to be a zero-sum game if*

$$F_1(s,y,\pi_1,\pi_2) + F_2(s,y,\pi_1,\pi_2) = 0,$$

for every $(s,y)\in\hat{S}$, $(\pi_1,\pi_2)\in\Pi_1\times\Pi_2$

The above definition suggests that we only need to consider one payoff function in our problem, $F := F_1 = -F_2$. Namely, from the definition of equilibrium, we can see that the goal of player 1 is to maximize the function F over Π_1, while the second player will try to minimize it over Π_2. Using this fact, conditions (2.3.1) and (2.3.2) are reduced to

$$F(s,y,\pi_1,\pi_2^*) \leqslant F(s,y,\pi_1^*,\pi_2^*) \leqslant F(s,y,\pi_1^*,\pi_2), \tag{2.4.1}$$

for all $\pi_1\in\Pi_1$,$\pi_2\in\Pi_2$ and $(s,y)\in\hat{S}$. In this case, the pair (π_1^*,π_2^*) receives the name of saddle point.

For $(s,y)\in\hat{S}$, let

$$L(s,y) := \sup_{\pi_1\in\Pi_1}\inf_{\pi_2\in\Pi_2} F(s,y,\pi_1,\pi_2), \tag{2.4.2}$$

and

$$U(s,y) := \inf_{\pi_2\in\Pi_2}\sup_{\pi_1\in\Pi_1} F(s,y,\pi_1,\pi_2). \tag{2.4.3}$$

$L(s,y)$ is called the lower value of the game, while $U(s,y)$ is the upper value of the game.

The definitions of L and U can be interpreted as follows: Player 1 observes the best response of player 2 for each $\pi_1 \in \Pi_1$ and then picks a strategy in Π_1 that produces the best revenue in the worst case, ensuring the payoff $L(s,y)$. The interpretation for the upper value $U(s,y)$ is analogous.

It is obvious that

$$L(s,y) \leqslant U(s,y) \qquad \forall (s,y) \in \hat{S}, \tag{2.4.4}$$

and if the reverse inequality holds, then the game is said to have a value, which is denoted by $V(s,y)$; that is,

$$V(s,y) = L(s,y) = U(s,y) \qquad \forall (s,y) \in \hat{S}.$$

This last fact together with (2.4.1) leads to the next result.

***Proposition* 2.8.** *Let Γ be a zero-sum game with payoff function F. If (π_1^*, π_2^*) is a saddle point of the game, then*

$$V(s,y) = F(s,y,\pi_1^*,\pi_2^*) \qquad \forall (s,y) \in \hat{S}. \tag{2.4.5}$$

Proof. Using the last inequality in (2.4.1)

$$F(s,y,\pi_1^*,\pi_2^*) \leqslant \sup_{\pi_1 \in \Pi_1} \inf_{\pi_2 \in \Pi_2} F(s,y,\pi_1,\pi_2) \qquad \forall (s,y) \in \hat{S}. \tag{2.4.6}$$

Now, using the first inequality in (2.4.1) we obtain

$$\inf_{\pi_2 \in \Pi_2} \sup_{\pi_1 \in \Pi_1} F(s,y,\pi_1,\pi_2) \leqslant F(s,y,\pi_1^*,\pi_2^*) \qquad \forall (s,y) \in \hat{S}. \tag{2.4.7}$$

The two last inequalities imply that

$$U(s,y) \leqslant L(s,y) \qquad \forall (s,y) \in \hat{S}. \tag{2.4.8}$$

Combining (2.4.7) and (2.4.4) we obtain (2.4.5). ■

The following proposition gives sufficient conditions for a pair of strategies to be a saddle point.

***Proposition* 2.9.** *If a pair (π_1^*, π_2^*) in $\Pi_1 \times \Pi_2$ is such that, for every $(s,y) \in \hat{S}$,*

$$F(s,y,\pi_1^*,\pi_2^*) = \sup_{\pi_1 \in \Pi_1} F(s,y,\pi_1,\pi_2^*) \tag{2.4.9}$$

$$= \inf_{\pi_2 \in \Pi_2} F(s,y,\pi_1^*,\pi_2), \tag{2.4.10}$$

then (π_1^*, π_2^*) is a saddle point.

Proof. Using (2.4.9) we obtain that

$$F(s,y,\pi_1^*,\pi_2^*) \geqslant F(s,y,\pi_1,\pi_2^*) \qquad \forall \pi_1 \in \Pi_1.$$

On the other hand (2.4.10) yields

$$F(s,y,\pi_1^*,\pi_2^*) \leqslant F(s,y,\pi_1^*,\pi_2) \qquad \forall \pi_2 \in \Pi_2,$$

which implies that (π_1^*, π_2^*) is a saddle point. ■

Chapter 3

Optimality Results

3.1 Introduction

This chapter is the most relevant of this thesis. Here, we establish a link between optimal values of players with the solution of a system of functional equations. Furthermore, with the use of these equations, it is possible to deduce the existence of noncooperative (Nash) equilibria. The previously mentioned is actually the well-known dynamic programming method that is generalized here to the case of a general Markov model. We shall work with two different type of payoff functions; namely, the discounted and the average payoff criteria.

This chapter is divided in three parts. Firstly, we study the discounted case. A verification theorem is provided for both the zero and nonzero-sum scenario. As for the second part, we define the average criterion and propose a set of sufficient conditions to get optimality results; in particular, as in the discounted case, we present its correspondent verification theorem. Finally, the third part is concerned with the passing from the discounted to the average case. This is possible thanks to the well-known vanishing discount factor technique. We prove that under suitable conditions, it is possible to regard the average dynamic programming equation as a limit of discounted dynamic programming equations when the discount factor approaches to zero.

3.2 Games with discounted payoff function

Let Γ be a Markov Game as in (2.2.1) that satisfies Assumptions 2.3 and 2.4. In this section we analyze two types of payoff functions for games with a *discount factor*, one of which the game ends in a fixed time (finite horizon) and another in which the game runs along an infinite period of time.

3.2.1 Finite-horizon payoff function

For this type of games the payoff function, for each player i ($i = 1, 2$), is given by

$$V_\tau^i(s, y, \pi_1, \pi_2) := E_{s,y}^{\pi_1,\pi_2}\left[\int_s^\tau e^{-\rho(t-s)} r_i^{\pi_1,\pi_2}(t, \mathrm{y}(t))dt + e^{\rho(\tau-s)} K_i(\tau, \mathrm{y}(\tau))\right], \quad (3.2.1)$$

for which $(s, y) \in S_\tau$, $K_i \in \mathcal{M}(\hat{S})$ and $\rho \in \mathbb{R}$. The time $\tau > 0$ is usually called the game's horizon, whereas $K_i(\tau, \mathrm{y}(\tau))$ is the terminal reward of player i. When $\rho \geqslant 0$, it can be interpreted as a discount factor.

The following proposition links a solution of a functional equation with the payoff function (3.2.1).

***Proposition* 3.1.** *For a fixed $\rho \in \mathbb{R}$, $i = 1, 2$ and a pair $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$, let $r_i^{\pi_1,\pi_2}(s, y) \in M_0(\hat{S})$ and $K(s, y)$ be measurable functions on $S_\tau := [0, \tau] \times S$, where $\tau > 0$ is a fixed time. If $v_i \in \mathcal{D}_{\mathcal{L}^{\pi_1,\pi_2}}(\hat{S})$ satisfies the equation*

$$\rho v_i(s, y) = r_i^{\pi_1,\pi_2}(s, y) + \mathcal{L}^{\pi_1,\pi_2} v_i(s, y) \qquad \forall (s, y) \in S_\tau, \quad (3.2.2)$$

and the terminal condition

$$v_i(\tau, y) = K_i(\tau, y), \quad (3.2.3)$$

for all $y \in S$. Then, for every $(s, y) \in S_\tau$

$$v_i(s, y) = E_{s,y}^{\pi_1,\pi_2}\left[\int_s^\tau e^{-\rho(t-s)} r_i^{\pi_1,\pi_2}(t, \mathrm{y}(t))dt + e^{\rho(\tau-s)} K_i(\tau, \mathrm{y}(\tau))\right]. \quad (3.2.4)$$

Moreover, if the equality in (3.2.2) is replaced by an inequality ($\geqslant$ or $\leqslant$) then the equality in (3.2.4) is replaced by the same inequality.

Proof. Fix $i = 1, 2$, and let $v_\rho^i(s, y) := e^{-\rho s} v_i(s, y)$, as in Lemma 1.7($iii$). By (3.2.2), we get

$$\mathcal{L}^{\pi_1,\pi_2} v_\rho^i(s, y) = e^{-\rho s}\left[\mathcal{L}^{\pi_1,\pi_2} v_i(s, y) - \rho v_i(s, y)\right] = -e^{-\rho s} r_i^{\pi_1,\pi_2}(s, y). \quad (3.2.5)$$

Now, applying Dynkin's formula to v_ρ and using (3.2.5), we obtain

$$\begin{aligned} E_{s,y}^{\pi_1,\pi_2}&\left[e^{-\rho(s+t)} v_i(s+t, \mathrm{y}(s+t))\right] - e^{-\rho s} v_i(s, y) \\ &= -E_{s,y}^{\pi_1,\pi_2}\left[\int_0^t e^{-\rho(s+w)} r_i^{\pi_1,\pi_2}(s+w, \mathrm{y}(s+w))dw\right] \\ &= -E_{s,y}^{\pi_1,\pi_2}\left[\int_s^{s+t} e^{-\rho w} r_i^{\pi_1,\pi_2}(w, \mathrm{y}(w))dw\right] \end{aligned} \quad (3.2.6)$$

Taking $\tau = t + s$ in (3.2.6) and using (3.2.3) we deduce that

$$E_{s,y}^{\pi_1,\pi_2}\left[e^{-\rho(\tau-s)} K_i(\tau, \mathrm{y}(\tau))\right] - v_i(s, y) = -E_{s,y}^{\pi_1,\pi_2}\left[\int_s^\tau e^{-\rho(w-s)} r_i^{\pi_1,\pi_2}(w, \mathrm{y}(w))dw\right]. \quad (3.2.7)$$

Multiplying (3.2.7) by $e^{\rho s}$, (3.2.4) follows.

The proof of the inequalities is similar, so we shall omit it. ■

Note that in Proposition 3.1 the number ρ is arbitrary but, for our present purposes, we will only require $\rho = 0$ or $\rho > 0$. When ρ is positive we will call it a *discount factor*.

If the function r is thought as a *reward rate*, then (3.2.4) can be interpreted as the expected reward during the time interval $[s, \tau]$ with the initial condition $\mathtt{y}(s) = y$ and terminal reward K.

The following result is a verification theorem of games with finite-horizon discounted payoff criterion.

***Theorem* 3.2.** *Let $\rho \in \mathbb{R}$ and $\tau > 0$ fixed. For each $i = 1, 2$ and each pair $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$, assume that $r_1^{\pi_1,\pi_2} \in \mathcal{M}_0(\hat{S})$ and K_i measurable on S_τ. Additionally, suppose that for each player there are functions $v_i(s, y) \in \boldsymbol{D}(\hat{S})$ and a pair $(\pi_1^*, \pi_2^*) \in \Pi_1 \times \Pi_2$ such that, for every $(s, y) \in S_\tau$,*

$$\rho v_1(s,y) = \max_{\pi_1 \in \Pi_1} \{ r_1^{\pi_1,\pi_2^*}(s,y) + \mathcal{L}^{\pi_1,\pi_2^*} v_1(s,y) \} \tag{3.2.8}$$

$$= r_1^{\pi_1^*,\pi_2^*}(s,y) + \mathcal{L}^{\pi_1^*,\pi_2^*} v_1(s,y), \tag{3.2.9}$$

$$\rho v_2(s,y) = \max_{\pi_2 \in \Pi_2} \{ r_2^{\pi_1^*,\pi_2}(s,y) + \mathcal{L}^{\pi_1^*,\pi_2} v_2(s,y) \} \tag{3.2.10}$$

$$= r_2^{\pi_1^*,\pi_2^*}(s,y) + \mathcal{L}^{\pi_1^*,\pi_2^*} v_2(s,y), \tag{3.2.11}$$

and the terminal conditions

$$v_1(\tau, y) = K_1(\tau, y) \quad \text{and} \quad v_2(\tau, y) = K_2(\tau, y) \qquad \forall y \in S. \tag{3.2.12}$$

Then (π_1^*, π_2^*) is a Nash equilibrium and for each player $i = 1, 2$, the expected payoff is

$$v_i(s,y) = V_\tau^i(s, y, \pi_1^*, \pi_2^*) \quad \forall (s,y) \in S_\tau. \tag{3.2.13}$$

Proof. First, using (3.2.9) and the terminal condition from (3.2.12) for player 1 altogether with Proposition 3.1 we get

$$v_1(s,y) = V_\tau^1(s, y, \pi_1^*, \pi_2^*)$$

for every $(s, y) \in S_\tau$. Analogously, using (3.2.11) and the other terminal condition we get (3.2.13) for $i = 2$.

Now, from (3.2.8) we get that

$$\rho v_1(s,y) \geqslant r_1^{\pi_1,\pi_2^*}(s,y) + \mathcal{L}^{\pi_1,\pi_2^*} v_1(s,y) \qquad \forall \pi_1 \in \Pi_1.$$

This fact together with the last statement of Proposition 3.1 yields

$$V_\tau^1(s, y, \pi_1^*, \pi_2^*) = v_1(s,y) \geqslant V_\tau^1(s, y, \pi_1, \pi_2^*) \qquad \forall \pi_1 \in \Pi_1. \tag{3.2.14}$$

This is exactly (2.3.1). In a similar way we get (2.3.2) for player 2, which means that (π_1^*, π_2^*) is a Nash equilibrium. ■

3.2.2 Zero-sum case: finite horizon

Here we present the analogue of Theorem 3.2 for the zero-sum context. First, remember that the payoff functions are such that

$$V_\tau(s,y,\pi_1,\pi_2) := V_\tau^1(s,y,\pi_1,\pi_2) = -V_\tau^2(s,y,\pi_1,\pi_2) \quad \forall (s,y,\pi_1,\pi_2), \tag{3.2.15}$$

where V_τ^i are as in (3.2.1), for $i = 1,2$. We have the following result.

Theorem 3.3. *Let $\rho \in \mathbb{R}$ and $\tau > 0$ fixed. Suppose that for each pair $(\pi_1,\pi_2) \in \Pi_1 \times \Pi_2$, $r^{\pi_1,\pi_2} \in \mathcal{M}_0(\hat{S})$ and K is measurable on S_τ. Additionally suppose that there is a function $v(s,y) \in \boldsymbol{D}(\hat{S})$ and a pair $(\pi_1^*,\pi_2^*) \in \Pi_1 \times \Pi_2$ such that, for every $(s,y) \in S_\tau$,*

$$\rho v(s,y) = \inf_{\pi_2 \in \Pi_2} \{ r^{\pi_1^*,\pi_2}(s,y) + \mathcal{L}^{\pi_1^*,\pi_2} v(s,y) \} \tag{3.2.16}$$

$$= \sup_{\pi_1 \in \Pi_1} \{ r^{\pi_1,\pi_2^*}(s,y) + \mathcal{L}^{\pi_1,\pi_2^*} v(s,y) \} \tag{3.2.17}$$

$$= r^{\pi_1^*,\pi_2^*}(s,y) + \mathcal{L}^{\pi_1^*,\pi_2^*} v(s,y), \tag{3.2.18}$$

and the terminal condition

$$v(\tau,y) = K(\tau,y) \qquad \forall y \in S. \tag{3.2.19}$$

Then (π_1^*,π_2^*) is a saddle point and the value of the game is

$$v(s,y) = V_\tau(s,y,\pi_1^*,\pi_2^*) \quad \forall (s,y) \in S_\tau. \tag{3.2.20}$$

Proof. By using expressions (3.2.18) and (3.2.19) together with Proposition 3.1 we easily deduce (3.2.20).

Now let define

$$F(s,y,\pi_1,\pi_2) := r^{\pi_1,\pi_2}(s,y) + \mathcal{L}^{\pi_1,\pi_2} v(s,y). \tag{3.2.21}$$

Interpreting this function as the payoff function of a game, from (3.2.16)-(3.2.18) along with Proposition 2.9 we get that the pair (π_1^*,π_2^*) is a saddle point, that is

$$r^{\pi_1,\pi_2^*}(s,y) + \mathcal{L}^{\pi_1,\pi_2^*} v(s,y) \leqslant F(s,y,\pi_1^*,\pi_2^*) \leqslant r^{\pi_1^*,\pi_2}(s,y) + \mathcal{L}^{\pi_1^*,\pi_2} v(s,y), \tag{3.2.22}$$

for every $(\pi_1,\pi_2) \in \Pi_1 \times \Pi_2$. From (3.2.18) we have that $F(s,y,\pi_1^*,\pi_2^*) = \rho v(s,y)$, using this fact and (3.2.22) along with the last statement of Proposition 3.1, we deduce that

$$V_\tau(s,y,\pi_1,\pi_2^*) \leqslant v(s,y) \leqslant V_\tau(s,y,\pi_1^*,\pi_2) \quad \forall \pi_1 \in \Pi_1, \pi_2 \in \Pi_2.$$

This means that $v(s,y)$ is a saddle point. Finally from Proposition 2.8 we get that $v(s,y)$ is the value of the game. ∎

3.2.3 Infinite-horizon payoff function

In this case, the payoff functions for each player i ($i = 1, 2$), are given by

$$V_i(s, y, \pi_1, \pi_2) := E_{s,y}^{\pi_1,\pi_2}\left[\int_s^\infty e^{-\rho(t-s)} r_i^{\pi_1,\pi_2}(t, \mathrm{y}(t))dt\right], \tag{3.2.23}$$

where $(s, y) \in \hat{S}$ and $\rho > 0$ a fixed discount factor.

The games with this type of payoff functions are slightly different from the ones mentioned in the previous section.

To make notation easier, let us remember that for each $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$,

$$\begin{aligned} T_t^{\pi_1,\pi_2} v(s, y) :=& \int_S P^{\pi_1,\pi_2}(s, t, s+t, dz) v(s+t, z) \\ =& E_{s,y}^{\pi_1,\pi_2}[v(s+t, \mathrm{y}(s+t))], \end{aligned} \tag{3.2.24}$$

for $v \in \mathcal{M}(\hat{S})$.

The following result relates the functional equation (3.2.25) with the payoff function (3.2.23).

***Proposition* 3.4.** *Given a fixed pair* $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$ *and* $i = 1, 2$, *assume that* $r_i^{\pi_1,\pi_2}$ *belongs to* $M_0(S)$ *and take* $\rho > 0$. *If* $v_i \in \mathcal{D}_{\mathcal{L}^{\pi_1,\pi_2}}(\hat{S})$ *is such that*

$$\rho v_i(s, y) = r_i^{\pi_1,\pi_2}(s, y) + \mathcal{L}^{\pi_1,\pi_2} v_i(s, y) \qquad \forall (s, y) \in \hat{S}, \tag{3.2.25}$$

and

$$e^{-\rho t} E_{s,y}^{\pi_1,\pi_2}[v_i(s+t, \mathrm{y}(s+t))] = e^{-\rho t} T_t^{\pi_1,\pi_2} v_i(s, y) \longrightarrow 0 \text{ as } t \to \infty \tag{3.2.26}$$

for every $(s, y) \in \hat{S}$, *then*

$$\begin{aligned} v_i(s, y) =& E_{s,y}^{\pi_1,\pi_2}\left[\int_s^\infty e^{-\rho(t-s)} r_i^{\pi_1,\pi_2}(t, \mathrm{y}(t))dt\right] \\ =& \int_0^\infty e^{-\rho t} T_t^{\pi_1,\pi_2} r_i^{\pi_1,\pi_2}(s, y)dt. \end{aligned} \tag{3.2.27}$$

Furthermore, if instead of the equality in (3.2.25) we have " $\leqslant$*" or "* $\geqslant$*", then the equality in (3.2.27) is replaced by the respective inequality.*

Proof. Using the same arguments from the proof of Proposition 3.1 we have that v_i satisfies (3.2.6). Multiplying this equation by $e^{\rho s}$ we get

$$v_i(s, y) = e^{-\rho t} E_{s,y}^{\pi_1,\pi_2}\left[v_i(s+t, \mathrm{y}(s+t)\right] + E_{s,y}^{\pi_1,\pi_2}\left[\int_s^{s+t} e^{-\rho(w-s)} r_i^{\pi_1,\pi_2}(w, \mathrm{y}(w))dw\right]. \tag{3.2.28}$$

Letting $t \to \infty$ in (3.2.28) and using (3.2.26) we obtain,

$$v_i(s, y) = E_{s,y}^{\pi_1,\pi_2}\left[\int_s^\infty e^{-\rho(w-s)} r_i^{\pi_1,\pi_2}(w, \mathrm{y}(w))dw\right], \tag{3.2.29}$$

which is (3.2.27). The treatment is the same if we replace the equality in (3.2.25) for an inequality. ■

In this case, the analogue of Theorem 3.2 is as follows.

***Theorem* 3.5.** *Let $\rho > 0$ be a fixed number. For each $i = 1, 2$ and each pair $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$, assume that $r_i^{\pi_1,\pi_2} \in \mathcal{M}_0(\hat{S})$. Suppose also that there are functions $v_i(s,y) \in \boldsymbol{D}(\hat{S})$ and a pair $(\pi_1^*, \pi_2^*) \in \Pi_1 \times \Pi_2$ that satisfy, for every $(s,y) \in \hat{S}$, the equations (3.2.8)-(3.2.11) along with the condition*

$$e^{-\rho t} T_t^{\pi_1,\pi_2} v_i(s,y) \to 0 \quad as \quad t \to \infty, \tag{3.2.30}$$

for every $i = 1, 2$, $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$, $(s,y) \in \hat{S}$. Then (π_1^*, π_2^*) is a Nash equilibrium of the game and for each player $i = 1, 2$, the expected payoff is

$$v_i(s,y) = V_i(s,y,\pi_1^*,\pi_2^*) \quad \forall (s,y) \in \hat{S}. \tag{3.2.31}$$

Proof. Using (3.2.9), (3.2.30) for v_1, and Proposition 3.4 we get (3.2.31) for the first player. A similar analysis with (3.2.11) and v_2 yields that (3.2.31) is satisfied for player 2.

Now, from (3.2.8) and using again the last statement of Proposition 3.4, we have

$$v_1(s,y) \geqslant V_1(s,y,\pi_1,\pi_2^*) \qquad \forall \pi_1 \in \Pi_1.$$

Similarly, by (3.2.10), we can deduce

$$v_2(s,y) \geqslant V_2(s,y,\pi_1^*,\pi_2) \qquad \forall \pi_2 \in \Pi_2, \tag{3.2.32}$$

which means that (π_1^*, π_2^*) is a Nash equilibrium. ■

3.2.4 Zero-sum case: infinite horizon

The analogous of Theorem 3.5 to the infinite-horizon zero sum case is presented in this subsection. As before, we have that

$$V(s,y,\pi_1,\pi_2) := V_1(s,y,\pi_1,\pi_2) = -V_2(s,y,\pi_1,\pi_2), \tag{3.2.33}$$

whit V as in (3.2.23).

***Theorem* 3.6.** *Let $\rho > 0$ a fixed number. Suppose that for every fixed pair $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$, $r \in \mathcal{M}_0(\hat{S})$. Suppose also that there is a function $v(s,y) \in \boldsymbol{D}(\hat{S})$ and a pair $(\pi_1^*, \pi_2^*) \in \Pi_1 \times \Pi_2$ under which the following relation is satisfied for every $(s,y) \in \hat{S}$*

$$\rho v(s,y) = \inf_{\pi_2 \in \Pi_2} \{ r^{\pi_1^*,\pi_2}(s,y) + \mathcal{L}^{\pi_1^*,\pi_2} v(s,y) \} \tag{3.2.34}$$

$$= \sup_{\pi_1 \in \Pi_1} \{ r^{\pi_1,\pi_2^*}(s,y) + \mathcal{L}^{\pi_1,\pi_2^*} v(s,y) \} \tag{3.2.35}$$

$$= r^{\pi_1^*,\pi_2^*}(s,y) + \mathcal{L}^{\pi_1^*,\pi_2^*} v(s,y) \tag{3.2.36}$$

along with the condition

$$e^{-\rho t} T_t^{\pi_1,\pi_2} v(s,y) \to 0 \quad as \quad t \to \infty, \tag{3.2.37}$$

for every $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$ *and all* $(s,y) \in \hat{S}$. *Then* (π_1^*, π_2^*) *is a Nash equilibrium (or saddle pint) and the value of the game is*

$$v(s,y) = V(s,y,\pi_1^*,\pi_2^*) \quad \forall (s,y) \in \hat{S}. \tag{3.2.38}$$

Proof. Using (3.2.36), (3.2.37) and Proposition 3.4 we obtain (3.2.38).

To verify that (π_1^*, π_2^*) is a saddle point, just follow the same steps as in Theorem 3.3, replacing the space S_τ with $\hat{S}$ and using Proposition 3.4 instead of Proposition 3.1. ■

3.3 Games with long-run average payoff function

In this class of games, the Markov game Γ will be considered to be time-homogeneous (see Chapter 2 for further details on this class of games). The function

$$J_\tau^i(y,\pi_1,\pi_2) := E_y^{\pi_1,\pi_2}\left[\int_0^\tau r_i^{\pi_1,\pi_2}(\mathtt{y}(t))dt\right], \tag{3.3.1}$$

represents the total expected payoff for player i on the time interval $[0,\tau]$ when player 1 uses $\pi_1 \in \Pi_1$ and player 2 uses $\pi_2 \in \Pi_2$, given the initial state $\mathtt{y}(0) = y$.

Then, the long-run expected average reward per unit time is defined as

$$J_i(y,\pi_1,\pi_2) := \liminf_{\tau\to\infty} \frac{1}{\tau} J_\tau^i(y,\pi_1,\pi_2), \tag{3.3.2}$$

for each player $i = 1,2$.

Notice that the inferior limit in (3.3.2) is always well defined, even when it can be infinite.

The following result will be helpful to prove the verification theorem in the average context.

***Lemma* 3.7.** *Fix a pair of strategies* $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$.

(i) Suppose that the function $r_1^{\pi_1,\pi_2}(y)$ *belongs to* $\mathcal{M}_0(S)$. *Furthermore, assume the existence of a constant* k_1 *and a time-invariant function* $h_1^{\pi_1,\pi_2} \in \boldsymbol{D}(S)$ *such that the pair* $(k_1, h_1^{\pi_1,\pi_2})$ *satisfies the Poisson equation*

$$k_1 = r_1^{\pi_1,\pi_2}(y) + \mathcal{L}^{\pi_1,\pi_2} h_1^{\pi_1,\pi_2}(y), \tag{3.3.3}$$

for every $y \in S$, *and also*

$$\lim_{t\to\infty} T_t^{\pi_1,\pi_2} h_1^{\pi_1,\pi_2}(y)/t = 0. \tag{3.3.4}$$

Then

$$k_1 = J_1(y,\pi_1,\pi_2), \tag{3.3.5}$$

for every $y \in S$.

(ii) *If the equality in (3.3.3) is replaced by an inequality then the equality in (3.3.5) is replaced by the same inequality.*

(iii) *Equivalently, suppose that the function* $r_2^{\pi_1,\pi_2}(y)$ *belongs to* $\mathcal{M}_0(S)$ *and assume the existence of a pair* $(k_2, h_2^{\pi_1,\pi_2})$ *consisting in a constant* k_2 *and a function* $h_2^{\pi_1,\pi_2} \in \boldsymbol{D}(S)$*, satisfying*

$$k_2 = r_2^{\pi_1,\pi_2}(y) + \mathcal{L}^{\pi_1,\pi_2} h_2^{\pi_1,\pi_2}(y), \tag{3.3.6}$$

for every $y \in S$*, and also*

$$\lim_{t\to\infty} T_t^{\pi_1,\pi_2} h_2^{\pi_1,\pi_2}(y)/t = 0. \tag{3.3.7}$$

Then

$$k_2 = J_2(y, \pi_1, \pi_2), \tag{3.3.8}$$

for every $y \in S$*.*

(iv) *Analogously, if the equality in (3.3.6) is replaced by an inequality then the equality in (3.3.8) is replaced by the same inequality.*

Proof. (i) Since $h_1^{\pi_1,\pi_2} \in \mathbf{D}(S)$, using Lemma 1.7 and (3.3.3) we obtain

$$\begin{aligned}
T_t^{\pi_1,\pi_2} h_1^{\pi_1,\pi_2}(y) - h_1^{\pi_1,\pi_2}(y) &= \int_0^t T_r^{\pi_1,\pi_2} \mathcal{L}^{\pi_1,\pi_2} h_1^{\pi_1,\pi_2}(y) dr \\
&= \int_0^t E_y^{\pi_1,\pi_2} \left[\mathcal{L}^{\pi_1,\pi_2} h_1^{\pi_1,\pi_2}(\mathtt{y}(r))\right] dr \\
&= E_y^{\pi_1,\pi_2} \left\{ \int_0^t [k_1 - r_1^{\pi_1,\pi_2}(\mathtt{y}(r))] dr \right\} \\
&= t k_1 - \int_0^t T_r^{\pi_1,\pi_2} r_1^{\pi_1,\pi_2}(y) dr,
\end{aligned}$$

which implies that

$$k_1 = \frac{1}{t} \int_0^t T_r^{\pi_1,\pi_2} r_1^{\pi_1,\pi_2}(y) dr + \frac{1}{t}[T_t^{\pi_1,\pi_2} h_1^{\pi_1,\pi_2}(y) - h_1^{\pi_1,\pi_2}(y)]. \tag{3.3.9}$$

Taking inferior limit as $t \to \infty$ in (3.3.9), altogether with (3.3.4), yields

$$\begin{aligned}
k_1 &= \liminf_{t\to\infty} \frac{1}{t} \int_0^t T_r^{\pi_1,\pi_2} r_1^{\pi_1,\pi_2}(y) dy \\
&= \liminf_{t\to\infty} \frac{1}{t} E_y^{\pi_1,\pi_2} \left[\int_0^t r_1^{\pi_1,\pi_2}(\mathtt{y}(r)) dr \right] = J_1(y, \pi_1, \pi_2),
\end{aligned}$$

for every $y \in S$.

Statements (ii)-(iv) use a similar argument, so we omit their proofs. ■

***Remark* 3.8.** *The choice of the constants* k_1 *and* k_2 *in Lemma 3.7 may depend on the pair of strategies* $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$*. However, for notational purposes, we omit such dependence.*

In the following definition, we shall use the notation

$$r^{a_1,\pi_2}(y) := \int_{A_2} r^{a_1,a_2}(y)\pi_2(da_2|y) \quad r^{\pi_1,a_2}(y) := \int_{A_1} r^{a_1,a_2}(y)\pi_1(da_1|y).$$

Definition* 3.9.** *A quadruple (j_1^*, h_1, j_2^*, h_2), in which $j_i^* : \Pi_k \to \mathbb{R}$ $(i \neq k)$ and the functions $h_1, h_2 \in \boldsymbol{D}(S)$, will be called a solution of the* ***average payoff optimality equation *(shortened to APOE) if*

$$j_1^*(\pi_2) = \sup_{a_1 \in A_1} \{r_1^{a_1,\pi_2}(y) + \mathcal{L}^{a_1,\pi_2} h_1(y)\} \tag{3.3.10}$$

for every $y \in S$ and $\pi_2 \in \Pi_2$, and

$$j_2^*(\pi_1) = \sup_{a_2 \in A_2} \{r_2^{\pi_1,a_2}(y) + \mathcal{L}^{\pi_1,a_2} h_2(y)\} \tag{3.3.11}$$

for every $y \in S$ and $\pi_1 \in \Pi_1$.

It is really important to notice that if (j_1^*, h_1, j_2^*, h_2) is solution of the APOE, then

$$j_1^*(\pi_2) \geqslant r_1^{\pi_1,\pi_2}(y) + \mathcal{L}^{\pi_1,\pi_2} h_1(y) \quad \forall \pi_1 \in \Pi_1.$$

Moreover, if $\pi_1 \in \Pi_1$ and $\pi_2 \in \Pi_2$ fulfill a transversality property, say

$$\lim_{t \to \infty} T_t^{\pi_1,\pi_2} h_1(y)/t = 0, \tag{3.3.12}$$

then, Lemma 3.7 gives

$$j_1^*(\pi_2) \geqslant J_1(y, \pi_1, \pi_2) \quad \forall y \in S \tag{3.3.13}$$

for every pair (π_1, π_2) for which (3.3.12) holds. These arguments apply, of course, if we replace j_1^*, h_1, r_1 and J_1 by j_2^*, h_2, r_2 and J_2.

Given a solution of the APOE, define $\Pi_1^* \times \Pi_2^*$ as the subset of strategies in $\Pi_1 \times \Pi_2$ for which (3.3.12) holds for h_1 and h_2.

Given a fixed $\hat{\pi}_2 \in \Pi_2$, define

$$\Pi_1^* := \{\pi_1 \in \Pi_1 \mid (\pi_1, \hat{\pi}_2) \in \Pi_1^* \times \Pi_2^*\}.$$

Note that the set Π_1^* depends on the fixed $\hat{\pi}_2 \in \Pi_2$ but for notational ease said dependence will be implicit. Equivalently we can define, for a fixed $\hat{\pi}_1 \in \Pi_1$, the set

$$\Pi_2^* := \{\pi_2 \in \Pi_2 \mid (\hat{\pi}_1, \pi_2) \in \Pi_1^* \times \Pi_2^*\}.$$

With the previous ideas we can claim the result below.

***Theorem* 3.10.** *Let (j_1^*, h_1, j_2^*, h_2) be a solution of the APOE and $\Pi_1^* \times \Pi_2^*$ as before. Then, for every $y \in S$,*

$$j_1^*(\pi_2) \geqslant \sup_{\pi_1 \in \Pi_1^*} J_1(y, \pi_1, \pi_2) \tag{3.3.14}$$

and

$$j_2^*(\pi_1) \geqslant \sup_{\pi_2 \in \Pi_2^*} J_2(y,\pi_1,\pi_2). \tag{3.3.15}$$

Additionally, if (π_1^*,π_2^*) *is a pair of strategies in* $\Pi_1^* \times \Pi_2^*$ *such that, for all* $y \in S$,

$$j_1^*(\pi_2^*) = r_1^{\pi_1^*,\pi_2^*}(y) + \mathcal{L}^{\pi_1^*,\pi_2^*} h_1(y) \tag{3.3.16}$$

and

$$j_2^*(\pi_1^*) = r_2^{\pi_1^*,\pi_2^*}(y) + \mathcal{L}^{\pi_1^*,\pi_2^*} h_2(y). \tag{3.3.17}$$

Then

$$j_1^*(\pi_2^*) = J_1(y,\pi_1^*,\pi_2^*) \geqslant J_1(y,\pi_1,\pi_2^*) \quad \forall \pi_1 \in \Pi_1^*, \tag{3.3.18}$$

$$j_2^*(\pi_1^*) = J_2(y,\pi_1^*,\pi_2^*) \geqslant J_2(y,\pi_1^*,\pi_2) \quad \forall \pi_2 \in \Pi_2^*; \tag{3.3.19}$$

in other words, the pair (π_1^*,π_2^*) *is a Nash equilibrium in the set* $\Pi_1^* \times \Pi_2^*$.

Proof. First observe that the use of (3.3.10) yields to

$$j_1^*(\pi_2) \geqslant r_1^{\pi_1,\pi_2}(y) + \mathcal{L}^{\pi_1,\pi_2} h_1(y) \qquad \forall \pi_1 \in \Pi_1^*, \tag{3.3.20}$$

thus, by using part (ii) of Lemma 3.7 ,we get

$$j_1^*(\pi_2) \geqslant J_1(y,\pi_1,\pi_2) \qquad \forall \pi_1 \in \Pi_1^*, \tag{3.3.21}$$

which yields (3.3.14). By a similar argument we obtain (3.3.15).

Now, to prove (3.3.18) just note that (3.3.16) and part (i) of Lemma 3.7 imply that

$$j_1^*(\pi_2^*) = J_1(y,\pi_1^*,\pi_2^*). \tag{3.3.22}$$

Hence by (3.3.14) we obtain (3.3.18). The analysis for (3.3.19) is similar. ■

3.3.1 Zero-Sum Case: Average Reward

The payoff functions in this scenario are such that

$$J(y,\pi_1,\pi_2) = J_1(y,\pi_1,\pi_2) = -J_2(y,\pi_1,\pi_2), \tag{3.3.23}$$

where each J_i is as in (3.3.2). The verification theorem for this case is as follows.

***Theorem* 3.11.** *Denote by* $\Pi_1' \times \Pi_2'$ *the subset of pairs of strategies under which*

$$\lim_{t\to\infty} T_t^{\pi_1,\pi_2} h(y)/t = 0 \tag{3.3.24}$$

holds. Let $(\pi_1^*,\pi_2^*) \in \Pi_1' \times \Pi_2'$ *be a pair of strategies. Suppose that there exist a constant* j^* *and a time-invariant function* $h \in \boldsymbol{D}$ *such that, for every* $y \in S$,

$$\begin{aligned} j^* &= \sup_{a_1 \in A_1} \{ r^{a_1,\pi_2^*}(y) + \mathcal{L}^{a_1,\pi_2^*} h(y) \} \\ &= \inf_{a_2 \in A_2} \{ r^{\pi_1^*,a_2}(y) + \mathcal{L}^{\pi_1^*,a_2} h(y) \}, \end{aligned} \tag{3.3.25}$$

If the pair $(\pi_1^*, \pi_2^*) \in \Pi_1' \times \Pi_2'$ *is such that,*

$$j^* = r^{\pi_1^*,\pi_2^*}(y) + \mathcal{L}^{\pi_1^*,\pi_2^*} h(y), \tag{3.3.26}$$

for every $y \in S$*, then*

$$j^* = J(y, \pi_1^*, \pi_2^*) \tag{3.3.27}$$

and (π_1^*, π_2^*) *is a saddle point in the set* $\Pi_1' \times \Pi_2'$*, i.e.,*

$$J(y, \pi_1, \pi_2^*) \leqslant j^* \leqslant J(y, \pi_1^*, \pi_2) \quad \forall (\pi_1, \pi_2) \in \Pi_1' \times \Pi_2'. \tag{3.3.28}$$

The proof of this result uses essentially the same arguments of Theorem 3.10 to deduce the first part, and of Theorem 3.3 to verify the saddle point condition.

3.4 Relation between discounted games and average reward games

Theorems 3.10 and 3.11 allow us to verify if a pair of strategies is a Nash equilibrium by means of the APOE; that is, by knowing the existence of solutions of such APOE as well as the existence of a pair of strategies that optimize the APOE, one gets Nash equilibria for the average payoff game. However, can we actually ensure a solution for the APOE equations? One way to find a solution is the use of the *vanishing discount-factor approach*, which consists in using the ρ-discounted criteria studied in Section 3.2 and regard the APOE as limiting equations, as $\rho \downarrow 0$, of those equations associated to the ρ-discounted case.

In the following lines we present a corollary of Theorem 1.8 that will be useful along this section.

In what follows we will assume that the reward rates $r_i^{\pi_1,\pi_2}$ satisfy the following assumption for each pair $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$.

***Assumption* 3.12.** *For* $i = 1, 2$*, the reward rate* $r_1^{\pi_1,\pi_2}$ *belongs to* $\mathcal{M}_0(S)$*, it is time-invariant, and non-negative.*

We define

$$\alpha(t) = \int_0^t T_w^{\pi_1,\pi_2} r_i^{\pi_1,\pi_2}(y) dw.$$

From the previous assumption, it is easy to verify that V_i in (3.2.23) is time-invariant and non-negative for each $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$. Furthermore, it satisfies that

$$V_i(y, \pi_1, \pi_2) = \int_0^\infty e^{-\rho t} T_t^{\pi_1,\pi_2} r_i^{\pi_1,\pi_2}(y) dt = \int_0^\infty e^{-\rho t} d\alpha(t) \qquad \forall y \in S \tag{3.4.1}$$

(note that $\alpha(t)$ depends on y). The function $J_i(y, \pi_1, \pi_2)$ in (3.3.2) becomes

$$J_i(y, \pi_1, \pi_2) = \liminf_{t \to \infty} \alpha(t)/t. \tag{3.4.2}$$

With these observations we can rewrite Theorem 1.8 in the following way.

***Corollary* 3.13.** *Let $(\pi_1,\pi_2) \in \Pi_1^* \times \Pi_2^*$ be a given pair. If the function $r_i^{\pi_1,\pi_2} \in \mathcal{M}_0(S)$ is time-invariant and nonnegative, and*

$$\limsup_{\tau\to\infty} \frac{1}{\tau}\int_0^\tau T_t^{\pi_1,\pi_2} r^{\pi_1,\pi_2}(y)dt < \infty$$

for every $y \in S$, then

(i) for each $\rho > 0$,

$$V_i^\rho(y,\pi_1,\pi_2) = \rho\int_0^\infty e^{-\rho t}\left[\int_0^t T_w^{\pi_1,\pi_2} r^{\pi_1,\pi_2}(y)dw\right]dt \tag{3.4.3}$$

where V_i^ρ is the function in (3.4.1),

(ii)
$$\begin{aligned} J_i(y,\pi_1,\pi_2) = \liminf_{t\to\infty}\frac{1}{t}\int_0^t T_w^{\pi_1,\pi_2} r^{\pi_1,\pi_2}(y)dw &\leqslant \liminf_{\rho\to 0^+} \rho V_i^\rho(y,\pi_1,\pi_2) \\ &\leqslant \limsup_{\rho\to 0^+} \rho V_i^\rho(y,\pi_1,\pi_2) \\ &\leqslant \limsup_{t\to\infty}\frac{1}{t}\int_0^t T_w^{\pi_1,\pi_2} r^{\pi_1,\pi_2}(y)dw, \end{aligned}$$

(iii) if the limit in (3.4.2) exists, then

$$J_i(y,\pi_1,\pi_2) = \lim_{\rho\to 0^+} \rho V_i^\rho(y,\pi_1,\pi_2).$$

In the following results we only analyze the case for the first player because analogous results apply to the second player.

The following assumption will be necessary.

***Assumption* 3.14.** *Let $\pi_2 \in \Pi_2$ be a fixed strategy. Then, for every pair such that (π_1,π_2) is in $\Pi_1^* \times \Pi_2^*$ we have that*

$$\sup_{\pi_1\in\Pi_1^*}\left[\limsup_{t\to\infty}\frac{1}{t}\int_0^t T_w^{\pi_1,\pi_2} r_1^{\pi_1,\pi_2}(y)dw\right] < \infty. \tag{3.4.4}$$

Now, given a fixed $\pi_2 \in \Pi_2$, let

$$V_{\rho,1}^*(y,\pi_2) := \sup_{\pi_1\in\Pi_1} V_{\rho,1}(y,\pi_1,\pi_2), \tag{3.4.5}$$

where $V_{\rho,1}$ is as in (3.2.23) (the subindex is just to emphasize the dependence on $\rho > 0$). Then the ρ-discount dynamic programming equation turns out to be

$$\rho V_{\rho,1}^*(y,\pi_2) = \sup_{a_1\in A_1}\{r_1^{a_1,\pi_2}(y) + \mathcal{L}^{a_1,\pi_2} V_{\rho,1}^*(y,\pi_2)\}. \tag{3.4.6}$$

Now, let $y' \in S$ be a fixed state and define

$$h_{\rho,1}(y,\pi_2) := V_{\rho,1}^*(y,\pi_2) - V_{\rho,1}^*(y',\pi_2) \quad \forall y \in S. \tag{3.4.7}$$

Then (3.4.6) becomes

$$\rho h_{\rho,1}(y,\pi_2) + \rho V^*_{\rho,1}(y',\pi_2) = \sup_{a_1\in A_1}\{r_1^{a_1,\pi_2}(y) + \mathcal{L}^{a_1,\pi_2}h_{\rho,1}(y,\pi_2)\}. \tag{3.4.8}$$

The last equation suggests to let ρ tend to zero to obtain (3.3.10) in the limit. Let us consider the following lemma that allow us to do so.

***Lemma* 3.15.** *Let $\pi_2 \in \Pi_2$ be a fixed strategy. Suppose that there exist $\rho_0 > 0$ and $b : S \times \Pi_2 \to \mathbb{R}$ such that*

$$|h_{\rho,1}(y,\pi_2)| \leqslant b_1(y,\pi_2) \quad \forall y \in S, \forall 0 < \rho < \rho_0. \tag{3.4.9}$$

Then there exist $\bar{j}_1 : \Pi_2 \to \mathbb{R}$ and a subsequence $\rho(n) \downarrow 0$ such that, for every $y \in S$,

(i) $\lim_{n\to\infty} \rho(n)h_{\rho(n),1}(y,\pi_2) = 0,$

(ii) $\lim_{n\to\infty} \rho(n)V^*_{\rho(n),1}(y,\pi_2) = \bar{j}_1(\pi_2),$

(iii) $J_1(y,\pi_1,\pi_2) \leqslant \bar{j}_1(\pi_2) \qquad \forall \pi_1 \in \Pi_1^*$

Proof. First notice that, by the third inequality in Corollary 3.13 and Assumption 3.14,

$$0 \leqslant \limsup_{\rho\downarrow 0} \rho V^*_{\rho,1}(y',\pi_2) < \infty, \tag{3.4.10}$$

with $y' \in S$ the fixed state in (3.4.7).

Let $\bar{j}_1(\pi_2)$ be a limit point of $\rho V^*_{\rho,1}(y',\pi_2)$ as $\rho \downarrow 0$ and let $\rho(n) \downarrow 0$ be a subsequence such that

$$\rho(n)V^*_{\rho(n),1}(y',\pi_2) \longrightarrow \bar{j}_1(\pi_2). \tag{3.4.11}$$

Then, we are now in conditions to prove (i) to (ii); namely:

(i) From (3.4.9) we get that

$$|\rho(n)h_{\rho(n),1}(y,\pi_2)| \leqslant \rho(n)|b_1(y,\pi_2)| \longrightarrow 0 \quad \text{as } n \to \infty, \tag{3.4.12}$$

which proves (i).

(ii) By (3.4.11) and (3.4.7) we obtain

$$\begin{aligned}\lim_{\rho(n)\downarrow 0} \rho(n)V^*_{\rho(n),1}(y,\pi_2) &= \lim_{\rho(n)\downarrow 0}\left[\rho(n)V^*_{\rho(n),1}(y',\pi_2) + \rho(n)h_{\rho(n),1}(y,\pi_2)\right] \\ &= \bar{j}_1(\pi_2),\end{aligned}$$

yielding (ii).

(*iii*) Using Corollary 3.13(*ii*) we can deduce

$$\begin{aligned} J_1(y,\pi_1,\pi_2) &\leqslant \lim_{\rho(n)\downarrow 0} \rho(n) V_{\rho(n),1}(y,\pi_1,\pi_2) \\ &\leqslant \lim_{\rho(n)\downarrow 0} \rho(n) V^*_{\rho(n),1}(y,\pi_2) \\ &= \bar{j}_1(\pi_2) \qquad \forall \pi_1 \in \Pi_1^*, \end{aligned}$$

which is what we wanted. ■

Note that the previous lemma is true for every $\pi_2 \in \Pi_2$. The following theorem approximates the solution of the APOE equation for the first player. The second player case is analogous.

Before establishing the theorem, we remark that if we know in advance the existence of a Nash equilibrium for the ρ-discounted game, then

$$v_1^\rho(y) = V_{\rho,1}(y,\pi_1^*,\pi_2^*) \qquad \forall y \in S,$$

where $v_1^\rho(y)$ is as in Theorem 3.5 (time-invariant version). Thus,

$$\lim_{\rho(n)\downarrow 0} \rho(n) V^*_{\rho(n),1}(y,\pi_2^*) = \bar{j}_1(\pi_2^*)$$

becomes

$$\lim_{\rho(n)\downarrow 0} \rho(n) v_1(y) = \bar{j}_1(\pi_2^*).$$

Now, we just need a function $h_1 \in \mathbf{D}(S)$ that satisfies the APOE. In specific cases we can give hypotheses that ensure the existence of such a function h_1 (as in Chapter 4), but since we are working in a more general case we will just suppose such existence.

The following verification theorem uses the previous ideas.

***Theorem* 3.16.** *Fix $\pi_2 \in \Pi_2$. Suppose that there exists a function $b : S \times \Pi_2 \to \mathbb{R}$ as in Lemma 3.15. Additionally, suppose that there exist $h_1 \in \boldsymbol{D}(S)$ and $\overline{\pi_1} \in \Pi_1$ a stationary strategy satisfying,*

$$\bar{j}_1(\pi_2) \leqslant r_1^{\overline{\pi_1},\pi_2}(y) + \mathcal{L}^{\overline{\pi_1},\pi_2} h_1(y) \qquad \forall y \in S, \tag{3.4.13}$$

and

$$\lim_{t\to\infty} \frac{1}{t} T_t^{\overline{\pi_1},\pi_2} h_1(y) = 0. \tag{3.4.14}$$

Then, we have that

$$J_1(y,\overline{\pi_1},\pi_2) \geqslant \bar{j}_1(\pi_2) \geqslant J_1(y,\pi_1,\pi_2) \qquad \forall \pi_1 \in \Pi_1^*. \tag{3.4.15}$$

Proof. Using Lemma 3.15 we know that

$$\bar{j}_1(\pi_2) \geqslant J_1(y,\pi_1,\pi_2) \qquad \forall \pi_1 \in \Pi_1^*. \tag{3.4.16}$$

On the other hand, from (3.4.13) and Lemma 3.7(ii), we obtain

$$\bar{j}_1(\pi_2) \leqslant J_1(y, \overline{\pi_1}, \pi_2), \tag{3.4.17}$$

implying (3.4.15). ■

The same analysis can be done for the second player, with a function $b_2 : S \times \Pi_1 \to \mathbb{R}$ as in Lemma 3.15 (second player version). As a consequence we can deduce the following corollary from Theorem 3.16.

***Corollary* 3.17.** *Suppose that there exist functions $b_1 : S \times \Pi_2 \to \mathbb{R}$ and $b_2 : S \times \Pi_1 \to \mathbb{R}$ as in Lemma 3.15, two functions $h_1, h_2 \in \boldsymbol{D}(S)$, and a pair of stationary strategies $(\overline{\pi_1}, \overline{\pi_2}) \in \Pi_1^* \times \Pi_2^*$ that satisfy, for every $y \in S$,*

$$\bar{j}_1(\overline{\pi_2}) \leqslant r_1^{\overline{\pi_1}, \overline{\pi_2}}(y) + \mathcal{L}^{\overline{\pi_1}, \overline{\pi_2}} h_1(y) \tag{3.4.18}$$

and

$$\bar{j}_2(\overline{\pi_1}) \leqslant r_2^{\overline{\pi_1}, \overline{\pi_2}}(y) + \mathcal{L}^{\overline{\pi_1}, \overline{\pi_2}} h_2(y). \tag{3.4.19}$$

Also, the pair $(\overline{\pi_1}, \overline{\pi_2})$ satisfies

$$\lim_{t \to \infty} \frac{1}{t} T_t^{\overline{\pi_1}, \overline{\pi_2}} h_i(y) = 0 \qquad \text{for } i = 1, 2. \tag{3.4.20}$$

Then,

$$J_1(y, \overline{\pi_1}, \overline{\pi_2}) \geqslant J_1(y, \pi_1, \overline{\pi_2}) \qquad \forall \pi_1 \in \Pi_1^* \tag{3.4.21}$$

and

$$J_2(y, \overline{\pi_1}, \overline{\pi_2}) \geqslant J_1(y, \overline{\pi_1}, \pi_2) \qquad \forall \pi_2 \in \Pi_2^*. \tag{3.4.22}$$

In other words, $(\overline{\pi_1}, \overline{\pi_2})$ is a Nash equilibrium.

Chapter 4

A Special Case

4.1 Introduction

In this chapter we apply our previous results to the case when the Markov process evolves as a diffusion process. The aim here is to illustrate that the general theory introduced in the previous sections apply to this case; in particular, we shall show the existence of a non-cooperative equilibrium for the nonzero-sum case. This type of games are based under the assumption that the drift of the diffusion as well as the cost rate have an additive structure.

4.2 Stochastic differential game

Let $\mathtt{y}(\cdot)$ be an m-dimensional *diffusion process* that is controlled by two players. More explicitly, the process evolves according to the stochastic differential equation

$$d\mathtt{y}(t) = b(\mathtt{y}(t), a_1(t), a_2(t))dt + \sigma(\mathtt{y}(t))dW(t), \quad \mathtt{y}(0) = 0, \ t \geqslant 0, \tag{4.2.1}$$

where $b : \mathbb{R}^m \times A_1 \times A_2 \to \mathbb{R}^m$, $\sigma : \mathbb{R}^m \to \mathbb{R}^{m\times d}$ are given functions, so-named the *drift* and the *dispersion* matrix, respectively, and $W(\cdot)$ is a d-dimensional standard Brownian motion. The action sets are $A_1 \subset R^{m_1}$, $A_2 \subset \mathbb{R}^{m_2}$. Finally, $a_i(\cdot)$ is an A_i-valued stochastic process that gives the action of player i at each time $t \geqslant 0$, for each $i = 1, 2$.

Throughout this work we shall assume the following.

***Assumption* 4.1.** *(i) The drift coefficient $b(y, a_1, a_2)$ is continuous on $\mathbb{R}^m \times A_1 \times A_2$ and there exists a positive constant K_1 such that for each $x, y \in \mathbb{R}^m$,*

$$\sup_{(a_1,a_2)\in A_1\times A_2} |b(y, a_1, a_2) - b(x, a_1, a_2)| \leqslant K_1|y - x|. \tag{4.2.2}$$

(ii) There exist measurable functions $b_1 : \mathbb{R}^m \times A_1 \to \mathbb{R}^m$, $b_2 : \mathbb{R}^m \times A_2 \to \mathbb{R}^m$ such that the drift coefficient in (4.2.1) satisfies that

$$b(y, a_1, a_2) = b_1(y, a_1) + b_2(y, a_2), \tag{4.2.3}$$

in which b_1 and b_2 satisfy (i).

(iii) *There exists a constant $K_2 > 0$ such that, for every $x, y \in \mathbb{R}^m$,*

$$|\sigma(y) - \sigma(x)| \leqslant K_2|y - x|. \tag{4.2.4}$$

(iv) *The matrix $a(y) := \sigma(y)\sigma'(y)$ satisfies that, for some constant $K_3 > 0$,*

$$x'a(y)x \geqslant K_3|x|^2 \quad \forall x, y \in \mathbb{R}^m. \tag{4.2.5}$$

(v) *The actions sets A_1 and A_2 are compact.*

Statement (iv) in Assumption 4.1 is usually known as *uniform ellipticity*. Some normed spaces that we will need are defined below.

In the following definitions ∇g is the gradient of the function g, $\mathbb{H}v$ is the Hessian matrix of v, b_i is the ith component of b, a_{ij} is the (i,j)-component of matrix $a(\cdot)$ (defined in Assumption 4.1(iv)), and

$$D^\lambda h := \frac{\partial^{|\lambda|}h}{\partial x_1^{\lambda_1}, \ldots, \partial x_m^{\lambda_m}}, \quad \text{with} \quad \lambda = (\lambda_1, \ldots, \lambda_m), \quad |\lambda| := \sum_{i=1}^m \lambda_i. \tag{4.2.6}$$

***Definition* 4.2.** *For a fixed open set $\mathcal{O} \subset \mathbb{R}^m$ we define:*

i) *$\mathcal{W}^{l,p}(\mathcal{O})$ as the Sobolev space of all real-valued measurable functions $h : \mathcal{O} \to \mathbb{R}$ such that $D^\lambda h$:*

- *a)* *exists for every $|\lambda| \leqslant l$ in the weak sense,*
- *b)* *belongs to $L^p(\mathcal{O})$.*

ii) *$\mathcal{C}^k(\mathcal{O})$ as the space of all real-valued functions on $\mathcal{O}$ with continuous l-th partial derivatives in $x_i \in \mathbb{R}$, for $i = 1, \ldots, m$, $l = 0, 1, \ldots, k$. When $k = 0$, $\mathcal{C}^0(\mathcal{O})$ is the space of real-valued continuous functions on $\mathcal{O}$, which we simply denote by $\mathcal{C}(\mathcal{O})$.*

iii) *$\mathcal{C}^{k,\beta}(\mathcal{O})$ as de subspace of $\mathcal{C}^k(\mathcal{O})$ consisting of all functions h such that $D^\lambda h$ satisfies a Hölder condition with exponent $\beta \in (0,1]$, for all $|\lambda| \leqslant k$. In others words, that there exists a constant K such that*

$$|D^\lambda h(x) - D^\lambda h(y)| \leqslant K|x - y|^\beta \tag{4.2.7}$$

iv) *$\mathcal{C}_b(\mathcal{O} \times A_i)$, for $i = 1, 2$, as the space of all continuous bounded functions on $\mathcal{O} \times A_i$.*

The spaces in Definition 4.2 that will be frequently used throughout this work are: $\mathcal{C}^{0,\beta}(\mathcal{O})$, $\mathcal{C}^{1,\beta}(\mathcal{O})$ and $\mathcal{W}^{2,p}(\mathcal{O})$, endowed with the following norms.

***Definition* 4.3.** *(i) for* $f \in \mathcal{C}^{0,\beta}(\mathcal{O})$,

$$\|f\|_{\mathcal{C}^{0,\beta}(\mathcal{O})} := \sup_{x\in\mathcal{O}} |f(x)| + \sup_{x,y\in\mathcal{O},x\neq y} \frac{|f(x)-f(y)|}{|x-y|^\beta}, \tag{4.2.8}$$

(ii) for $g \in \mathcal{C}^{1,\beta}(\mathcal{O})$,

$$\begin{aligned} \|g\|_{\mathcal{C}^{1,\beta}(\mathcal{O})} := &\max\left\{\sup_{x\in\mathcal{O}} |g(x)|, \sup_{x\in\mathcal{O}} |\nabla g(x)|\right\} \\ &+ \max\left\{\sup_{x,y\in\mathcal{O},x\neq y} \frac{|g(x)-g(y)|}{|x-y|^\beta}, \sup_{x,y\in\mathcal{O},x\neq y} \frac{|\nabla g(x)-\nabla g(y)|}{|x-y|^\beta}\right\}, \end{aligned} \tag{4.2.9}$$

(iii) for $h \in \mathcal{W}^{2,p}(\mathcal{O})$,

$$\|h\|_{\mathcal{W}^{2,p}(\mathcal{O})} := \left(\int_{\mathcal{O}} \left[|h(x)|^p + \sum_{i=1}^m |\partial_{x_i} h(x)|^p + \sum_{i,j=1}^m |\partial^2_{x_i x_j} h(x)|^p\right] dx\right)^{1/p}. \tag{4.2.10}$$

Let $\mathcal{O} \equiv \mathbb{R}^m$. For a fixed pair $(a_1,a_2) \in A_1 \times A_2$ and $v \in \mathcal{W}^{2,p}(\mathbb{R}^m)$, $p \geqslant 1$, define

$$\begin{aligned} \mathcal{L}^{a_1,a_2} v(y) :=& \langle \nabla v, b(y,a_1,a_2)\rangle + Tr[(\mathbb{H}v)a](y) \\ =& \sum_{i=1}^m b_i(y,a_1,a_2)\partial_i v(y) + \frac{1}{2}\sum_{i,j=1}^m a_{ij}(y)\partial^2_{ij} v(y), \end{aligned} \tag{4.2.11}$$

When players use the randomized strategies $(\pi_1,\pi_2) \in \Pi_1 \times \Pi_2$, the drift coefficient b in (4.2.1) is such that

$$b_i(y,\pi_i) := \int_{A_i} b(y,a_i)\pi_i(da_i|y), \quad i=1,2. \tag{4.2.12}$$

Furthermore, from (4.2.11), we have that for each $h \in \mathcal{W}^{2,p}(\mathbb{R}^m)$, $p \geqslant 1$,

$$\mathcal{L}^{\pi_1,\pi_2} h(y) := \int_{A_1}\int_{A_2} \mathcal{L}^{a_1,a_2} h(y)\pi_2(da_2|y)\pi_1(da_1|y). \tag{4.2.13}$$

***Remark* 4.4.** *Assumption 4.1 ensures that, for each pair of strategies* $(\pi_1,\pi_2) \in \Pi_1 \times \Pi_2$*, there exists an almost surely unique strong solution of (4.2.1) which is a Markov-Feller process (see [2], Theorem 2.2.12).*

Let us denote by $P^{\pi_1,\pi_2}(t,y,\cdot)$ the corresponding transition probability of the process $\mathrm{y}^{\pi_1,\pi_2}(\cdot)$, and recall that $E_y^{\pi_1,\pi_2}[\cdot]$ is its corresponding expectation.

***Remark* 4.5.** *By Theorem 4.3 in [1], for each pair* $(\pi_1,\pi_2) \in \Pi_1 \times \Pi_2$ *the probability measure* $P^{\pi_1,\pi_2}(t,y,\cdot)$ *is absolutely continuous with respect to the Lebesgue measure* $\lambda(\cdot)$*, for every* $y \in \mathbb{R}^m$*, and* $t \geqslant 0$*. Then, there exist a transition density function* $p^{\pi_1,\pi_2}(t,y,z) \geqslant 0$ *such that*

$$P^{\pi_1,\pi_2}(t,y,C) = \int_C p^{\pi_1,\pi_2}(t,y,z)dz, \tag{4.2.14}$$

for every $C \in \mathcal{B}(\mathbb{R}^m)$.

Using Assumption 4.1 and Theorem 7.3.8 in [4], we get that the transition density function $p^{\pi_1,\pi_2}(t,y,z)$ is strictly positive and according to the construction of *fundamental solutions* in Section 6.4 of [7], for each $t > 0$, $p^{a_1,a_2}(t,y,z)$ is continuous in $y,z \in \mathbb{R}^m$ and continuous in $(a_1,a_2) \in A_1 \times A_2$.

To talk about the topology of the strategies sets we need to define the following convergence notion.

Recall the definition of (randomized) Markov strategies Π_1, Π_2 defined in Section 2.2.

***Definition* 4.6.** *A sequence $\{\pi_1^n\} \subset \Pi_1$ converges to $\pi_1 \in \Pi_1$, denoted $\pi_1^n \xrightarrow{W} \pi_1$, if and only if*

$$\int_{\mathbb{R}^m} g(y) \int_{A_1} h(y,a_1)\pi_1^n(da_1|y)dy \longrightarrow \int_{\mathbb{R}^m} g(y) \int_{A_1} h(y,a_1)\pi_1(da_1|y)dy, \quad (4.2.15)$$

for every $g \in L^1(\mathbb{R}^m)$ and $h \in \mathcal{C}_b(\mathbb{R}^m \times A_1)$. Convergence in Π_2 is defined similarly.

An important remark is that under this notion of convergence, both Π_1 and Π_2 are compact sets (see [2], Section 2.4).

Now we present some assumptions about the ergodicity of our system.

***Assumption* 4.7.** *There exist a function $w \in \mathcal{C}^2(\mathbb{R}^m)$ $w \geqslant 1$, and constants $d \geqslant c > 0$ that satisfy the following:*

(i) $\lim_{|y|\to\infty} w(y) = +\infty$.

(ii) $\mathcal{L}^{\pi_1,\pi_2} w(y) \leqslant -cw(y) + d, \qquad (\pi_1,\pi_2) \in \Pi_1 \times \Pi_2,\ y \in \mathbb{R}^m.$

Given Assumption 4.7, for each pair $(\pi_1,\pi_2) \in \Pi_1 \times \Pi_2$, the Markov process $\mathrm{y}^{\pi_1,\pi_2}(\cdot)$ has a unique invariant probability measure μ_{π_1,π_2} for which

$$\mu^{\pi_1,\pi_2}(w) := \int_{\mathbb{R}^m} w(y)\mu_{\pi_1,\pi_2}(dy) < \infty, \quad (4.2.16)$$

(see [2] for more details). Furthermore, applying Dynkin's formula to the function $v(t,y) := e^{ct}w(y)$ and using Assumption 4.7(ii) we obtain

$$\mu^{\pi_1,\pi_2}(w) \leqslant \frac{d}{c} \quad (4.2.17)$$

and

$$E_y^{\pi_1,\pi_2}\left[w(\mathrm{y}(t))\right] \leqslant e^{-ct}w(y) + \frac{d}{c}(1 - e^{-ct}), \quad (4.2.18)$$

for every $(\pi_1,\pi_2) \in \Pi_1 \times \Pi_2$, $y \in \mathbb{R}^m$ and $t \geqslant 0$.

In order to ensure a type of exponential ergodicity, we shall assume the following.

***Assumption* 4.8.** *There exist $\rho > 0$, $T > 0$, and a bounded open set $\mathcal{O}$ for which the transition density function in (4.2.14) satisfies*

$$p^{\pi_1,\pi_2}(T,y,z) \geqslant \rho \qquad \forall y,z \in \overline{\mathcal{O}}, (\pi_1,\pi_2) \in \Pi_1 \times \Pi_2, \tag{4.2.19}$$

($\overline{\mathcal{O}}$ denotes the closure of $\mathcal{O}$) and T satisfies that

$$\frac{d}{c}(1 - e^{-cT}) \leqslant \rho\lambda(\mathcal{O}). \tag{4.2.20}$$

***Definition* 4.9.** *Consider an open set $\mathcal{O} \subset \mathbb{R}^m$. $\mathbb{B}_w(\mathcal{O})$ denotes the Banach space of al real-valued measurable functions v on $\mathcal{O}$ with finite w-norm defined as*

$$\|v\|_w = \sup_{y \in \mathcal{O}} \frac{|v(y)|}{w(y)}. \tag{4.2.21}$$

***Remark* 4.10.** *Under (4.2.18), it is easy to see that the transversality property condition (3.2.26) (see also (3.2.30), (3.3.4), (3.3.7), (3.3.12) and (3.3.24)) holds for every pair of strategies $(\pi_1,\pi_2) \in \Pi_1 \times \Pi_2$ and every function $h \in \mathbb{B}_w(\mathbb{R}^m)$. Indeed,*

$$\begin{aligned}\left|e^{-\rho t} E_y^{\pi_1,\pi_2}\left[h(\mathrm{y}(t))\right]\right| &\leqslant e^{-\rho t} E_y^{\pi_1,\pi_2}\left[\|h\|_w w(\mathrm{y}(t))\right] \\ &\leqslant e^{-\rho t}\|h\|_w E_y^{\pi_1,\pi_2}\left[w(\mathrm{y}(t))\right] \\ &\leqslant e^{-\rho t}\|h\|_w \left(e^{-ct} w(y) + \frac{d}{c}(1 - e^{-ct})\right).\end{aligned}$$

Hence, taking limit as $t \to \infty$ we obtain (3.2.26). Note that the same argument works to show (3.3.4) if we replace $e^{-\rho t}$ by $1/t$.

From Assumptions 4.1, 4.7 and 4.8 we can deduce that for every $(\pi_1,\pi_2) \in \Pi_1 \times \Pi_2$, the process $\mathrm{y}^{\pi_1,\pi_2}(\cdot)$ is *w-exponentially ergodic*; that is, there exist $c > 0$ and $\delta > 0$ such that

$$\sup_{(\pi_1,\pi_2) \in \Pi_1 \times \Pi_2} \left|E_y^{\pi_1,\pi_2}\left[v(\mathrm{y}(t)\right] - \mu^{\pi_1,\pi_2}(v)\right| \leqslant c e^{-\delta t}\|v\|_w w(y), \tag{4.2.22}$$

for every $y \in \mathbb{R}^m$, $t \geqslant 0$ and $v \in \mathbb{B}_w(\mathbb{R}^m)$, where μ^{π_1,π_2} is as in (4.2.16). For details see [13] ,Theorem 2.7.

Now we impose some assumptions on the reward rate of the game.

For each $i = 1, 2$, the reward rate function is a real-valued function on $\mathbb{R}^m \times A_1 \times A_2$ that satisfies the following.

***Assumption* 4.11.** *For each $i = 1, 2$:*

(a) There exist functions $r_{i1} : \mathbb{R}^m \times A_1 \to \mathbb{R}$ and $r_{i2} : \mathbb{R}^m \times A_2 \to \mathbb{R}$ such that

$$r_i^{a_1,a_2}(y) = r_{i1}^{a_1}(y) + r_{i2}^{a_2}(y). \tag{4.2.23}$$

(b) For each j ($j = 1, 2$), the function $r_{ij}^{a_j}(y)$ is continuous in $a_j \in A_j$, and for each $y \in \mathbb{R}^m$ there exist a neighbourhood of y, denoted by $\mathcal{O}_y$, and a constant $K_{ij}(y) > 0$ such that, for every $z \in \mathcal{O}_y$,

$$\sup_{a_j \in A_j} |r_{ij}^{a_j}(y) - r_{ij}^{a_j}(z)| \leqslant K_{ij}(y)|y - z|. \tag{4.2.24}$$

(c) There exist a constant $M > 0$ such that, for every $y \in \mathbb{R}^m$,

$$\sup_{a_1 \in A_1} |r_{i1}^{a_1}(y)| \leqslant M w(y) \tag{4.2.25}$$

and

$$\sup_{a_2 \in A_2} |r_{i2}^{a_2}(y)| \leqslant M w(y). \tag{4.2.26}$$

That is, $r_{i1}^{a_1}(\cdot)$ and $r_{i2}^{a_2}(\cdot)$ are in $\mathbb{B}_w(\mathbb{R}^m)$ uniformly in a_1 and a_2 respectively.

As in previous comments, when the players use randomized strategies $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$, the payoff rate is defined as

$$r_i^{\pi_1,\pi_2}(y) := \int_{A_1} \int_{A_2} r_i^{a_1,a_2}(y) \pi_2(da_2|y) \pi_1(da_1|y) \tag{4.2.27}$$

$$= \int_{A_1} r_{i1}^{a_1}(y) \pi_1(da_1|y) + \int_{A_2} r_{i2}^{a_2}(y) \pi_2(da_2|y) \tag{4.2.28}$$

$$=: r_{i1}^{\pi_1}(y) + r_{i2}^{\pi_2}(y). \tag{4.2.29}$$

4.3 Discounted case, infinite horizon

The payoff function, for each player i ($i = 1, 2$), is given as in (3.2.23) in the time-homogeneous case, so according to (4.2.27)-(4.2.29) it can be rewritten as

$$V_i^\rho(y, \pi_1, \pi_2) = E_y^{\pi_1,\pi_2}\left[\int_0^\infty e^{-\rho t} r_{i1}^{\pi_1}(\mathtt{y}(t))dt\right] + E_y^{\pi_1,\pi_2}\left[\int_0^\infty e^{-\rho t} r_{i2}^{\pi_2}(\mathtt{y}(t))dt\right]. \tag{4.3.1}$$

Now, for a given strategy $\pi_2 \in \Pi_2$, consider the functions

$$v_{1,\pi_2}^\rho(y) := \sup_{\pi_1 \in \Pi_1} V_1^\rho(y, \pi_1, \pi_2). \tag{4.3.2}$$

Similarly, for a fixed $\pi_1 \in \Pi_1$, consider

$$v_{2,\pi_1}^\rho(y) := \sup_{\pi_2 \in \Pi_2} V_2^\rho(y, \pi_1, \pi_2). \tag{4.3.3}$$

Under Assumptions 4.7 and 4.11(c) we obtain that, for each $i = 1, 2$,

$$\sup_{(\pi_1,\pi_2) \in \Pi_1 \times \Pi_2} |V_i^\rho(y, \pi_1, \pi_2)| \leqslant 2M \frac{\rho + d}{\rho c} w(y), \tag{4.3.4}$$

which implies that $V_i^\rho(\cdot, \pi_1, \pi_2) \in \mathbb{B}_w(\mathbb{R}^m)$ for each $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$, $i = 1, 2$.

Note that (4.3.2)-(4.3.3) imply that the functions v_{1,π_2}^ρ and v_{2,π_1}^ρ also belong to $\mathbb{B}_w(\mathbb{R}^m)$.

***Definition* 4.12.** *Let $\pi_2 \in \Pi_2$ be a given strategy. A strategy $\pi_1^* \in \Pi_1$ is a ρ-**discounted optimal response**, to π_2, from player 1 if*

$$V_1^\rho(y, \pi_1^*, \pi_2) = v_{1,\pi_2}^\rho(y). \tag{4.3.5}$$

*Analogously, given $\pi_1 \in \Pi_1$, a strategy $\pi_2^* \in \Pi_2$ is a ρ-**discounted optimal response** from player 2 if*

$$V_2^\rho(y, \pi_1, \pi_2^*) = v_{2,\pi_1}^\rho(y). \tag{4.3.6}$$

***Proposition* 4.13.** *Under Assumptions 4.1, 4.7 and 4.11 the following is satisfied:*

(i) For each fixed $\pi_2 \in \Pi_2$ there exists a function $v_1 \in \mathcal{W}^{2,p}(\mathbb{R}^m) \cap \mathbb{B}_w(\mathbb{R}^m)$ $(p > m)$ and a strategy $\delta_{f_1}^* \in \Pi_1$ ($f_1 \in \mathbb{F}_1$ may depend on π_2) such that equations (3.2.8)-(3.2.9) hold.

(ii) $v_1(y) = v_{1,\pi_2}^\rho(y)$ for every $y \in \mathbb{R}^m$.

(iii) A strategy $\pi_1^* \in \Pi_1$ is an optimal response from player 1, to the strategy $\pi_2 \in \Pi_2$, if and only if (3.2.8)-(3.2.9) are satisfied.

(iv) Analogously, for each fixed strategy $\pi_1 \in \Pi_1$, there exist a function $v_2 \in \mathcal{W}^{2,p}(\mathbb{R}^m) \cap \mathbb{B}_w(\mathbb{R}^m)$ $(p > m)$ and a strategy $\delta_{f_2}^* \in \Pi_2$ ($f_2 \in \mathbb{F}_2$) such that (3.2.10)-(3.2.11) hold.

(v) $v_2(y) = v_{2,\pi_1}^\rho(y)$ for every $y \in \mathbb{R}^m$.

(vi) A strategy $\pi_2^* \in \Pi_2$ is an optimal response of player 2 if and only if (3.2.10)-(3.2.11) are satisfied.

Proof. Parts (i) and (iv) follow from [14, Proposition 3.6], whereas the rest of the statements are special cases of the proof of Theorem 3.2 ■

The following proposition helps to ensure the existence of a Nash equilibrium for the ρ-discounted game. Its proof uses the weak convergence given in Definition 4.6 and a technical result on interchanging limits with respect to the generators of type (4.2.11). The elements needed for the proof are outside of the scope of this work, so for more details we refer the reader to reference [14].

***Proposition* 4.14.** *Suppose that Assumptions 4.1, 4.7 and 4.11 are satisfied. Let $\{v_1^n\}$ and $\{v_2^n\}$ be sequences of functions in $\mathcal{W}^{2,p}(\mathbb{R}^m) \cap \mathbb{B}_w(\mathbb{R}^m)$ $(p > m)$ and $\{(\pi_1^n, \pi_2^n)\}$ a sequence in $\Pi_1 \times \Pi_2$.*

(a) If v_1^n satisfies, for every $n \geqslant 1$ and $y \in \mathbb{R}^n$,

$$\begin{aligned}\rho v_1^n(y) &= r_1^{\pi_1^n, \pi_2^n}(y) + \mathcal{L}^{\pi_1^n, \pi_2^n} v_1^n(y) \\ &= \sup_{\pi_1 \in \Pi_1} \left\{ r_1^{\pi_1, \pi_2^n}(y) + \mathcal{L}^{\pi_1, \pi_2^n} v_1^n(y) \right\},\end{aligned}$$

and $(\pi_1^n, \pi_2^n) \xrightarrow{W} (\pi_1, \pi_2)$*, then there exists a function* $v_1 \in \mathcal{W}^{2,p}(\mathbb{R}^m) \cap \mathbb{B}_w(\mathbb{R}^m)$ *such that* $v_1^n \to v_1$ *uniformly and*

$$\begin{aligned} \rho v_1(y) &= r_1^{\pi_1,\pi_2}(y) + \mathcal{L}^{\pi_1,\pi_2} v_1(y) \\ &= \sup_{\pi_1 \in \Pi_1} \{ r_1^{\pi_1,\pi_2}(y) + \mathcal{L}^{\pi_1,\pi_2} v_1(y) \}. \end{aligned}$$

(b) *If* v_2^n *is such that, for every* $n \geqslant 1$ *and* $y \in \mathbb{R}^m$,

$$\begin{aligned} \rho v_2^n(y) &= r_2^{\pi_1^n,\pi_2^n}(y) + \mathcal{L}^{\pi_1^n,\pi_2^n} v_2^n(y) \\ &= \sup_{\pi_2 \in \Pi_2} \{ r_2^{\pi_1^n,\pi_2}(y) + \mathcal{L}^{\pi_1^n,\pi_2} v_2^n(y) \}, \end{aligned}$$

and $(\pi_1^n, \pi_2^n) \xrightarrow{W} (\pi_1, \pi_2)$*, then there exists a function* $v_2 \in \mathcal{W}^{2,p}(\mathbb{R}^m) \cap \mathbb{B}_w(\mathbb{R}^m)$ $(p > m)$ *such that* $\{v_2^n\}$ *converges uniformly to* v_2 *and*

$$\begin{aligned} \rho v_2(y) &= r_2^{\pi_1,\pi_2}(y) + \mathcal{L}^{\pi_1,\pi_2} v_2(y) \\ &= \sup_{\pi_2 \in \Pi_2} \{ r_2^{\pi_1,\pi_2}(y) + \mathcal{L}^{\pi_1,\pi_2} v_2(y) \}. \end{aligned}$$

Now, we define a pair of sets that are useful to prove the existence of a Nash equilibrium in this type of games.

For a fixed $\pi_2 \in \Pi_2$ and the function v_{1,π_2}^ρ as in (4.3.5), define

$$S_1(\pi_2) := \{ \pi_1 \in \Pi_1 \mid \rho v_{1,\pi_2}^\rho(y) = r_1^{\pi_1,\pi_2}(y) + \mathcal{L}^{\pi_1,\pi_2} v_{1,\pi_2}^\rho(y) \}. \tag{4.3.7}$$

In a similar way, for a fixed $\pi_1 \in \Pi_1$ and v_{2,π_1}^ρ as in (4.3.6), let

$$S_2(\pi_1) := \{ \pi_2 \in \Pi_2 \mid \rho v_{2,\pi_1}^\rho(y) = r_2^{\pi_1,\pi_2}(y) + \mathcal{L}^{\pi_1,\pi_2} v_{2,\pi_1}^\rho(y) \}. \tag{4.3.8}$$

These sets will be called *sets of optimal responses*. The following proposition shows important properties of the sets of optimal responses, $S_1(\pi_2)$ and $S_2(\pi_1)$.

***Proposition* 4.15.** *Under Assumptions 4.1, 4.7 and 4.11, the sets of optimal responses are convex compact sets for each* $\pi_1 \in \Pi_1$ *and* $\pi_2 \in \Pi_2$.

Proof. First, let us show that $S_1(\pi_2)$ and $S_2(\pi_1)$ are closed sets. Let $\{\pi_1^{n*}\}_{n=1}^\infty$ be a sequence in $S_1(\pi_2)$ such that $\pi_1^{n*} \xrightarrow{W} \pi_1^*$. From 4.3.7 we have that, for each $n \geqslant 1$,

$$v_{1,\pi_2}^\rho(y) = r_1^{\pi_1^{n*},\pi_2}(y) + \mathcal{L}^{\pi_1^{n*},\pi_2} v_{1,\pi_2}^\rho(y).$$

Letting $n \to \infty$ and using Proposition 4.14 yields

$$v_{1,\pi_2}^\rho(y) = r_1^{\pi_1^*,\pi_2}(y) + \mathcal{L}^{\pi_1^*,\pi_2} v_{1,\pi_2}^\rho(y), \tag{4.3.9}$$

which implies that π_1^* belongs to $S_1(\pi_2)$, that is, $S_1(\pi_2)$ is closed. A similar argument gives that $S_2(\pi_1)$ is closed. Now, since Π_1 and Π_2 are compact sets, then $S_1(\pi_2)$ and $S_2(\pi_1)$ are compact sets. Finally, from [16, Lemma 4.6] we get that both sets $S_1(\pi_2)$ and $S_2(\pi_1)$ are convex. ■

Finally, using all previous assumptions and results, we can present the following result.

***Theorem* 4.16.** *Suppose that Assumptions 4.1, 4.7 and 4.11 hold. Then there exist a Nash equilibrium* $(\pi_1^*, \pi_2^*) \in \Pi_1 \times \Pi_2$ *for the* ρ*-discounted payoff game.*

Proof. From Proposition 4.13 we obtain that the sets $S_1(\pi_2)$ and $S_2(\pi_1)$ are non-empty for each $\pi_2 \in \Pi_2$ and $\pi_1 \in \Pi_1$, respectively. Define the multifunction $F : \Pi_1 \times \Pi_2 \longleftrightarrow 2^{\Pi_1 \times \Pi_2}$ as

$$F(\pi_1, \pi_2) := S_1(\pi_2) \times S_2(\pi_1). \tag{4.3.10}$$

We show that there exist a fixed point of F.

Let $(\pi_1', \pi_2') \in \Pi_1 \times \Pi_2$ be an arbitrary fixed pair of strategies. Using Proposition 4.13 we obtain that there exist a pair of strategies $(\delta^*_{f_1^1}, \delta^*_{f_2^1}) \in \Pi_1 \times \Pi_2$ that belongs to $F(\pi_1', \pi_2')$. Inductively , we can get two sequences of strategies $\{\delta^*_{f_1^n}\}_{n \geqslant 1}$ and $\{\delta^*_{f_2^n}\}_{n \geqslant 1}$ such that $(\delta^*_{f_1^n}, \delta^*_{f_2^n}) \in F(\delta^*_{f_1^{n-1}}, \delta^*_{f_2^{n-1}})$; that is, for every $y \in \mathbb{R}^m$,

$$\rho v_1^{n-1}(y) = r_1^{\delta^*_{f_1^n}, \delta^*_{f_2^{n-1}}}(y) + \mathcal{L}^{\delta^*_{f_1^n}, \delta^*_{f_2^{n-1}}} v_1^{n-1}(y) \tag{4.3.11}$$

$$= \sup_{\pi_1 \in \Pi_1} \left\{ r_1^{\pi_1, \delta^*_{f_2^{n-1}}}(y) + \mathcal{L}^{\pi_1, \delta^*_{f_2^{n-1}}} v_1^{n-1}(y) \right\} \tag{4.3.12}$$

and

$$\rho v_2^{n-1}(y) = r_2^{\delta^*_{f_1^{n-1}}, \delta^*_{f_2^n}}(y) + \mathcal{L}^{\delta^*_{f_1^{n-1}}, \delta^*_{f_2^n}} v_2^{n-1}(y) \tag{4.3.13}$$

$$= \sup_{\pi_2 \in \Pi_2} \left\{ r_2^{\delta^*_{f_1^{n-1}}, \pi_2}(y) + \mathcal{L}^{\delta^*_{f_1^{n-1}}, \pi_2} v_2^{n-1}(y) \right\} \tag{4.3.14}$$

Since $\Pi_1 \times \Pi_2$ is compact, there exist a subsequence $(\delta^*_{f_1^{n_k}}, \delta^*_{f_2^{n_k}}) \equiv (\delta^*_{f_1^n}, \delta^*_{f_2^n})$ such that $(\delta^*_{f_1^n}, \delta^*_{f_2^n}) \xrightarrow{W} (\pi_1^*, \pi_2^*)$ for some $(\pi_1^*, \pi_2^*) \in \Pi_1 \times \Pi_2$. Taking limit as $n \to \infty$ in equations (4.3.11)-(4.3.14), Proposition 4.14 allows us to show the existence of two functions v_1 and v_2, both belonging to $\mathcal{W}^{2,p}(\mathbb{R}^m) \cap \mathbb{B}_w(\mathbb{R}^m)$, that satisfy

$$\begin{aligned} \rho v_1(y) &= r_1^{\pi_1^*, \pi_2^*}(y) + \mathcal{L}^{\pi_1^*, \pi_2^*} v_1(y) \qquad \forall y \in \mathbb{R}^m \\ &= \sup_{\pi_1 \in \Pi_1} \left\{ r_1^{\pi_1, \pi_2^*}(y) + \mathcal{L}^{\pi_1, \pi_2^*} v_1(y) \right\} \end{aligned}$$

and

$$\begin{aligned} \rho v_2(y) &= r_2^{\pi_1^*, \pi_2^*}(y) + \mathcal{L}^{\pi_1^*, \pi_2^*} v_2(y) \qquad \forall y \in \mathbb{R}^m \\ &= \sup_{\pi_2 \in \Pi_2} \left\{ r_2^{\pi_1^*, \pi_2}(y) + \mathcal{L}^{\pi_1^*, \pi_2} v_2(y) \right\}. \end{aligned}$$

Finally, from Theorem 3.5 we obtain that (π_1^*, π_2^*) is a Nash equilibrium in $\Pi_1 \times \Pi_2$, with v_1 and v_2 the value function of player 1 and player 2 respectively. ∎

4.4 Average case

In this case, we have that the payoff function for each player i ($i = 1, 2$) is defined as in (3.3.2), which is sometimes called *ergodic payoff*.

Given a pair $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$ we define the constant

$$J_i(\pi_1, \pi_2) := \mu^{\pi_1,\pi_2}(r_i(\cdot, \pi_1, \pi_2)) = \int_{\mathbb{R}^m} r_i(y, \pi_1, \pi_2)\mu_{\pi_1,\pi_2}(dy), \tag{4.4.1}$$

in which μ^{π_1,π_2} is the function in (4.2.16). Under Assumption 4.11(c) and (4.2.17) we obtain that, for every $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$,

$$|J_i(\pi_1, \pi_2)| \leqslant \int_{\mathbb{R}^m} |r_i(y, \pi_1, \pi_2)|\mu_{\pi_1,\pi_2}(dy) \leqslant 2M\mu^{\pi_1,\pi_2}(w) \leqslant 2M\frac{d}{c}, \tag{4.4.2}$$

which means that $J_i(\pi_1, \pi_2)$ is uniformly bounded in $\Pi_1 \times \Pi_2$.

***Remark* 4.17.** *Notice that under Assumptions 4.1, 4.7, 4.8 and 4.11 the average payoff function (3.3.2) coincides with the constant $J_i(\pi_1, \pi_2)$ in (4.4.1). Rewriting (3.3.1) as*

$$J^i_\tau(y, \pi_1, \pi_2) = \tau J_i(\pi_1, \pi_2) + \int_0^\tau \left[E_y^{\pi_1,\pi_2} r_i^{\pi_1,\pi_2}(\mathtt{y}(t)) - J_i(\pi_1, \pi_2)\right] dt, \tag{4.4.3}$$

then multiplying both sides of the equation by $1/\tau$ and letting $\tau \to \infty$, using (4.2.22) we obtain

$$J_i(y, \pi_1, \pi_2) = \liminf_{t\to\infty} \frac{1}{\tau} J^i_\tau(y, \pi_1, \pi_2) = J_i(\pi_1, \pi_2) \qquad \forall y \in \mathbb{R}^m. \tag{4.4.4}$$

From this fact, we can simply refer to (3.3.2) as $J_i(\pi_1, \pi_2)$.

4.4.1 Vanishing discount technique

In this section we present results that prove the existence of a Nash equilibrium for the *long-run average payoff* (3.3.2) using the vanishing discount technique as we saw in Section 3.4

Define, for every $y \in \mathbb{R}^m$,

$$h^\rho_{1,\pi_2}(y) := v^\rho_{1,\pi_2}(y) - v^\rho_{1,\pi_2}(0) \tag{4.4.5}$$

and

$$h^\rho_{2,\pi_1}(y) := v^\rho_{2,\pi_1}(y) - v^\rho_{2,\pi_1}(0). \tag{4.4.6}$$

In this case, the fixed state is $0 \in \mathbb{R}^m$.

The following result is a special version of Lemma 3.15. For a proof see Proposition 4.4 in [14].

***Proposition* 4.18.** *Let $R > 0$ and $\rho > 0$ be given numbers. Suppose that Assumptions 4.1, 4.7, 4.8 and 4.11 hold. Let $\{\rho_n\} \subset \mathbb{R}$ be a sequence of positive numbers such that $\rho_n \downarrow 0$ and a sequence $\{(\pi_1^{\rho_n}, \pi_2^{\rho_n})\} \subset \Pi_1 \times \Pi_2$ such that $(\pi_1^{\rho_n}, \pi_2^{\rho_n})$ is a Nash equilibrium for the ρ_n-discounted payoff game. Also, suppose that a pair of strategies $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$ is such that $(\pi_1^{\rho_n}, \pi_2^{\rho_n}) \xrightarrow{W} (\pi_1, \pi_2)$ as $n \to \infty$. Then:*

(a) There exist a constant g_1 (depending on π_2) and a subsequence $\{\rho_n^1\} \subset \{\rho_n\}$ such that

$$\rho_n^1 v_{1,\pi_2^{\rho_n^1}}^{\rho_n^1}(0) \longrightarrow g_1 \quad \text{as } \rho_n^1 \to 0.$$

(b) For some constant $K_3 > 0$, the function $h_{1,\pi_2^\rho}^\rho$ is such that

$$\|h_{1,\pi_2^\rho}^\rho\|_{\mathcal{W}^{2,p}(B_R)} \leqslant K_3$$

holds, where $B_R := \{y \in \mathbb{R}^m : |y| < R\}$.

(c) From (a) and (b) we obtain

$$\|\rho_1^n h_{1,\pi_2^{\rho_1^n}}^{\rho_1^n}\|_{\mathcal{W}^{2,p}(B_R)} \longrightarrow 0 \qquad \text{as } \rho_n^1 \to 0.$$

For the second player we have the following.

(d) There exist $g_2 \in \mathbb{R}$ (depending on π_1) and a subsequence $\{\rho_n^2\} \subset \{\rho_n\}$ such that

$$\rho_n^2 v_{2,\pi_1}^{\rho_n^2}(0) \longrightarrow g_2 \qquad \text{as } \rho_n^2 \to 0.$$

(e) For some constant $K_4 > 0$, the function h_{2,π_1}^ρ satisfies

$$\|h_{2,\pi_1^\rho}^\rho\|_{\mathcal{W}^{2,p}(B_R)} \leqslant K_4.$$

(f) Using (d) and (e) we obtain

$$\|\rho_n^2 h_{2,\pi_1^{\rho_n^2}}^{\rho_n^2}\|_{\mathcal{W}^{2,p}(B_R)} \longrightarrow 0 \qquad \text{as } \rho_n^2 \to 0.$$

Finally, the last important result of this section says as follows.

***Theorem* 4.19.** *Suppose that Assumptions 4.1, 4.7, 4.8 and 4.11 are fulfilled. Let $\{\rho_n\}_{n=1}^\infty \subset \mathbb{R}$ be a sequence of positive numbers such that $\rho_n \downarrow 0$, let $(\pi_1^{\rho_n}, \pi_2^{\rho_n}) \in \Pi_1 \times \Pi_2$ be a Nash equilibrium for the ρ_n-discounted game with the payoff function (4.3.1). If $\pi_1^{\rho_n} \xrightarrow{W} \pi_1^*$ and $\pi_2^{\rho_n} \xrightarrow{W} \pi_1^*$ as $\rho_n \downarrow 0$ then (π_1^*, π_2^*) is a Nash equilibrium for the game with ergodic payoff function given by (3.3.2).*

Proof. Let $(\pi_1^{\rho_n}, \pi_2^{\rho_n})$ be a Nash equilibrium for the ρ_n-discounted case. Since $\pi_1^{\rho_n}$ is an ρ_n-discounted optimal response to $\pi_2^{\rho_n}$, which means that $\pi_1^{\rho_n} \in S_1(\pi_2^{\rho_n})$ (analogously $\pi_2^{\rho_n}$ belongs to $S_2(\pi_1^{\rho_n})$), using Proposition 4.13 we obtain that the pair $(\pi_1^{\rho_n}, \pi_2^{\rho_n})$ satisfies,

$$\begin{aligned}\rho_n v_{1,\pi_2^{\rho_n}}^{\rho_n}(y) &= r_1^{\pi_1^{\rho_n},\pi_2^{\rho_n}}(y) + \mathcal{L}^{\pi_1^{\rho_n},\pi_2^{\rho_n}} v_{1,\pi_2^{\rho_n}}^{\rho_n}(y)\\ &= \sup_{\pi_1\in\Pi_1}\left\{r_1^{\pi_1,\pi_2^{\rho_n}}(y) + \mathcal{L}^{\pi_1,\pi_2^{\rho_n}} v_{1,\pi_2^{\rho_n}}^{\rho_n}(y)\right\}\end{aligned}$$

and

$$\begin{aligned}\rho_n v_{2,\pi_1^{\rho_n}}^{\rho_n}(y) &= r_2^{\pi_1^{\rho_n},\pi_2^{\rho_n}}(y) + \mathcal{L}^{\pi_1^{\rho_n},\pi_2^{\rho_n}} v_{2,\pi_1^{\rho_n}}^{\rho_n}(y)\\ &= \sup_{\pi_2\in\Pi_2}\left\{r_2^{\pi_1^{\rho_n},\pi_2}(y) + \mathcal{L}^{\pi_1^{\rho_n},\pi_2} v_{2,\pi_1^{\rho_n}}^{\rho_n}(y)\right\}.\end{aligned}$$

Replacing $h_{1,\pi_2^{\rho_n}}^{\rho_n}$ and $h_{2,\pi_1^{\rho_n}}^{\rho_n}$ from (4.4.5) and (4.4.6) in the previous equations we obtain,

$$\begin{aligned}\rho_n v_{1,\pi_2^{\rho_n}}^{\rho_n}(0) &= r_1^{\pi_1^{\rho_n},\pi_2^{\rho_n}}(y) + \mathcal{L}^{\pi_1^{\rho_n},\pi_2^{\rho_n}} h_{1,\pi_2^{\rho_n}}^{\rho_n}(y) - \rho_n h_{1,\pi_2^{\rho_n}}^{\rho_n}(y)\\ &= \sup_{\pi_1\in\Pi_1}\left\{r_1^{\pi_1,\pi_2^{\rho_n}}(y) + \mathcal{L}^{\pi_1,\pi_2^{\rho_n}} h_{1,\pi_2^{\rho_n}}^{\rho_n}(y)\right\} - \rho_n h_{1,\pi_2^{\rho_n}}^{\rho_n}(y) \quad \forall y \in B_R \qquad (4.4.7)\end{aligned}$$

and

$$\begin{aligned}\rho_n v_{1,\pi_2^{\rho_n}}^{\rho_n}(0) &= r_2^{\pi_1^{\rho_n},\pi_2^{\rho_n}}(y) + \mathcal{L}^{\pi_1^{\rho_n},\pi_2^{\rho_n}} h_{2,\pi_1^{\rho_n}}^{\rho_n}(y) - \rho_n h_{2,\pi_1^{\rho_n}}^{\rho_n}(y)\\ &= \sup_{\pi_2\in\Pi_2}\left\{r_2^{\pi_1^{\rho_n},\pi_2}(y) + \mathcal{L}^{\pi_1^{\rho_n},\pi_2} h_{2,\pi_1^{\rho_n}}^{\rho_n}(y)\right\} - \rho_n h_{2,\pi_1^{\rho_n}}^{\rho_n}(y) \quad \forall y \in B_R. \qquad (4.4.8)\end{aligned}$$

Let $\{\rho_n^1\}$ and $\{\rho_n^2\}$ be the subsequences of $\{\rho_n\}$ whose existence was proved in Proposition 4.18(a),(c). From parts (a), (c), (d) and (f) of said result we obtain that, for $R > 0$ fixed:

$$\rho_n^1 v_{1,\pi_2^{\rho_n^1}}^{\rho_n^1}(0) \longrightarrow g_1, \qquad \rho_n^2 v_{1,\pi_2^{\rho_n^2}}^{\rho_n^2}(0) \longrightarrow g_2, \qquad (4.4.9)$$

and

$$\rho_n^1 h_{1,\pi_2^{\rho_n^1}}^{\rho_n^1}(y) \to 0, \qquad \rho_n^2 h_{2,\pi_1^{\rho_n^2}}^{\rho_n^2}(y) \to 0, \qquad \text{in} \quad \mathcal{W}^{2,p}(B_R), \qquad (4.4.10)$$

as $\rho_n \to 0$.

Proposition 4.18 and equations (4.4.5) and (4.4.6) yield that the hypothesis of Theorems 5.3-5.5 from [14] hold, which allows interchanging limits with the supremum and the infinitesimal operator. These theorems together with (4.4.9) and (4.4.10) enable us to assert the existence of two functions $h_{1,\pi_2^*}, h_{2,\pi_1^*} \in \mathcal{W}^{2,p}(B_R)$

such that, if $\rho_n \to 0$ and $(\pi_1^{\rho_n}, \pi_2^{\rho_n}) \xrightarrow{W} (\pi_1^*, \pi_2^*)$, then $h^{\rho_n}_{1,\pi_2^{\rho_n}} \to h_{1,\pi_2^*}$ and $h^{\rho_n}_{2,\pi_1^{\rho_n}} \to h_{2,\pi_1^*}$ uniformly on B_R. Moreover,

$$g_1 = r_1^{\pi_1^*,\pi_2^*}(y) + \mathcal{L}^{\pi_1^*,\pi_2^*} h_{1,\pi_2^*}(y) \tag{4.4.11}$$

$$= \sup_{\pi_1 \in \Pi_1} \left\{ r_1^{\pi_1,\pi_2^*}(y) + \mathcal{L}^{\pi_1,\pi_2^*} h_{1,\pi_2^*}(y) \right\} \qquad \forall y \in B_R \tag{4.4.12}$$

and

$$g_2 = r_2^{\pi_1^*,\pi_2^*}(y) + \mathcal{L}^{\pi_1^*,\pi_2^*} h_{2,\pi_1^*}(y) \tag{4.4.13}$$

$$= \sup_{\pi_2 \in \Pi_2} \left\{ r_2^{\pi_1^*,\pi_2}(y) + \mathcal{L}^{\pi_1^*,\pi_2} h_{2,\pi_1^*}(y) \right\} \qquad \forall y \in B_R. \tag{4.4.14}$$

The same argumentation can be done to extend to all of $y \in \mathbb{R}^m$ due to the fact that $R > 0$ was arbitrary. Now, since the previous convergence was uniform, we can claim that h_{1,π_2^*} and h_{2,π_1^*} are in $\mathcal{W}^{2,p}(\mathbb{R}^m)$ (see [14, Proposition 5.1]).

Finally, using (4.4.11) and (4.4.14) together with Theorem 3.10 we obtain that (π_1^*, π_2^*) is a Nash equilibrium for the average payoff case. ■

Bibliography

[1] A. Arapostathis and V. S. Borkar. Uniform recurrence properties of control diffusions and applications to optimal control. *SIAM J. Control Optim.*, no. 7(**48**):4181–4223, 2010.

[2] A. Arapostathis, M. K. Ghosh, and V. S. Borkar. *Ergodic Control of Diffusion Processes*. Cambridge University Press, 2012.

[3] E. Dockner, S. Jorgensen, N. V. Long, and G. Sorger. *Differential Games in Economics and Management Science*. Cambridge University Press, Cambridge, U.K., 2000.

[4] R. Durrett. *Stochastic Calculus*. CRC Press, 1996.

[5] E. B. Dynkin. *Markov Processes*, volume 1. Springer, 1965.

[6] S. N. Ethier and T. G. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley, New York, 1986.

[7] A. Friedman. Stochastic differential games. *J. Differential Equations*, vol. 1(**11**):79–108, 1972.

[8] X. P. Guo and O. Hernández-Lerma. Zero-sum games for continuous-time jump markov processes in polish spaces and discounted payoff. *Adv. Appl. Probab.*, (**39**):pp. 645–668, 2007.

[9] A. Haurie, J. B. Krawczyk, and G. Zaccour. Games and dynamic games. *World Scientific Books*, 2012.

[10] O. Hernández-Lerma. *Lectures on Continuous-Time Markov Control Processes*. Sociedad Matemática Mexicana, Mexico City, 1994.

[11] H. Jasso-Fuentes. Non-cooperative markov games. *Morfismos*, (**9**):39–54, 2004.

[12] H. Jasso-Fuentes. Non-cooperative markov games. Master's thesis, Cinvestav-IPN, 2004.

[13] H. Jasso-Fuentes and O. Hernández-Lerma. Characterizations of overtaking optimality for controlled diffusion processes. *Appl. Math. Optim.*, (**57**):349–369, 2008.

[14] H. Jasso-Fuentes, J. D. López-Barrientos, and B. A. Escobedo-Trujillo. Infinite-horizon non-zero-sum stochastic differential games with additive structure. *IMA J. Math. Control Inform.*, pages 1–27, 2015. doi: 10.1093/imamci/dnv045.

[15] J. L. López-Barrientos. *Basic and Advanced Optimality Criteria for Zero-Sum Stochastic Differential Games*. PhD thesis, Cinvestav-IPN, 2012.

[16] T. Prieto-Rumeau and O. Hernández-Lerma. Bias and overtaking equilibria for zero-sum continuous-time markov games. *Math. Meth. Oper. Res.*, (**61**):437–454, 2005.

[17] T. Prieto-Rumeau and O. Hernández-Lerma. *Selected Topics on Continuous-time Controlled Markov Chains and Markov Games*. Imperial College Press, London, 2012.

[18] K. Schmüdgen. *Unbounded Self-adjoint Operators on Hilbert Space*. Springer, 2012.

[19] D. V. Widder. *The Laplace Transform*. Princeton University Press, Princeton N. J., 1946.