Morfismos, Vol. 9, No. 1, 2005, pp. 39–54

Noncooperative continuous-time Markov games *

Héctor Jasso-Fuentes

Abstract

This work concerns noncooperative continuous-time Markov games with Polish state and action spaces. We consider finite-horizon and infinite-horizon discounted payoff criteria. Our aim is to give a unified presentation of optimality conditions for general Markov games. Our results include zero-sum and nonzero-sum games.

2000 Mathematics Subject Classification: 91A25, 91A15, 91A10.
Keywords and phrases: Continuous-time Markov games, noncooperative games.

1 Introduction

Continuous-time Markov games form a class of dynamic stochastic games in which the state evolves as a Markov process. The class of Markov games includes (deterministic) differential games, stochastic differential games, jump Markov games and many others, but they are usually studied as separate, different, types of games. In contrast, we propose here a *unified* presentation of optimality conditions for *general* Markov games. In fact, we only consider *noncooperative* games but the same ideas can be extended in an obvious manner to the cooperative case.

As already mentioned, our presentation and results hold for general Markov games but we have to pay a price for such a generality; namely, we restrict ourselves to *Markov strategies*, which depend only

*Research partially supported by a CONACYT scolarship. This paper is part of the author's M. Sc. thesis presented at the Department of Mathematics of CINVESTAV-IPN.

on the current state. More precisely, at each decision time t, the players choose their corresponding actions (independently and simultaneously) depending only on the current state $X(t)$ of the game. Hence, this excludes some interesting situations, for instance, some hierarchical games in which some players "go first".

Our references are mainly on noncooperative continuous-time games. However, for *cooperative* games the reader may consult Filar/Petrosjan [2], Gaidov [3], Haurie [5] and their references. For *discrete time* games see, for instance, Basar/Oldsder [1], Gonzalez-Trejo et al. [4].

A remark on terminology: The Borel σ-algebra of a topological space S is denoted by $\mathcal{B}(S)$. A complete and separable metric space is called a *Polish space.*

2 Preliminaries

Throughout this section we let S be a Polish space, and $X(\cdot) = \{X(t), t \geq 0\}$ a S-valued Markov process defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Denote by $\mathbb{P}(s, x, t, B) := \mathbb{P}(X(t) \in B | X(s) = x)$ for all $t \geq s \geq 0$, $x \in S$ and $B \in \mathcal{B}(S)$, the transition probability function of $X(\cdot)$.

2.1 Semigroups

Definition 2.1 *Let M be the linear space of all real-valued measurable functions v on $\hat{S} := [0, \infty) \times S$ such that*

$$\int_S \mathbb{P}(s, x, t, dy)\ |v(s, y)| < \infty \quad \text{for all } 0 \leq s \leq t \text{ and } x \in S.$$

For each $t \geq 0$ and $v \in M$, we define a function $T_t v$ on $\hat{S}$ as

$$T_t v(s, x) := \int_S \mathbb{P}(s, x, s + t, dy)\ v(s + t, y). \tag{1}$$

Proposition 2.2 *The operators T_t, $t \geq 0$, defined by* (1)*, form a semigroup of operators on M, that is,*

(i) $T_0 = I$, the identity, and

(ii) $T_{t+r} = T_t T_r$.

For a proof of this proposition see, for instance, Jasso-Fuentes [7], Proposition 1.2.2.

2.2 The extended generator

Definition 2.3 *Let $M_0 \subset M$ be the family of functions $v \in M$ for which the following conditions hold:*

a) $\lim_{t\downarrow 0} T_t v(s,x) = v(s,x)$ *for all* $(s,x) \in \hat{S}$*;*

b) there exist $t_0 > 0$ and $u \in M$ such that

$$T_t|v|(s,x) \le u(s,x) \quad \text{for all } (s,x) \in \hat{S} \text{ and } 0 \le t \le t_0.$$

Now let $\mathcal{D}(L) \subset M_0$ be the set of functions $v \in M_0$ for which:

a) the limit

$$\begin{aligned} Lv(s,x) : &= \lim_{t\downarrow 0} \frac{[T_t v(s,x) - v(s,x)]}{t} \\ &= \lim_{t\downarrow 0} \frac{1}{t} \int_S \mathbb{P}(s,x,s+t,dy)[v(s+t,y) - v(s,x)] \end{aligned} \tag{2}$$

exists for all $(s,x) \in \hat{S}$,

b) $Lv \in M_0$, and

c) there exist $t_0 > 0$ and $u \in M$ such that

$$\frac{|T_t v(s,x) - v(s,x)|}{t} \le u(s,x)$$

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

The operator L in (2) will be referred to as the *extended generator* of the semigroup T_t, and the set $\mathcal{D}(L)$ is called the *domain* of L.

The following lemma (which is proved in [7], Lemma 1.3.2, for instance) summarizes some properties of L.

Lemma 2.4 *For each $v \in \mathcal{D}(L)$, the following conditions hold:*

a) $\frac{d^+}{dt} T_t v := \lim_{h\downarrow 0} h^{-1}[T_{t+h}v - T_t v] = T_t L v$*,*

b) $T_t v(s,x) - v(s,x) = \int_0^t T_r(Lv)(s,x)\, dr$*,*

c) if $\rho > 0$ and $v_\rho(s,x) := e^{-\rho s} v(s,x)$, then v_ρ is in $\mathcal{D}(L)$ and

$$Lv_\rho(s,x) = e^{-\rho s}[Lv(s,x) - \rho v(s,x)].$$

2.3 Expected rewards

Let $X(\cdot) = \{X(t), t \geq 0\}$ be as in the previous paragraphs, that is, a Markov process with values in a Polish space S and with transition probabilities $\mathbb{P}(s, x, t, B)$ for all $t \geq s \geq 0$, $x \in S$ and $B \in \mathcal{B}(S)$. Recalling the Definitions 2.1 and 2.3 the semigroup defined in (1) becomes

$$T_t v(s, x) = \mathbb{E}_{sx}[v(s + t, X(s + t))],$$

where $\mathbb{E}_{sx}(\cdot) := \mathbb{E}[\,\cdot\,|X(s) = x]$ is the conditional expectation given $X(s) = x$. Similarly, we can rewrite part b) of Lemma 2.4 as

$$\mathbb{E}_{sx}[v(s + t, X(s + t))] - v(s, x) = \mathbb{E}_{sx}\left[\int_0^t Lv(s + r, X(s + r))dr\right] \tag{3}$$

for each $v \in \mathcal{D}(L)$. We shall refer to (3) as *Dynkin's formula.* The extended generator L of the semigroup $\{T_t\}$ will also be referred to as the *extended generator of the Markov process* $X(\cdot)$.

The following fact will be useful in later sections.

Proposition 2.5 *Fix numbers $\rho \in \mathbb{R}$ and $\tau > 0$. Let $R(s, x)$ and $K(s, x)$ be measurable functions on $S_\tau := [0, \tau] \times S$, and suppose that R is in M_0. If a function $v \in \mathcal{D}(L)$ satisfies the equation*

$$\rho v(s, x) = R(s, x) + Lv(s, x) \tag{4}$$

on S_τ, with the "terminal" condition

$$v(\tau, x) = K(\tau, x), \tag{5}$$

then, for every $(s, x) \in S_\tau$,

$$v(s, x) = \mathbb{E}_{sx}\left[\int_s^\tau e^{-\rho(t-s)}R(t, X(t))dt + e^{-\rho(\tau-s)}K(\tau, X(\tau))\right]. \tag{6}$$

If the equality in (4) is replaced with the inequality "$\leq$" or "$\geq$", then the equality in (6) is replaced with the same inequality, that is, "$\leq$" or "$\geq$" respectively.

Proof: Suppose that v satisfies (4) and let $v_\rho(s, x) := e^{-\rho s}v(s, x)$. Then, by (4) and Lemma 2.4 c), we obtain

$$\begin{aligned} Lv_\rho(s, x) &= e^{-\rho s}[Lv(s, x) - \rho v(s, x)] \\ &= -e^{-\rho s}R(s, x). \end{aligned} \tag{7}$$

Therefore, applying Dynkin's formula (3) to v_ρ and using (7),

$$\begin{aligned} &\mathbb{E}_{sx}\left[e^{-\rho(s+t)}v(s+t,X(s+t))\right]-e^{-\rho s}v(s,x) \\ &= -\mathbb{E}_{sx}\left[\int_0^t e^{-\rho(s+r)}R(s+r,X(s+r))dr\right] \qquad (8) \\ &= -\mathbb{E}_{sx}\left[\int_s^{s+t} e^{-\rho r}R(r,X(r))dr\right]. \end{aligned}$$

The latter expression, with $s+t=\tau$, and (5) give

$$\begin{aligned} &\mathbb{E}_{sx}\left[e^{-\rho\tau}K(\tau,X(\tau))\right]-e^{-\rho s}v(s,x) \\ &= -\mathbb{E}_{sx}\left[\int_s^{\tau} e^{-\rho r}R(r,X(r))dr\right]. \end{aligned}$$

Finally, multiply both sides of this equality by $e^{\rho s}$ and then rearrange terms to obtain (6).

Concerning the last statement in the proposition, suppose that instead of (4) we have $\rho v \geq R + Lv$. Then (7) becomes

$$-e^{-\rho s}R(s,x) \geq Lv_\rho(s,x)$$

and the same calculations in the previous paragraph show that the equality in (6) should be replaced with "$\geq$". For "$\leq$", the result is obtained similarly. □

Observe that the number ρ in Proposition 2.5 can be arbitrary, but in most applications in later sections we will require either $\rho = 0$ or $\rho > 0$. In the latter case ρ is called a "discount factor".

On the other hand, if the function $R(s,x)$ is interpreted as a "reward rate", then (6) represents an expected total reward during the time interval $[s,\tau]$ with initial condition $X(s)=x$ and terminal reward K. This expected reward will be associated with *finite-horizon* games. In contrast, the expected reward in (11), below, will be associated with *infinite-horizon* games.

Proposition 2.6 *Let $\rho > 0$ be a given number, and $R \in M_0$ a function on $\hat{S} := [0,\infty) \times S$. If a function $v \in \mathcal{D}(L)$ satisfies*

$$\rho v(s,x) = R(s,x) + Lv(s,x) \quad \textit{for all } (s,x) \in \hat{S} \tag{9}$$

and is such that, as $t \to \infty$,

$$e^{-\rho t}T_t v(s,x) = e^{-\rho t}\mathbb{E}_{sx}\left[v(s+t,X(s+t))\right] \to 0, \tag{10}$$

then

$$(11) \qquad v(s,x) = \mathbb{E}_{sx}\left[\int_s^\infty e^{-\rho(t-s)} R(t, X(t))dt\right] = \int_0^\infty e^{-\rho t} T_t R(s,x)dt.$$

Moreover, if the equality in (9) *is replaced with the inequality "$\leq$" or "$\geq$", then the equality in* (11) *should be replaced with the same inequality.*

Proof: Observe that the equations (9) and (4) are essentially the same, the only difference being that the former is defined on $\hat{S}$ and the latter on S_τ. At any rate, the calculations in (7)-(8) are also valid in the present case. Hence, multiplying both sides of (8) by $e^{\rho s}$ and then letting $t \to \infty$ and using (10) we obtain (11). The remainder of the proof is as in Proposition 2.5. □

3 The game model and strategies

For notational case, we shall restrict ourselves to the two-player situation. However, the extension to any finite number ≥ 2 of players is completely analogous.

3.1 The game model

Some of the main features of a (two-player) continuous-time Markov game can be described by means of the *game model*

$$(12) \qquad GM := \{S, (A_i, R_i)_{i=1,2}, L^{a_1,a_2}\}$$

with the following components.

- S denotes the game's *state space*, which is assumed to be a Polish space.
- Associated with each player $i = 1, 2$, we have

  $$(13) \qquad (A_i, R_i)$$

  where A_i is a Polish space that stands for the *action space* (or *control set*) for player i. Let

  $$\mathbf{A} := A_1 \times A_2, \quad \text{and} \quad \mathbf{K} := S \times \mathbf{A}.$$

The second component in (13) is a real-valued measurable function R_i on

$$[0,\infty) \times \mathbf{K} = [0,\infty) \times S \times \mathbf{A} = \hat{S} \times \mathbf{A} \quad (\hat{S} := [0,\infty) \times S),$$

which denotes the *reward rate* function for player i. (Observe that $R_i(s,x,a_1,a_2)$ depends on the actions $(a_1,a_2) \in \mathbf{A}$ of both players.)

- For each pair $\mathbf{a} = (a_1,a_2) \in \mathbf{A}$ there is a linear operator $L^{\mathbf{a}}$ with domain $\mathcal{D}(L^{\mathbf{a}})$, which is the extended generator of a S-valued Markov process with transition probability $\mathbb{P}^{\mathbf{a}}(s,x,t,B)$.

The game model (12) is said to be *time-homogeneous* if the reward rates are time-invariant and the transition probabilities are time-homogeneous, that is,

$$R_i(s,x,\mathbf{a}) = R_i(x,\mathbf{a}) \quad \text{and} \quad \mathbb{P}^{\mathbf{a}}(s,x,t,B) = \mathbb{P}^{\mathbf{a}}(t-s,x,B).$$

Summarizing, the game model (12) tells us *where* the game lives (the state space S) and *how* it moves (according to the players' actions $\mathbf{a} = (a_1,a_2)$ and the Markov process associated to $L^{\mathbf{a}}$). The reward rates R_i are used to define the payoff function that player i ($i = 1,2$) wishes to "optimize"— see for instance (15) and (16) below. To do this optimization each player uses, when possible, suitable "strategies", such as those defined next.

3.2 Strategies

We will only consider *Markov* (also known as *feedback*) *strategies*, namely, for each player $i = 1,2$, measurable functions π_i from $\hat{S} := [0,\infty) \times S$ to A_i. Thus, $\pi_i(s,x) \in A_i$ denotes the action of player i prescribed by the strategy π_i if the state is $x \in S$ at time $s \geq 0$. In fact, we will restrict ourselves to classes Π_1, Π_2 of Markov strategies that satisfy the following.

Assumption 3.1 *For each pair $\pi = (\pi_1,\pi_2) \in \Pi_1 \times \Pi_2$, there exists a strong Markov process $X^\pi(\cdot) = \{X^\pi(t), t \geq 0\}$ such that:*

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

b) The extended generator L^π of $X^\pi(\cdot)$ satisfies that

$$L^\pi = L^{\mathbf{a}} \quad \textit{if} \quad (\pi_1(s,x), \pi_2(s,x)) = (a_1, a_2) = \mathbf{a}.$$

The set $\Pi_1 \times \Pi_2$ in Assumption 3.1 is called the family of *admissible* pairs of Markov strategies. A pair $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$ is said to be *stationary* if $\pi_i(s,x) \equiv \pi_i(x)$ does not depend on $s \geq 0$.

Clearly, the function spaces $M \supset M_0 \supset \mathcal{D}(L)$ introduced in Section 2 depend on the pair $\pi = (\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$ of strategies being used, because so does $\mathbb{P}^\pi$. Hence, these spaces will now be written as M^π, M_0^π, $\mathcal{D}(L^\pi)$, and they are supposed to verify the following conditions.

Assumption 3.2 *a) There exist nonempty spaces $\mathcal{M} \supset \mathcal{M}_0 \supset \mathbf{D}$, which do not depend on π, such that, for all $\pi = (\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$*

$$\mathcal{M} \subset M^\pi \ , \quad \mathcal{M}_0 \subset M_0^\pi \ , \quad \mathbf{D} \subset \mathcal{D}(\mathbf{L}^\pi)$$

and, in addition, the operator L^π is the closure of its restriction to $\mathbf{D}$.

b) For $\pi = (\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$ and $i = 1, 2$, the reward rate $R_i(s, x, a_1, a_2)$ is such that R_i^π is in $\mathcal{M}_0$, where

$$R_i^\pi(s,x) := R_i(s, x, \pi_1(s,x), \pi_2(s,x)).$$

Sometimes we shall use the notation

$$R_i^\pi(s,x) := R_i(s, x, \pi_1, \pi_2) \quad \text{for } \pi = (\pi_1, \pi_2),\ i = 1, 2. \tag{14}$$

If the game model is time-homogeneous and the pair (π_1, π_2) is stationary, then (14) reduces to

$$R_i^\pi(x) := R_i(x, \pi_1(x), \pi_2(x)) = R_i(x, \pi_1, \pi_2).$$

Throughout the remainder of this paper we consider the game model GM in (12) under Assumptions 3.1 and 3.2.

4 Noncooperative equilibria

Let GM be as in (12). In this work, we are concerned with the following two types of payoff functions, where we use the notation (14). For each pair of strategies $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$ and each player $i = 1, 2$:

- *The finite-horizon payoff*

$$
\begin{aligned}
V_\tau^i(s,x,\pi_1,\pi_2) : \quad &= \quad \mathbb{E}_{sx}^{\pi_1,\pi_2}\left[\int_s^\tau e^{-\rho(t-s)} R_i(t,X(t),\pi_1,\pi_2)dt\right. \\
&\qquad \left. + e^{-\rho(\tau-s)} K_i(\tau, X(\tau)) \right]
\end{aligned} \tag{15}
$$

where $0 \le s \le \tau$, $x \in S$, K_i is a function in $\mathcal{M}$ (the space in Assumption 3.2 a)), and $\rho \ge 0$ is a "discount factor". The time $\tau > 0$ is called the game's horizon or "terminal time", and K_i is a "terminal reward".

- *The infinite-horizon discounted payoff*

$$
V^i(s,x,\pi_1,\pi_2) := \mathbb{E}_{sx}^{\pi_1,\pi_2}\left[\int_s^\infty e^{-\rho(t-s)} R_i(t,X(t),\pi_1,\pi_2)dt\right] \tag{16}
$$

where $s \ge 0$, $x \in S$, and $\rho > 0$ is a (fixed) discount factor.

Each player $i = 1,2$ wishes to "optimize" his payoff in the following sense.

Definition 4.1 *For $i = 1,2$, let V_τ^i be as in* (15)*, and define $S_\tau := [0,\tau] \times S$. A pair $(\pi_1^*, \pi_2^*) \in \Pi_1 \times \Pi_2$ of admissible strategies is said to be a* noncooperative equilibrium*, also known as a* Nash equilibrium*, if for all $(s,x) \in S_\tau$*

$$
V_\tau^1(s,x,\pi_1^*,\pi_2^*) \ge V_\tau^1(s,x,\pi_1,\pi_2^*) \quad \textit{for all } \pi_1 \in \Pi_1 \tag{17}
$$

and

$$
V_\tau^2(s,x,\pi_1^*,\pi_2^*) \ge V_\tau^2(s,x,\pi_1^*,\pi_2) \quad \textit{for all } \pi_2 \in \Pi_2. \tag{18}
$$

Hence, (π_1^*, π_2^*) is a Nash equilibrium if for each $i = 1,2$, π_i^* maximizes over Π_i the payoff function V_τ^i of player i when the other player, say $j \ne i$, uses the strategy π_j^*.

For the infinite-horizon payoff function in (16), the definition of Nash equilibrium is the same as in Definition 4.1 with V^i and $\hat{S} := [0,\infty) \times S$ in lieu of V_τ^i and S_τ, respectively.

Zero-sum games. For $i = 1,2$, let $F_i(s,x,\pi_1,\pi_2)$ be the payoff function in either (15) or (16). The game is called a *zero-sum game* if

$$
F_1(s,x,\pi_1,\pi_2) + F_2(s,x,\pi_1,\pi_2) = 0 \quad \text{for all } s,x,\pi_1,\pi_2,
$$

that is, $F_1 = -F_2$. Therefore, if we define $F := F_1 = -F_2$, it follows from (17) and (18) that player 1 wishes to *maximize* $F(s,x,\pi_1,\pi_2)$ over Π_1, whereas player 2 wishes to *minimize* $F(s,x,\pi_1,\pi_2)$ over Π_2, so (17) and (18) become

$$F(s,x,\pi_1,\pi_2^*) \leq F(s,x,\pi_1^*,\pi_2^*) \leq F(s,x,\pi_1^*,\pi_2) \tag{19}$$

for all $\pi_1 \in \Pi_1$ and $\pi_2 \in \Pi_2$, and all (s,x). In this case the Nash equilibrium (π_1^*,π_2^*) is called a *saddle point*.

In the zero-sum case, the functions

$$L(s,x) := \sup_{\pi_1\in\Pi_1} \inf_{\pi_2\in\Pi_2} F(s,x,\pi_1,\pi_2) \tag{20}$$

and

$$U(s,x) := \inf_{\pi_2\in\Pi_2} \sup_{\pi_1\in\Pi_1} F(s,x,\pi_1,\pi_2) \tag{21}$$

play an important role. The function $L(s,x)$ is called the game's *lower value* (with respect to the payoff $F(s,x,\pi_1,\pi_2)$) and $U(s,x)$ is the game's *upper value*. Clearly, we have

$$L(s,x) \leq U(s,x) \quad \text{for all } (s,x). \tag{22}$$

If the upper and lower values coincide, then the game is said to have a *value*, and the *value of the game*, call it $\mathcal{V}(s,x)$ is the common value of $L(s,x)$ and $U(s,x)$, i.e.

$$\mathcal{V}(s,x) := L(s,x) = U(s,x) \quad \text{for all } (s,x).$$

On the other hand, if (π_1^*,π_2^*) satisfies (19), a trivial calculation yields

$$U(s,x) \leq F(s,x,\pi_1^*,\pi_2^*) \leq L(s,x) \quad \text{for all } (s,x),$$

which together with (22) gives the following.

Proposition 4.2 *If the zero-sum game with payoff function F has a saddle point (π_1^*,π_2^*), then the game has the value*

$$\mathcal{V}(s,x) = F(s,x,\pi_1^*,\pi_2^*) \quad \textit{for all } (s,x).$$

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

Proposition 4.3 *Suppose that there is a pair of admissible strategies* π_1^*, π_2^* *that satisfy, for all* (s,x),

$$
\begin{aligned}
F(s,x,\pi_1^*,\pi_2^*) &= \sup_{\pi_1\in\Pi_1} F(s,x,\pi_1,\pi_2^*) \\
&= \inf_{\pi_2\in\Pi_2} F(s,x,\pi_1^*,\pi_2).
\end{aligned}
\tag{23}
$$

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

Proof: Let (π_1^*,π_2^*) be a pair of admissible strategies that satisfy (23). Then, for all (s,x), from the first equality in (23) we obtain

$$F(s,x,\pi_1^*,\pi_2^*) \geq F(s,x,\pi_1,\pi_2^*) \quad \text{for all } \pi_1\in\Pi_1,$$

which is the first inequality in (19). Similarly, the second equality in (23) yields the second inequality in (19), and it follows that (π_1^*,π_2^*) is a saddle point. □

In the next section we give conditions for a pair of strategies to be a saddle point, and in Section 6 we study the so-called *nonzero-sum* case as in (17), (18).

5 Zero-sum games

In this section we study the existence of saddle points for the finite-horizon and infinite-horizon payoffs in (15) and (16), respectively.

Finite-horizon payoff

As in (19)-(21), the finite-horizon payoff (15), in the zero-sum case, does not depend on $i=1,2$. Hence, we have the payoff

$$
\begin{aligned}
V_\tau(s,x,\pi_1,\pi_2) : &= \mathbb{E}_{sx}^{\pi_1,\pi_2}\left[\int_s^\tau e^{-\rho(t-s)}R(t,X(t),\pi_1,\pi_2)dt\right. \\
&\qquad \left.+e^{-\rho(\tau-s)}K(\tau,X(\tau))\right].
\end{aligned}
$$

This function V_τ plays now the role of F in (19)-(23). Recall that the Assumptions 3.1 and 3.2 are supposed to hold.

Theorem 5.1 *Consider $\rho \in \mathbb{R}$ and $\tau > 0$ fixed. Moreover let $R(s,x,a_1,a_2)$ and $K(s,x,a_1,a_2)$ be measurable functions on $S_\tau \times \mathbf{A}$, where $S_\tau := [0,\tau] \times S$ and $\mathbf{A} := A_1 \times A_2$. Suppose that for each pair $(\pi_1,\pi_2) \in \Pi_1 \times \Pi_2$, the function $R(s,x,\pi_1,\pi_2)$ is in $\mathcal{M}_0$. In addition, suppose that there is a function $v(s,x) \in \mathbf{D}$ and a pair of strategies $(\pi_1^*,\pi_2^*) \in \Pi_1 \times \Pi_2$ such that, for all $(s,x) \in S_\tau$,*

$$\rho v(s,x) = \inf_{\pi_2 \in \Pi_2} \{R(s,x,\pi_1^*,\pi_2) + L^{\pi_1^*,\pi_2} v(s,x)\} \tag{24}$$

$$= \sup_{\pi_1 \in \Pi_1} \{R(s,x,\pi_1,\pi_2^*) + L^{\pi_1,\pi_2^*} v(s,x)\} \tag{25}$$

$$= R(s,x,\pi_1^*,\pi_2^*) + L^{\pi_1^*,\pi_2^*} v(s,x) \tag{26}$$

with the boundary condition

$$v(\tau,x) = K(\tau,x) \quad \textit{for all } x \in S. \tag{27}$$

Then

- *a) $v(s,x) = V_\tau(s,x,\pi_1^*,\pi_2^*)$ for all $(s,x) \in S_\tau$;*
- *b) (π_1^*,π_2^*) is a saddle point and $v(s,x)$ is the value of the game.*

Proof:

- a) Comparing (26)-(27) with (4)-(5), we conclude that part a) follows from Proposition 2.5.
- b) Assume for a moment that, for all $(s,x) \in S_\tau$ and all pairs (π_1,π_2) of admissible strategies, we have

$$V_\tau(s,x,\pi_1,\pi_2^*) \le v(s,x) \le V_\tau(s,x,\pi_1^*,\pi_2) \tag{28}$$

If this is indeed true, then b) will follow from part a) together with (19) and Proposition 4.2. Hence it suffices to prove (28).

To this end, let us call $F(s,x,\pi_1,\pi_2)$ the function inside the brackets in (24)-(25), i.e.

$$F(s,x,\pi_1,\pi_2) := R(s,x,\pi_1,\pi_2) + L^{\pi_1,\pi_2} v(s,x). \tag{29}$$

Interpreting this function as the payoff of a certain game, it follows from (24)-(26) and the Proposition 4.3 that the pair (π_1^*,π_2^*) is a saddle point, that is, $F(s,x,\pi_1^*,\pi_2^*) = \rho v(s,x)$ satisfies (19). More explicitly, from (29) and the equality $F(s,x,\pi_1^*,\pi_2^*) = \rho v(s,x)$, (19) becomes: for all $\pi_1 \in \Pi_1$ and $\pi_2 \in \Pi_2$,

$$R(s,x,\pi_1,\pi_2^*) + L^{\pi_1,\pi_2^*}v(s,x) \quad \leq \quad \rho v(s,x) \leq R(s,x,\pi_1^*,\pi_2) + L^{\pi_1^*,\pi_2}v(s,x).$$

These two inequalities together with the second part of Proposition 2.5 give (28). □

Infinite-horizon discounted payoff

We now consider the infinite-horizon payoff in (16), which in the zero-sum case can be interpreted as

$$V(s,x,\pi_1,\pi_2) = \mathbb{E}_{sx}^{\pi_1,\pi_2}\left[\int_s^\infty e^{-\rho(t-s)}R(t,X(t),\pi_1,\pi_2)dt\right].$$

Exactly the same arguments used in the proof of Theorem 5.1 but replacing Proposition 2.5 with Proposition 2.6, give the following result in the *infinite-horizon* case.

Theorem 5.2 *Suppose $\rho > 0$. Let $R(s,x,a_1,a_2)$ be as in Assumption 3.2 b). Suppose that there exist a function $v \in \mathbf{D}$ and a pair of strategies $(\pi_1^*,\pi_2^*) \in \Pi_1 \times \Pi_2$ such that, for all $(s,x) \in \hat{S} := [0,\infty) \times S$,*

$$\begin{aligned} \rho v(s,x) &= \inf_{\pi_2 \in \Pi_2}\{R(s,x,\pi_1^*,\pi_2) + L^{\pi_1^*,\pi_2}v(s,x)\} \\ &= \sup_{\pi_1 \in \Pi_1}\{R(s,x,\pi_1,\pi_2^*) + L^{\pi_1,\pi_2^*}v(s,x)\} \\ &= R(s,x,\pi_1^*,\pi_2^*) + L^{\pi_1^*,\pi_2^*}v(s,x) \end{aligned} \tag{30}$$

and, moreover, for all $(s,x) \in \hat{S}$ and $(\pi_1,\pi_2) \in \Pi_1 \times \Pi_2$,

$$e^{-\rho t}\mathbb{E}_{sx}^{\pi_1,\pi_2}[v(s+t,X(s+t))] \to 0 \quad \text{as } t \to \infty. \tag{31}$$

Then

a) $v(s,x) = V(s,x,\pi_1^*,\pi_2^*)$ for all $(s,x) \in \hat{S}$;

b) (π_1^*,π_2^*) is a saddle point for the infinite-horizon discounted payoff, and $v(s,x)$ is the value of the game.

Proof: Comparing (30)-(31) with (9)-(10) we can use Proposition 2.6 to obtain a).

To obtain b), we follow the same steps used in the proof of Theorem 5.1 but replacing Proposition 2.5 with Proposition 2.6, and S_τ with $\hat{S}$. □

6 Nonzero-sum games

An arbitrary game which does not satisfy the zero-sum condition is called a *nonzero-sum* game. In this section we are concerned with the existence of Nash equilibria for nonzero-sum continuous-time Markov games with the payoff functions (15) and (16).

Finite-horizon payoff

For $i = 1, 2$, let $V^i_\tau(s, x, \pi_1, \pi_2)$ be the finite-horizon payoff in (15). In this setting, the following theorem gives sufficient conditions for the existence of a Nash equilibrium — see Definition 4.1.

Theorem 6.1 *Suppose that for $i = 1, 2$, there are functions $v_i(s, x)$ in $\mathbf{D}$ and strategies $\pi^*_i \in \Pi_i$ that satisfy, for all $(s, x) \in S_\tau$, the equations*

$$(32) \qquad \begin{aligned} \rho v_1(s,x) &= \max_{\pi_1 \in \Pi_1} \{R_1(s,x,\pi_1,\pi^*_2) + L^{\pi_1,\pi^*_2} v_1(s,x)\} \\ &= R_1(s,x,\pi^*_1,\pi^*_2) + L^{\pi^*_1,\pi^*_2} v_1(s,x) \end{aligned}$$

and

$$(33) \qquad \begin{aligned} \rho v_2(s,x) &= \max_{\pi_2 \in \Pi_2} \{R_2(s,x,\pi^*_1,\pi_2) + L^{\pi^*_1,\pi_2} v_2(s,x)\} \\ &= R_2(s,x,\pi^*_1,\pi^*_2) + L^{\pi^*_1,\pi^*_2} v_2(s,x), \end{aligned}$$

as well as the boundary (or "terminal") conditions

$$(34) \quad v_1(\tau, x) = K_1(\tau, x) \quad \textit{and} \quad v_2(\tau, x) = K_2(\tau, x) \quad \textit{for all } x \in S.$$

Then (π^*_1, π^*_2) is a Nash equilibrium and for each player $i = 1, 2$ the expected payoff is

$$(35) \qquad v_i(s, x) = V^i_\tau(s, x, \pi^*_1, \pi^*_2) \quad \textit{for all } (s, x) \in S_\tau.$$

Proof: From the second equality in (32) together with the first boundary condition in (34), the Proposition 2.5 gives (35) for $i = 1$. A similar argument gives of course (35) for $i = 2$.

On the other hand, from the first equality in (32) we obtain

$$\rho v_1(s, x) \geq R_1(s, x, \pi_1, \pi^*_2) + L^{\pi_1,\pi^*_2} v_1(s, x) \quad \text{for all } \pi_1 \in \Pi_1.$$

Thus using again Proposition 2.5 we obtain

$$v_1(s,x) \geq V^1_\tau(s,x,\pi_1,\pi_2^*) \quad \text{for all } \pi_1 \in \Pi_1,$$

which combined with (35) for $i = 1$ yields (17). A similar argument gives (18) and the desired conclusion follows. □

Infinite-horizon discounted payoff

Let us now consider the infinite-horizon payoff $V^i(s,x,\pi_1,\pi_2)$ in (16). The corresponding analogue of Theorem 6.1 is as follows.

Theorem 6.2 *Suppose that, for $i = 1, 2$, there are functions $v_i(s,x) \in$ **D** and strategies $\pi_i^* \in \Pi_i$ that satisfy, for all $(s,x) \in \hat{S}$, the equations* (32) *and* (33) *together with the condition*

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

for all $\pi_1 \in \Pi_1$, $\pi_2 \in \Pi_2$, $i = 1, 2$, and $(s,x) \in \hat{S}$. Then (π_1^*, π_2^*) is a Nash equilibrium for the infinite-horizon discounted payoff (16) *and the expected payoff is*

$$v_i(s,x) = V_i(s,x,\pi_1^*,\pi_2^*) \quad \textit{for all } (s,x) \in \hat{S},\ i = 1, 2.$$

We omit the proof of this theorem because it is essentially the same as the proof of Theorem 6.1 (using Proposition 2.6 in lieu of Proposition 2.5).

7 Concluding remarks

In this paper we have presented a unified formulation of continuous-time Markov games, similar to the one-player (or control) case in Hernández-Lerma[6]. This formulation is quite general and it includes practically any kind of Markov games, but of course it comes at price because we have restricted ourselves to *Markov strategies*, which are memoryless. In other words, our players are not allowed to use past information; they base their decisions on the current state only. This is a serious restriction that needs to be eliminated, and so it should lead to future work.

Acknowledgement

Thanks to Prof. Onésimo Hernández-Lerma for valuable comments and discussions on this work.

Héctor Jasso-Fuentes
Departamento de Matemáticas,
CINVESTAV-IPN,
A.P. 14-470,
México D.F. 07000,
México.

References

[1] Basar T.; Olsder G.J., Dynamic Noncooperative Game Theory, Second Edition, SIAM, Philadelphia, 1999.

[2] Filar J.A.; Petrosjan L.A., *Dynamic cooperative games*, International Game Theory Review **2** (2000), 47–65.

[3] Gaidov S.D., *On the Nash-bargaining solution in stochastic differential games*, Serdica **16** (1990), 120–125.

[4] González-Trejo J.I.; Hernández-Lerma O.; Hoyos-Reyes L.F., *Minimax control of discrete-time stochastic system*, SIAM J. Control Optim. **41** (2003), 1626–1659.

[5] Haurie A., *A historical perspective on cooperative differential games*, in: Advances in Dynamic Games and Applications (Maastricht, 1998), 19–29, Birkhäuser Boston, 2001.

[6] Hernández-Lerma O., Lectures on Continuous-Time Markov Control Processes, Sociedad Matemática Mexicana, México D.F., 1994.

[7] Jasso-Fuentes H., *Noncooperative continuous-time Markov games*, Tesis de Maestría, CINVESTAV-IPN, México D.F., 2004.