Math Meth Oper Res (2005) 61: 437–454
DOI 10.1007/s001860400392

Mathematical Methods
of Operations Research

Bias and overtaking equilibria for zero-sum continuous-time Markov games

Tomás Prieto-Rumeau[1],*, Onésimo Hernández-Lerma[2],†

[1] Departamento de Economía Aplicada Cuantitativa II. Facultad de Ciencias Económicas y Empresariales, Universidad Nacional de Educación a Distancia, Calle Senda del Rey 11, 28040 Madrid, Spain. (e-mail: tprieto@cee.uned.es, Tel.: +34-913-986-399, Fax: +34-913-986-335)
[2] Departamento de Matemáticas, CINVESTAV-IPN. A. Postal 14-740, México D.F. 07000, Mexico (e-mail: ohernand@math.cinvestav.mx, Tel.: 52-55-5061-3871, Fax: 52-55-5061-3876)

Manuscript received: February 2004/Final version received: July 2004

Abstract. This paper deals with continuous-time zero-sum two-person Markov games with denumerable state space, general (Borel) action spaces and possibly unbounded transition and reward/cost rates. We analyze the bias optimality and the weakly overtaking optimality criteria. An example shows that, in contrast to control (or one-player) problems, these criteria are not equivalent for games.

Key words: Continuous-time zero-sum Markov games, Bias optimality, Overtaking optimality

AMS Mathematics Subject Classifications (2000): 91A15, 91A25, 60J27

1. Introduction

We are concerned with continuous-time two-person zero-sum stochastic games with denumerable state space. Most of the papers dealing with Markov games consider discrete-time models (e.g. [9, 10, 11]) but there are just a few references that analyze continuous-time games; see e.g. [6, 7, 17]. In [17], a general state space is considered but restrictive boundedness assumptions are made. On the contrary, in the papers by Guo and Hernández-Lerma [6, 7], for a denumerable state space, both the reward/cost rates and the transition rates are allowed to be unbounded. In this paper, we follow this approach.

* The research of this author was supported by a grant from the Spanish Secretaría de Estado de Educación y Universidades in cooperation with the European Social Funds.
† The research of this author was partially supported by CONACyT Grant 37355-E.

One of the most widely used optimality criteria is average optimality but, as it is well known, this criterion is very underselective and, thus, more restrictive optimality criteria have been proposed in the stochastic control literature as, for instance, bias optimality [8, 12, 14], overtaking optimality [1, 4, 8, 12, 15, 18] and sensitive discount optimality [13, 14], among others. The bias and overtaking optimality criteria are concerned with the asymptotic optimization of the total expected reward (or cost) on finite-horizon problems, as the time horizon goes to infinity. The bias optimality criterion, for stochastic games, is implicitly introduced in [10, 11]. Overtaking optimality for stochastic games is analyzed in [3, 4, 16] and also by Nowak in [10, 11]. Sensitive discount optimality for stochastic games is studied in [10].

In the previous papers dealing with stochastic games, the obtained results were more or less direct generalizations of the corresponding results in stochastic control. For instance, the existence of optimal strategies or the existence of the value of the game were obtained under quite similar assumptions; see e.g. [7, 9]. In this paper and for the first time (as far as we know), we exhibit an important discrepancy between control and game models. More precisely, in stochastic control, bias and weak overtaking optimality are essentially equivalent (see [8, Section 10.3] and [12]) whereas, under similar hypotheses on a stochastic game model, we prove that there exist bias optimal strategies though there might *not* exist weakly overtaking optimal strategies.

The rest of the paper is organized as follows. In Section 2 we define the game model and introduce our assumptions. In Section 3 we recall some results on the average optimality equations introduced in [7] and we make a more detailed analysis of the solutions to those equations. We define the bias optimality criterion in Section 4 and, by introducing the so-called bias optimality equations, we prove the existence of bias optimal strategies. Section 5 is devoted to the weak overtaking optimality criterion: the relations existing between this criterion and bias optimality are explored, and we provide a counterexample showing that weakly overtaking equilibria might not exist. Finally, in Section 6, we conclude with some remarks in which we point out that the sensitive discount optimality criteria, which have been extensively studied in control models, might be of limited interest in stochastic games.

2. Preliminaries

In this section we define the game model we will deal with and introduce our assumptions.

The game model. The continuous-time two-person zero-sum game we are concerned with is given by

- The state space S, assumed to be a denumerable set. We suppose without loss of generality that $S = \{0, 1, \ldots\}$.
- The action sets A and B for players 1 and 2, respectively, which are supposed to be Borel spaces. For each $i \in S$, the (nonempty) Borel set $A(i) \subseteq A$ (resp. $B(i) \subseteq B$) stands for the set of admissible control actions for player 1 (resp. player 2) in state i. Define

  $$K := \{(i, a, b) : i \in S, a \in A(i), b \in B(i)\}.$$

- The system's transition rates $q_{ij}(a,b)$, where $j \in S$ and $(i,a,b) \in K$. They verify that $q_{ij}(a,b) \geq 0$ whenever $j \neq i$, and they are assumed to be conservative, i.e.

  $$\sum_{j \in S} q_{ij}(a,b) = 0 \quad \text{for every } (i,a,b) \in K,$$

  and stable, which means that

  $$q(i) := \sup_{(a,b) \in A(i) \times B(i)} \{-q_{ii}(a,b)\}$$

  is finite for every $i \in S$. Finally, given i and j in S, we suppose that $(a,b) \mapsto q_{ij}(a,b)$ is measurable on $A(i) \times B(i)$.
- The reward/cost rate function $r : K \to \mathbb{R}$, assumed to be measurable on $A(i) \times B(i)$ for each $i \in S$ fixed. For player 1, r represents the reward rate whereas r is the cost rate for player 2.

The so-defined game model is written

$$\mathscr{M} := \{S, (A(i), B(i), i \in S), (q_{ij}(a,b)), (r(i,a,b))\}.$$

The game is played as follows. At each time $t \geq 0$ both players observe the state of the system, say $i \in S$, and they independently choose control actions $a_t \in A(i)$ and $b_t \in B(i)$. Then their reward/cost rate at time t is $r(i, a_t, b_t)$ and, also, the system moves to a state $j \neq i$ with a probability rate $q_{ij}(a_t, b_t)$.

The goal of player 1 (resp. player 2) is to maximize (resp. minimize) his/her reward (resp. cost) over the time horizon $[0, \infty)$ with respect to some suitably defined optimality criterion. We shall deal with three different optimality criteria: average optimality, bias optimality and weak overtaking optimality.

Strategies. In this paper we will restrict ourselves to stationary strategies. The reasons for this are, first of all, that our assumptions will ensure the existence of optimal stationary strategies for the average optimality criterion (see [7]). Second, the bias optimality criterion is usually defined (in stochastic control problems) only on the class of stationary strategies (see e.g. [12]) and, finally, one cannot expect to find an overtaking optimal policy in the class of nonstationary policies; see [2] and the comment after Theorem 5 in [11].

For each state $i \in S$, let $\mathscr{P}(A(i))$ be the space of probability measures on $A(i)$ endowed with the topology of weak convergence. The space $\mathscr{P}(B(i))$ is defined similarly. We will also use the notation $\overline{A}(i) := \mathscr{P}(A(i))$ and $\overline{B}(i) := \mathscr{P}(B(i))$ for $i \in S$.

A randomized stationary strategy π^1 for player 1 is a family of probability measures $\pi^1(\cdot|i)$ in $\mathscr{P}(A(i))$ for each $i \in S$. The set of stationary strategies for player 1 is denoted Π^1. When using policy $\pi^1 \in \Pi^1$, player 1 randomly chooses a control action which depends on the state of the system but not on the time t. We define similarly the stationary strategies for player 2, written $\pi^2 \in \Pi^2$.

For a general definition of admissible nonstationary strategies the interested reader is referred to [7].

When the players use strategies $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$ the (stationary) transition rates of the system are

$$q_{ij}(\pi^1,\pi^2) := \int_{B(i)} \int_{A(i)} q_{ij}(a,b)\pi^1(da|i)\pi^2(db|i) \quad \text{for } i,j \in S,$$

and the reward/cost rate is

$$r(i,\pi^1,\pi^2) := \int_{B(i)} \int_{A(i)} r(i,a,b)\pi^1(da|i)\pi^2(db|i) \quad \text{for } i \in S.$$

We also introduce the notation

$$q_{ij}(\phi,\psi) := \int_{B(i)} \int_{A(i)} q_{ij}(a,b)\phi(da)\psi(db) \tag{2.1}$$

and

$$r(i,\phi,\psi) := \int_{B(i)} \int_{A(i)} r(i,a,b)\phi(da)\psi(db) \tag{2.2}$$

for $i,j \in S$, $\phi \in \overline{A}(i)$ and $\psi \in \overline{B}(i)$. Our assumptions below ensure that the above integrals are well defined.

Assumptions. Now we state the assumptions we make on the game model $\mathscr{M}$. They are supposed to hold throughout the following.

Assumption A. There exist a sequence $\{S_m\}_{m\geq 1}$ of subsets of S, a nondecreasing function $w : S \to [1,\infty)$ and constants $c > 0$, $d \geq 0$ and $M > 0$ such that

(i) $S_m \uparrow S$ and $\sup_{i\in S_m} q(i) < \infty$;
(ii) $\lim_{m\to\infty} \inf_{j\notin S_m}\{w(j)\} = +\infty$;
(iii) for every $(i,a,b) \in K$, $\sum_{j\in S} q_{ij}(a,b)w(j) \leq -cw(i) + d\mathbf{1}_{\{i=0\}}$, where $\mathbf{1}$ denotes the indicator function;
(iv) $|r(i,a,b)| \leq Mw(i)$ for every $(i,a,b) \in K$.

Assumption A(i) and A(ii) are not necessary when the transition rates are bounded, that is, when $\sup_{i\in S} q(i)$ is finite. Assumption A guarantees, for each $(\pi^1,\pi^2) \in \Pi^1 \times \Pi^2$, the existence of a regular Q-process with conservative transition rate matrix

$$Q(\pi^1,\pi^2) := \{q_{ij}(\pi^1,\pi^2)\}_{i,j\in S}.$$

We denote by $\{x(t,\pi^1,\pi^2)\}_{t\geq 0}$ the homogeneous Markov process defined by $Q(\pi^1,\pi^2)$ and, for each initial state $i \in S$, let $E_i^{\pi^1,\pi^2}$ be the corresponding expectation operator. For a detailed construction of the Markov process $\{x(t,\pi^1,\pi^2)\}_{t\geq 0}$ we refer to [7].

Next we state the usual compactness-continuity conditions.

Assumption B.

(i) For each $i \in S$, the action sets $A(i)$ and $B(i)$ are compact.

(ii) Given $i, j \in S$, the functions

$$(a,b) \mapsto q_{ij}(a,b), \quad (a,b) \mapsto \sum_{j \in S} q_{ij}(a,b)w(j) \quad \text{and} \quad (a,b) \mapsto r(i,a,b)$$

are continuous on $A(i) \times B(i)$.

(iii) There exist $w' : S \to [0, \infty)$ and constants $c' > 0$, $d' \geq 0$ and $M' > 0$ for which

$$q(i)w(i) \leq M'w'(i) \quad \text{and} \quad \sum_{j \in S} q_{ij}(a,b)w'(j) \leq c'w'(i) + d'$$

for all $(i,a,b) \in K$.

Finally, we must impose a condition ensuring that the Markov processes $\{x(t, \pi^1, \pi^2)\}_{t \geq 0}$, for $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$, are irreducible and that they verify the *uniform exponential ergodic property*. To this end, we propose the following sufficient *monotonicity condition* that was introduced in [5].

Assumption C.

(i) For each $(a_i, b_i) \in A(i) \times B(i)$ and $(a_{i+1}, b_{i+1}) \in A(i+1) \times B(i+1)$,

$$\sum_{j \geq k} q_{ij}(a_i, b_i) \leq \sum_{j \geq k} q_{i+1,j}(a_{i+1}, b_{i+1})$$

for every $i, k \in S$, provided that $k \neq i+1$.

(ii) Given two states $i \neq j$, either $q_{ij}(a,b) > 0$ for every $(a,b) \in A(i) \times B(i)$, or there exist l states $i_1, i_2, \ldots, i_l$, with $i \neq i_1$ and $i_m \neq i_{m+1}$, for $m = 1, \ldots, l-1$, such that

$$q_{ii_1}(a,b)q_{i_1i_2}(a_{i_1}, b_{i_1}) \cdots q_{i_lj}(a_{i_l}, b_{i_l}) > 0$$

for every $(a,b) \in A(i) \times B(i)$ and $(a_{i_m}, b_{i_m}) \in A(i_m) \times B(i_m)$, for $m = 1, \ldots, l$.

(iii) For $j > i > 0$, either $q_{ij}(a,b) > 0$ for all $(a,b) \in A(i) \times B(i)$ or there exist n states $j_1, \ldots, j_n$ such that, defining $j_0 := i$, we have $j_{m-1} \neq j_m$ and $j_m \neq 0$ for $m = 1, \ldots, n$, and $j_n \geq j$. Moreover, for any $(a,b) \in A(i) \times B(i)$ and $(a_{j_m}, b_{j_m}) \in A(j_m) \times B(j_m)$, for $m = 1, \ldots, n-1$,

$$q_{ij_1}(a,b)q_{j_1j_2}(a_{j_1}, b_{j_1}) \cdots q_{j_{n-1}j_n}(a_{j_{n-1}}, b_{j_{n-1}}) > 0.$$

This assumption ensures that, given $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$, the Markov process $\{x(t, \pi^1, \pi^2)\}_{t \geq 0}$ is irreducible and thus its unique invariant probability measure, denoted μ_{π^1,π^2}, verifies

$$\mu_{\pi^1,\pi^2}\{i\} > 0 \quad \text{for every } i \in S. \tag{2.3}$$

Assumption C also implies the uniform exponential ergodic property (4.1) below; see [5].

To conclude this section we introduce some more notation. Let w be as in Assumption A and denote by $\mathbb{B}_w(S)$ the Banach space of real-valued functions u on S with finite w-norm defined as

$$||u||_w := \sup_{i \in S}\{|u(i)|/w(i)\}.$$

Our assumptions guarantee that, for every pair of strategies $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$, $\sum_{j \in S} w(j) \mu_{\pi^1,\pi^2}\{j\} < \infty$ and thus, for $u \in \mathbb{B}_w(S)$,

$$\mu_{\pi^1,\pi^2}(u) := \int_S u \, d\mu_{\pi^1,\pi^2}$$

is finite.

3. Average optimality criterion

Given $T > 0$ and a pair of stationary strategies $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$ define the total expected payoff of (π^1, π^2) over the time interval $[0, T]$ when the initial state is $i \in S$ as

$$J_T(i, \pi^1, \pi^2) := E_i^{\pi^1,\pi^2}[\int_0^T r(x(t, \pi^1, \pi^2), \pi^1, \pi^2) dt]. \tag{3.1}$$

By Lemma 7.1(a) in [7], the expectation and the integral in (3.1) are interchangeable. The average payoff of the pair (π^1, π^2) is then defined as

$$J(i, \pi^1, \pi^2) := \limsup_{T \to \infty} \frac{J_T(i, \pi^1, \pi^2)}{T} \quad \text{for } i \in S. \tag{3.2}$$

By standard arguments it follows that $J(i, \pi^1, \pi^2) = \mu_{\pi^1,\pi^2}(r(\cdot, \pi^1, \pi^2))$, that does not depend on the initial state i. Therefore, we will simply write (3.2) as $J(\pi^1, \pi^2)$.

Observe that, when dealing with the average payoff criterion, the situation is greatly simplified by just considering the family of stationary policies and, as shown in [7], we can indeed restrict our attention to stationary strategies without loss of generality.

We define the value of the game (for the average reward/cost criterion) as

$$V^* := \sup_{\pi^1 \in \Pi^1} \inf_{\pi^2 \in \Pi^2} J(\pi^1, \pi^2) = \inf_{\pi^2 \in \Pi^2} \sup_{\pi^1 \in \Pi^1} J(\pi^1, \pi^2),$$

which is well defined; see [7, Theorem 5.1(c)].

Definition 3.1. *Consider the stochastic game $\mathscr{M}$. We say that a pair of stationary strategies $(\pi^{*1}, \pi^{*2}) \in \Pi^1 \times \Pi^2$ is average optimal if*

$$J(\pi^1, \pi^{*2}) \le J(\pi^{*1}, \pi^{*2}) \le J(\pi^{*1}, \pi^2) \quad \text{for every } (\pi^1, \pi^2) \in \Pi^1 \times \Pi^2. \tag{3.3}$$

The set of average optimal strategies is denoted $\Pi^{*1} \times \Pi^{*2}$.

It is worth noting that if (π^{*1}, π^{*2}) is average optimal then $J(\pi^{*1}, \pi^{*2}) = V^*$, though the converse is not necessarily true. Observe also that the notation $\Pi^{*1} \times \Pi^{*2}$ suggests that the set of average optimal strategies is a rectangle in $\Pi^1 \times \Pi^2$. In fact, this property turns out to be true as a consequence of Theorem 3.3(ii) below. (See Lemma 4.6 and the paragraph after it.)

Now we introduce the so-called average optimality equations. For ease of notation we shall write, for $i, j \in S$, $(\phi, \psi) \in \overline{A}(i) \times \overline{B}(i)$ and $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$,

$$q_{ij}(\phi, \pi^2) := q_{ij}(\phi, \pi^2(\cdot|i)) \quad \text{and} \quad q_{ij}(\pi^1, \psi) := q_{ij}(\pi^1(\cdot|i), \psi),$$

and also

$$r(i, \phi, \pi^2) := r(i, \phi, \pi^2(\cdot|i)) \quad \text{and} \quad r(i, \pi^1, \psi) := r(i, \pi^1(\cdot|i), \psi).$$

Definition 3.2. *We say that a constant $g \in \mathbb{R}$, a function $h^0 \in \mathbb{B}_w(S)$ and a pair of strategies $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$ verify the average optimality equations if*

$$g = r(i, \pi^1, \pi^2) + \sum_{j \in S} q_{ij}(\pi^1, \pi^2) h^0(j) \tag{3.4}$$

$$= \sup_{\phi \in \overline{A}(i)} \{r(i, \phi, \pi^2) + \sum_{j \in S} q_{ij}(\phi, \pi^2) h^0(j)\} \tag{3.5}$$

$$= \inf_{\psi \in \overline{B}(i)} \{r(i, \pi^1, \psi) + \sum_{j \in S} q_{ij}(\pi^1, \psi) h^0(j)\}, \tag{3.6}$$

for every $i \in S$.

Our next theorem summarizes some useful results about the average optimality equations.

Theorem 3.3. *Suppose that the game model $\mathscr{M}$ verifies Assumptions A, B and C. Then:*

(i) *There exist solutions to the average optimality equations (3.4)–(3.6). Moreover, the constant $g = V^*$ (the value of the game) and the function h^0 is unique up to additive constants.*

(ii) *A pair of strategies is average optimal if and only if it satisfies the average optimality equations.*

Proof. (i). The first statement in (i) as well as the fact that $g = V^*$ is proved in [7, Theorem 5.1]. Let us show that h^0 is unique up to an additive constant. Suppose that (V^*, h^*) and (π^{*1}, π^{*2}), and also $(V^*, \overline{h})$ and $(\overline{\pi}^1, \overline{\pi}^2)$, satisfy (3.4)–(3.6). Then we have

$$V^* = \sup_{\pi^1 \in \Pi^1} J(\pi^1, \pi^{*2}) \quad \text{and} \quad V^* = \inf_{\pi^2 \in \Pi^2} J(\overline{\pi}^1, \pi^2),$$

and thus

$$V^* = J(\overline{\pi}^1, \pi^{*2}). \tag{3.7}$$

On the other hand, we know that

$$V^* = \sup_{\phi \in \overline{A}(i)} \{r(i, \phi, \pi^{*2}) + \sum_{j \in S} q_{ij}(\phi, \pi^{*2}) h^*(j)\} \quad \text{for } i \in S$$

and, in particular,

$$V^* \geq r(i, \overline{\pi}^1, \pi^{*2}) + \sum_{j \in S} q_{ij}(\overline{\pi}^1, \pi^{*2}) h^*(j) \quad \text{for } i \in S.$$

If the strict inequality holds in any of the above inequalities then, multiplying by $\mu_{\overline{\pi}^1, \pi^{*2}}\{i\}$, which is positive (recall (2.3)), and summing over $i \in S$ yields $V^* > J(\overline{\pi}^1, \pi^{*2})$, contradicting (3.7). Therefore,

$$V^* = r(i,\overline{\pi}^1,\pi^{*2}) + \sum_{j\in S} q_{ij}(\overline{\pi}^1,\pi^{*2})h^*(j) \quad \text{for each } i \in S.$$

Mutatis mutandis we obtain

$$V^* = r(i,\overline{\pi}^1,\pi^{*2}) + \sum_{j\in S} q_{ij}(\overline{\pi}^1,\pi^{*2})\overline{h}(j) \quad \text{for each } i \in S.$$

Hence, the functions h^* and $\overline{h}$ verify

$$\sum_{j\in S} q_{ij}(\overline{\pi}^1,\pi^{*2})(h^*(j) - \overline{h}(j)) = 0 \quad \text{for every } i \in S,$$

that is, $h^* - \overline{h}$ is harmonic and, as in the proof of [12, Theorem 3.3], this implies that h^* and $\overline{h}$ differ by a constant.

(ii). The *if* part is established in [7, Theorem 5.1(d)]. To prove the *only if* statement proceed by contradiction. Suppose that (π^{*1},π^{*2}) is a pair of average optimal strategies that does not verify the average optimality equations. Then, either (3.5) or (3.6) do not hold. Suppose for instance that (3.5) is not satisfied. We have, by Theorem 5.1(c) and Lemma 7.2 in [7], that there exists $\pi^1 \in \Pi^1$ such that

$$V^* \leq r(i,\pi^1,\pi^{*2}) + \sum_{j\in S} q_{ij}(\pi^1,\pi^{*2})h^0(j) \quad \text{for every } i \in S$$

with strict inequality for some $i \in S$. Multiplying by the invariant probability measure $\mu_{\pi^1,\pi^{*2}}$ and summing over $i \in S$ yields $V^* < J(\pi^1,\pi^{*2})$, and then, by Definition 3.1, we obtain $V^* < J(\pi^1,\pi^{*2}) \leq J(\pi^{*1},\pi^{*2}) = V^*$, which is not possible. This completes the proof. □

In stochastic control theory, strategies that satisfy (3.4)–(3.6) are called *canonical*. Hence Theorem 3.3 proves the equivalence between average optimal strategies and canonical strategies. It is a well known fact that, in general, there might exist optimal strategies that are not canonical. In our case, the irreducibility of the state Markov processes (recall Assumption C) implies that both classes coincide.

4. Bias optimality

In this section we are going to define the *bias* of a pair of stationary policies. We will give an interpretation of the bias in terms of the total expected reward/cost over finite time intervals as the time horizon goes to ∞ and we will introduce the bias optimality criterion.

The extended game model. For technical reasons it is useful to consider the stochastic game model $\overline{\mathscr{M}}$ in which the admissible control actions correspond to the randomized actions in model $\mathscr{M}$. More precisely, let

$$\overline{\mathscr{M}} := \{S, (\overline{A}(i), \overline{B}(i), i \in S), (q_{ij}(\phi,\psi)), (r(i,\phi,\psi))\}.$$

Proposition 4.1. *If the game model $\mathscr{M}$ verifies Assumptions A, B and C, then so does $\overline{\mathscr{M}}$.*

Proof. First of all observe that the transition rates of the system, i.e. the $q_{ij}(\phi, \psi)$, are measurable on $\overline{A}(i) \times \overline{B}(i)$ and that they are conservative and stable. The reward/cost rate function is also measurable.

Assumption A for $\overline{\mathscr{M}}$, with the same constants as for $\mathscr{M}$, is easily derived from (2.1) and (2.2).

Since $A(i)$ and $B(i)$, for $i \in S$, are compact Borel spaces then $\overline{A}(i)$ and $\overline{B}(i)$ (endowed with the weak convergence topology) are also compact Borel spaces for each $i \in S$. Hence, Assumption B(i) is satisfied. Assumption B(ii) is a consequence of [7, Lemma 7.2]. Assumption B(iii) is easily verified for model $\overline{\mathscr{M}}$.

Finally, it is trivial to check that Assumption C also holds for $\overline{\mathscr{M}}$. This completes the proof. □

Proposition 4.1 implies that the randomized stationary policies (and not just the deterministic stationary policies) verify the w-uniform exponential ergodic property, that is, there exists a constant $R > 0$ such that

$$\sup_{(\pi^1,\pi^2)\in\Pi^1\times\Pi^2} |E_i^{\pi^1,\pi^2} u(x(t,\pi^1,\pi^2)) - \mu_{\pi^1,\pi^2}(u)| \leq Re^{-ct}||u||_w w(i) \tag{4.1}$$

for each $i \in S$, $t \geq 0$ and $u \in \mathbb{B}_w(S)$, where the constant $c > 0$ is as in Assumption A.

The bias and the Poisson equations. We define the *bias* of $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$ as the function $\{h^0(i, \pi^1, \pi^2)\}_{i\in S} \in \mathbb{B}_w(S)$ given by

$$h^0(i,\pi^1,\pi^2) := E_i^{\pi^1,\pi^2} \int_0^\infty [r(x(t,\pi^1,\pi^2),\pi^1,\pi^2) - J(\pi^1,\pi^2)]dt \quad \text{for } i \in S. \tag{4.2}$$

Observe that (4.1) ensures that $h^0(\cdot, \pi^1, \pi^2)$ is indeed in $\mathbb{B}_w(S)$.

The bias of a stationary policy can be computed via the Poisson equations defined next.

Given a pair of stationary strategies (π^1, π^2) we say that $g \in \mathbb{R}$ and $h^0, h^1 \in \mathbb{B}_w(S)$ are a solution of the *Poisson equations* for (π^1, π^2) if

$$g = r(i,\pi^1,\pi^2) + \sum_{j\in S} q_{ij}(\pi^1,\pi^2)h^0(j) \quad \text{for every } i \in S \tag{4.3}$$

and

$$h^0(i) = \sum_{j\in S} q_{ij}(\pi^1,\pi^2)h^1(j) \quad \text{for every } i \in S. \tag{4.4}$$

The average payoff $J(\pi^1, \pi^2)$ and the bias of (π^1, π^2) are characterized by the Poisson equations in the following sense.

Proposition 4.2. *Let $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$ be given.*

(i) *The solution $(g, h^0) \in \mathbb{R} \times \mathbb{B}_w(S)$ of the Poisson equation (4.3) exists. Also, $g = J(\pi^1, \pi^2)$ and, moreover, $h^0(\cdot)$ coincides with the bias $h^0(\cdot, \pi^1, \pi^2)$ except for an additive constant, that is, there exists $z \in \mathbb{R}$ such that*

$$h^0(i) + z = h^0(i,\pi^1,\pi^2) \quad \text{for every } i \in S.$$

If in addition $\mu_{\pi^1,\pi^2}(h^0) = 0$, then $h^0(\cdot) = h^0(\cdot, \pi^1, \pi^2)$.

(ii) *The solution* $(g, h^0, h^1) \in \mathbb{R} \times \mathbb{B}_w(S) \times \mathbb{B}_w(S)$ *of the Poisson equations (4.3)–(4.4) exists and it verifies*

$$g = J(\pi^1, \pi^2) \quad \textit{and} \quad h^0(\cdot) = h^0(\cdot, \pi^1, \pi^2).$$

Proof. The proof goes along the same lines as that of [12, Proposition 3.4]. See also the proof of [13, Theorem 4.1]. □

Bias optimal policies. From (4.1) and the definition (4.2) of the bias we obtain that

$$J_T(i, \pi^1, \pi^2) = J(\pi^1, \pi^2)T + h^0(i, \pi^1, \pi^2) + \mathrm{O}(e^{-cT}) \tag{4.5}$$

as $T \to \infty$ for every $i \in S$ and $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$.

Therefore, to asymptotically maximize the total expected reward over finite time intervals, player 1 should attempt to maximize the average reward $J(\pi^1, \pi^2)$, for fixed $\pi^2 \in \Pi^2$, and then maximize the bias $h^0(\cdot, \pi^1, \pi^2)$ within the class of average optimal policies. Player 2 defines similarly his/her bias criterion. Now we give the precise definition of bias optimality.

Definition 4.3. *Consider the stochastic game* $\mathcal{M}$. *We say that a pair of average optimal stationary strategies* $(\pi^{*1}, \pi^{*2}) \in \Pi^{*1} \times \Pi^{*2}$ *is bias optimal if*

$$h^0(i, \pi^1, \pi^{*2}) \leq h^0(i, \pi^{*1}, \pi^{*2}) \leq h^0(i, \pi^{*1}, \pi^2) \tag{4.6}$$

for every $i \in S$ *and every pair of average optimal strategies* $(\pi^1, \pi^2) \in \Pi^{*1} \times \Pi^{*2}$.

Before proceeding to prove the existence of bias optimal policies we need two more preliminary results. The following result is perhaps well known but we could not find a reference. Hence we will provide a proof.

Lemma 4.4. *Let* $f : X \times Y \to \mathbb{R}$ *be a continuous function, where* X *and* Y *are compact Borel spaces. Then* $g : Y \to \mathbb{R}$ *given by* $g(y) := \max_{x \in X} f(x, y)$ *is continuous.*

Proof. To prove the continuity of g we will proceed by contradiction. Suppose that there exists $y \in Y$ and a sequence $\{y_n\}$ in Y verifying $y_n \to y$ when $n \to \infty$ and $g(y_n) \not\to g(y)$. Also, let $x \in X$ and $x_n \in X$ be such that $g(y) = f(x, y)$ and $g(y_n) = f(x_n, y_n)$. There exist $\varepsilon > 0$ and subsequences of $\{x_n\}$ and $\{y_n\}$ (not explicit in the notation) for which

$$|g(y_n) - g(y)| \geq \varepsilon \quad \text{and} \quad x_n \to x^*, \quad \text{for some } x^* \in X.$$

Hence, one of the following two conditions is satisfied: (i) $g(y_n) \geq g(y) + \varepsilon$ for infinitely many n; or (ii) $g(y) \geq g(y_n) + \varepsilon$ for infinitely many n.

Suppose that (i) holds. Then $g(y_n) = f(x_n, y_n)$ converges to $f(x^*, y)$, and thus $f(x^*, y) \geq g(y) + \varepsilon$, which is a contradiction. Else if (ii) holds, and since $f(x, y_n) \to g(y)$, observe that

$$f(x, y_n) \geq g(y_n) + \varepsilon/2 = f(x_n, y_n) + \varepsilon/2$$

infinitely often, which contradicts the definition of x_n. This establishes the stated result. □

Definition 4.5. *Consider the game model* $\mathcal{M}$ *and let* $(V^*, h^0) \in \mathbb{R} \times \mathbb{B}_w(S)$ *be a solution of the average optimality equations (3.4)–(3.6). Fix* $i \in S$ *and let* $\overline{A}_0(i)$ *be the set of* $\phi \in \overline{A}(i)$ *such that*

$$V^* = \inf_{\psi \in \overline{B}(i)} \{r(i,\phi,\psi) + \sum_{j \in S} q_{ij}(\phi,\psi)h^0(j)\}.$$

Define also $\overline{B}_0(i)$ as the set of $\psi \in \overline{B}(i)$ for which

$$V^* = \sup_{\phi \in \overline{A}(i)} \{r(i,\phi,\psi) + \sum_{j \in S} q_{ij}(\phi,\psi)h^0(j)\}.$$

Lemma 4.6. *The sets $\overline{A}_0(i)$ and $\overline{B}_0(i)$ in Definition 4.5 are convex compact Borel spaces for every $i \in S$ and, further, they do not depend on h^0.*

Proof. First of all, let us prove that $\overline{A}_0(i)$ and $\overline{B}_0(i)$ are compact Borel spaces. By Assumption B(i), the sets $A(i)$ and $B(i)$ are compact Borel spaces. Therefore, $\overline{A}(i)$ and $\overline{B}(i)$ are also compact Borel spaces. Thus, to prove our statement, it suffices to show that $\overline{A}_0(i)$ and $\overline{B}_0(i)$ are closed sets, which is true as a consequence of Lemma 4.4 and Lemma 7.2 in [7].

Let us now show that $\overline{A}_0(i)$ is a convex set. To this end observe that, for each $\psi \in \overline{B}(i)$, the function $\phi \mapsto r(i,\phi,\psi)$ is linear in the following sense:

$$r(i, \lambda\phi^1 + (1-\lambda)\phi^2, \psi) = \lambda r(i,\phi^1,\psi) + (1-\lambda) r(i,\phi^2,\psi)$$

for $\phi^1, \phi^2 \in \overline{A}(i)$ and $\lambda \in [0,1]$, and where $\lambda\phi^1 + (1-\lambda)\phi^2$ is a convex linear combination of probability measures on $A(i)$, which is itself a probability measure in $\overline{A}(i)$. Similarly we have that for a given $\psi \in \overline{B}(i)$, $\phi \mapsto \sum_{j \in S} q_{ij}(\phi,\psi)h^0(j)$ is linear. Therefore,

$$\phi \mapsto \inf_{\psi \in \overline{B}(i)} \{r(i,\phi,\psi) + \sum_{j \in S} q_{ij}(\phi,\psi)h^0(j)\},$$

which is the infimum of linear functions, is concave. On the other hand,

$$V^* = \sup_{\phi \in \overline{A}(i)} \inf_{\psi \in \overline{B}(i)} \{r(i,\phi,\psi) + \sum_{j \in S} q_{ij}(\phi,\psi)h^0(j)\},$$

and thus $\overline{A}_0(i)$ is the set of maxima of a concave function and so $\overline{A}_0(i)$ is convex.

Using the same arguments one can show that

$$\psi \mapsto \sup_{\phi \in \overline{A}(i)} \{r(i,\phi,\psi) + \sum_{j \in S} q_{ij}(\phi,\psi)h^0(j)\}$$

is convex and thus $\overline{B}_0(i)$ is convex.

To conclude the proof, observe that the solution h^0 of the average optimality equations (3.4)–(3.6) is unique up to additive constants (by Theorem 3.3(i)) and since the transition rates of $\overline{\mathcal{M}}$ are conservative (recall the proof of Proposition 4.1) then $\overline{A}_0(i)$ and $\overline{B}_0(i)$ do not depend on h^0. □

By Theorem 3.3(ii), (π^1, π^2) is in $\Pi^{*1} \times \Pi^{*2}$ if and only if $\pi^1(\cdot|i) \in \overline{A}_0(i)$ and $\pi^2(\cdot|i) \in \overline{B}_0(i)$ for each $i \in S$, and this justifies the use of the rectangle notation $\Pi^{*1} \times \Pi^{*2}$.

Suppose that $(V^*, h^0) \in \mathbb{R} \times \mathbb{B}_w(S)$ is a solution of the average optimality equations. To analyze the bias optimality criterion consider now the stochastic game $\overline{\mathcal{M}}_0$ with state space S, admissible actions $\overline{A}_0(i)$ and $\overline{B}_0(i)$, for $i \in S$, and the same transition rates as $\overline{\mathcal{M}}$. To determine the reward/cost rate in $\overline{\mathcal{M}}_0$ observe that by Proposition 4.2(i)

$$h^0(\cdot,\pi^1,\pi^2) = h^0(\cdot) + z \quad \text{for some } z \in \mathbb{R} \tag{4.7}$$

and, therefore,

$$z = \mu_{\pi^1,\pi^2}(-h^0), \tag{4.8}$$

where (π^1,π^2) is in $\Pi^{*1} \times \Pi^{*2}$. Consequently, to find bias optimal policies it suffices to consider the stochastic game $\overline{\mathscr{M}}_0$ with reward/cost rate $-h^0$ under the expected average reward/cost criterion. Summarizing, $\overline{\mathscr{M}}_0$ is defined as

$$\overline{\mathscr{M}}_0 := \{S, (\overline{A}_0(i), \overline{B}_0(i), i \in S), (q_{ij}(\phi,\psi)), (-h^0(i))\},$$

and observe that $\overline{\mathscr{M}}_0$ satisfies Assumptions A, B and C. In particular, the average value V^{0*} of $\overline{\mathscr{M}}_0$ exists and then, by (4.7) and (4.8),

$$H^*(i) := \inf_{\pi^2\in\Pi^{*2}} \sup_{\pi^1\in\Pi^{*1}} h^0(i,\pi^1,\pi^2) = \sup_{\pi^1\in\Pi^{*1}} \inf_{\pi^2\in\Pi^{*2}} h^0(i,\pi^1,\pi^2) = h^0(i) + V^{0*} \tag{4.9}$$

for every $i \in S$. Note that $V^{0*} \equiv V^{0*}(h^0)$ depends on the particular solution h^0 of the average optimality equations, though $h^0 + V^{0*}(h^0)$ does not depend on h^0.

The bias optimality equations. We give a characterization of bias optimal policies via the bias optimality equations defined below.

Definition 4.7. *We say that $g \in \mathbb{R}$, $h^0, h^1 \in \mathbb{B}_w(S)$ and $(\pi^1,\pi^2) \in \Pi^1 \times \Pi^2$ verify the bias optimality equations if*

$$g = r(i,\pi^1,\pi^2) + \sum_{j\in S} q_{ij}(\pi^1,\pi^2)h^0(j) \tag{4.10}$$

$$= \sup_{\phi\in\overline{A}(i)} \{r(i,\phi,\pi^2) + \sum_{j\in S} q_{ij}(\phi,\pi^2)h^0(j)\} \tag{4.11}$$

$$= \inf_{\psi\in\overline{B}(i)} \{r(i,\pi^1,\psi) + \sum_{j\in S} q_{ij}(\pi^1,\psi)h^0(j)\}, \tag{4.12}$$

for every $i \in S$ and, moreover,

$$0 = -h^0(i) + \sum_{j\in S} q_{ij}(\pi^1,\pi^2)h^1(j) \tag{4.13}$$

$$= \sup_{\phi\in\overline{A}_0(i)} \{-h^0(i) + \sum_{j\in S} q_{ij}(\phi,\pi^2)h^1(j)\} \tag{4.14}$$

$$= \inf_{\psi\in\overline{B}_0(i)} \{-h^0(i) + \sum_{j\in S} q_{ij}(\pi^1,\psi)h^1(j)\}, \tag{4.15}$$

for every $i \in S$.

Theorem 4.8. *Suppose that the game model $\mathscr{M}$ verifies Assumptions A, B and C. Then the following holds.*

(i) *The solutions of the bias optimality equations exist and, further,*

$$g = V^* \quad \text{and} \quad h^0(i) = H^*(i) \quad \text{for every } i \in S.$$

(ii) The stationary strategies $(\pi^1,\pi^2) \in \Pi^1 \times \Pi^2$ *are bias optimal if and only if they verify the bias optimality equations.*

Proof. (i). The equations (4.10)–(4.12) are the average optimality equations. By Theorem 3.3, they have a solution and we know that $g = V^*$.

Concerning equations (4.13)–(4.15), observe that the stochastic game $\overline{\mathscr{M}}_0$ verifies the hypotheses of [7, Theorem 5.1], though average optimal strategies for $\overline{\mathscr{M}}_0$ are randomized actions, that is, they are probability measures on $\overline{A}_0(i)$ and $\overline{B}_0(i)$ or, in other words, they belong to $\mathscr{P}(\mathscr{P}(A(i)))$ and $\mathscr{P}(\mathscr{P}(B(i)))$, respectively. Nevertheless, the convexity property proved in Lemma 4.6 allows us to "stay" in $\mathscr{P}(A(i))$ and $\mathscr{P}(B(i))$.

Indeed, for a given $i \in S$, define the following projection operator $\mathbf{p} : \mathscr{P}(\overline{A}_0(i)) \to \overline{A}_0(i)$ where, for $\overline{\phi} \in \mathscr{P}(\overline{A}_0(i))$, $\mathbf{p}\overline{\phi}$ is a probability measure on $A(i)$ defined by

$$(\mathbf{p}\overline{\phi})(F) := \int\limits_{\overline{A}_0(i)} \phi(F)\overline{\phi}(d\phi) \quad \text{for each measurable set } F \subseteq A(i).$$

Note that $\mathbf{p}\overline{\phi} \in \overline{A}_0(i)$ because $\overline{A}_0(i)$ is a convex set of probability measures; see Lemma 4.6. Similarly, we may define $\mathbf{p} : \mathscr{P}(\overline{B}_0(i)) \to \overline{B}_0(i)$ where, for simplicity, we will use the same notation.

Observe also that the transition rates of $(\overline{\phi}, \overline{\psi}) \in \mathscr{P}(\overline{A}_0(i)) \times \mathscr{P}(\overline{B}_0(i))$, i.e.

$$q_{ij}(\overline{\phi}, \overline{\psi}) := \int\limits_{\overline{B}_0(i)} \int\limits_{\overline{A}_0(i)} q_{ij}(\phi, \psi)\ \overline{\phi}(d\phi)\ \overline{\psi}(d\psi)$$

verify

$$q_{ij}(\overline{\phi}, \overline{\psi}) = q_{ij}(\mathbf{p}\overline{\phi}, \mathbf{p}\overline{\psi}) \quad \text{for } j \in S.$$

A similar result is true for the reward/cost rate $r(i, \overline{\phi}, \overline{\psi})$. Identifying $\phi \in \overline{A}_0(i)$ with the Dirac measure concentrated at ϕ, we have $\overline{A}_0(i) \subseteq \mathscr{P}(\overline{A}_0(i))$, and also $\overline{B}_0(i) \subseteq \mathscr{P}(\overline{B}_0(i))$.

Therefore, there exists a correspondence from $\mathscr{P}(\overline{A}_0(i))$ (resp. $\mathscr{P}(\overline{B}_0(i))$) onto $\overline{A}_0(i)$ (resp. $\overline{B}_0(i)$) with invariant transition and reward/cost rates, and thus invariant expected rewards/costs. As a consequence, the average optimality equations of $\overline{\mathscr{M}}_0$ may be written as

$$\begin{aligned} V^{0*} &= -\,h^0(i) + \sum_{j\in S} q_{ij}(\pi^1, \pi^2)h^1(j) \\ &= \sup_{\phi\in\overline{A}_0(i)} \{-h^0(i) + \sum_{j\in S} q_{ij}(\phi, \pi^2)h^1(j)\} \\ &= \inf_{\psi\in\overline{B}_0(i)} \{-h^0(i) + \sum_{j\in S} q_{ij}(\pi^1, \psi)h^1(j)\}, \end{aligned}$$

for $i \in S$ and some $(\pi^1, \pi^2) \in \Pi^{*1} \times \Pi^{*2}$. Hence, from (4.9) we deduce that H^* verifies equations (4.13)–(4.15). To prove the uniqueness property observe that h^0 in (4.10)–(4.12) is unique up to additive constants (recall Theorem 3.3(i)) and, therefore, the solution h^0 of (4.13)–(4.15) is necessarily unique and coincides with H^*.

(ii). This part follows from the equivalence between optimal and canonical policies of $\overline{\mathscr{M}}_0$ that was established in Theorem 3.3(ii). □

Expression (4.5) shows that bias optimality is related to the asymptotic maximization/minimization of the total expected reward/cost $J_T(i, \pi^1, \pi^2)$. This relation is further explored in the next section.

5. Weak overtaking optimality

We introduce the weak overtaking optimality criterion for continuous-time stochastic games, which is the extension to continuous-time games of the discrete-time definition given in [10, Definition 3].

Roughly speaking, we say that a pair of strategies is weakly overtaking optimal if, in the limit as $T \to \infty$, it is a saddle point of the finite-horizon total expected payoff $J_T(i, \pi^1, \pi^2)$. This is formalized below.

Definition 5.1. *A pair of strategies* $(\pi^{*1}, \pi^{*2}) \in \Gamma^1 \times \Gamma^2 \subseteq \Pi^1 \times \Pi^2$ *is weakly overtaking optimal in the class* $\Gamma^1 \times \Gamma^2$ *if for each* $(\pi^1, \pi^2) \in \Gamma^1 \times \Gamma^2$ *and* $i \in S$ *we have*

$$\liminf_{T\to\infty}[J_T(i, \pi^{*1}, \pi^{*2}) - J_T(i, \pi^1, \pi^{*2})] \geq 0 \tag{5.1}$$

and

$$\limsup_{T\to\infty}[J_T(i, \pi^{*1}, \pi^{*2}) - J_T(i, \pi^{*1}, \pi^2)] \leq 0.$$

Our next two results explore the relations existing between bias optimality and weak overtaking optimality. We then present an example showing that these relations are not as "strong" as for control (or single-player) problems. In fact, for continuous-time *controlled* Markov chains and under assumptions similar to ours, bias optimality and weak overtaking optimality are *equivalent*; see [12, Theorem 3.8].

Theorem 5.2. *Suppose that the stochastic game* $\mathcal{M}$ *verifies Assumptions A, B and C. If a pair of strategies* $(\pi^{*1}, \pi^{*2}) \in \Pi^1 \times \Pi^2$ *is bias optimal then it is weakly overtaking optimal in the class of average optimal strategies* $\Pi^{*1} \times \Pi^{*2}$.

Proof. Let $(\pi^1, \pi^2) \in \Pi^{*1} \times \Pi^{*2}$ be a pair of average optimal strategies. Recalling (4.5) we have

$$J_T(i, \pi^1, \pi^2) = V^*T + h^0(i, \pi^1, \pi^2) + \mathrm{O}(e^{-cT}) \quad \text{for all } i \in S,$$

and thus Definition 4.3 yields

$$\lim_{T\to\infty}[J_T(i, \pi^{*1}, \pi^{*2}) - J_T(i, \pi^1, \pi^{*2})] = h^0(i, \pi^{*1}, \pi^{*2}) - h^0(i, \pi^1, \pi^{*2}) \geq 0$$

and also

$$\lim_{T\to\infty}[J_T(i, \pi^{*1}, \pi^{*2}) - J_T(i, \pi^{*1}, \pi^2)] = h^0(i, \pi^{*1}, \pi^{*2}) - h^0(i, \pi^{*1}, \pi^2) \leq 0$$

for every $i \in S$. This proves that bias optimal strategies are weakly overtaking optimal in $\Pi^{*1} \times \Pi^{*2}$. □

Theorem 5.3 *Suppose that the stochastic game* $\mathcal{M}$ *verifies Assumptions A, B and C. If a pair of strategies* $(\pi^{*1}, \pi^{*2}) \in \Pi^1 \times \Pi^2$ *is weakly overtaking optimal in* $\Pi^1 \times \Pi^2$ *then it is bias optimal.*

Proof. Using (4.5) and recalling Definition 5.1 it follows that

$$\lim_{T\to\infty}[(J(\pi^{*1}, \pi^{*2}) - J(\pi^1, \pi^{*2}))T + h^0(i, \pi^{*1}, \pi^{*2}) - h^0(i, \pi^1, \pi^{*2})] \geq 0 \tag{5.2}$$

and

$$\lim_{T\to\infty}[(J(\pi^{*1},\pi^{*2})-J(\pi^{*1},\pi^{2}))T+h^0(i,\pi^{*1},\pi^{*2})-h^0(i,\pi^{*1},\pi^{2})]\le 0 \quad (5.3)$$

for every $(\pi^1,\pi^2)\in\Pi^1\times\Pi^2$ and $i\in S$.

Dividing by T and letting $T\to\infty$ in (5.2) and (5.3) yields precisely condition (3.3) in Definition 3.1, that is, (π^{*1},π^{*2}) is average optimal. Suppose now that $(\pi^1,\pi^2)\in\Pi^{*1}\times\Pi^{*2}$. Then (5.2) and (5.3) become (4.6) in Definition 4.3, completing the proof. □

An example. The result of Theorem 5.2 cannot be extended to weak overtaking optimality in the class of *all* stationary policies or, in other words, the converse of Theorem 5.3 needs not to be true. Indeed, as shown by the example below, there might not exist weakly overtaking optimal policies in the class of stationary policies.

Consider the following zero-sum stochastic game. The state space is

$$S=\{0,1\}$$

and the admissible control actions are

$$A(0)=\{0\},\quad A(1)=\{0,1\},\quad B(0)=\{0\},\quad B(1)=\{0,1\}.$$

The reward/cost rates and the transition rates are given by

$$r(0,0,0)=4,\ r(1,0,0)=1,\ r(1,0,1)=-2,\ r(1,1,0)=0,\ r(1,1,1)=2$$

and

$$q_{00}(0,0)=-2,\ q_{11}(0,0)=-1,\ q_{11}(0,1)=q_{11}(1,0)=q_{11}(1,1)=-2,$$

respectively.

Randomized stationary policies for player 1, denoted π_x^1, are parametrized by $x\in[0,1]$, where $\pi_x^1(\cdot|1)$ takes values 0 and 1 with probabilities x and $1-x$, respectively. We will denote by π_y^2, with $0\le y\le 1$, a randomized stationary strategy for player 2, where $\pi_y^2(\cdot|1)$ takes values 0 and 1 with probabilities y and $1-y$, respectively.

It is easily verified that the so-defined game model satisfies Assumptions A, B and C in Section 2.

Let us compute the average reward/cost of the stationary policies (π_x^1,π_y^2), for x and y in $[0,1]$. Direct calculations show that the expected reward/cost rates for stationary policies are

$$r(0,\pi_x^1,\pi_y^2)=4 \quad\text{and}\quad r(1,\pi_x^1,\pi_y^2)=5xy-4x-2y+2$$

whereas the transition rates matrices are

$$Q(\pi_x^1,\pi_y^2)=\begin{pmatrix}-2 & 2\\ 2-xy & xy-2\end{pmatrix}.$$

Hence the invariant probability measures are given by

$$\mu_{\pi_x^1,\pi_y^2}\{0\}=\frac{2-xy}{4-xy}\quad\text{and}\quad \mu_{\pi_x^1,\pi_y^2}\{1\}=\frac{2}{4-xy}\quad\text{for } 0\le x,y\le 1,$$

and thus the expected reward/cost of the stationary strategy (π_x^1,π_y^2) is

$$J(x,y) := J(\pi^1_x, \pi^2_y) = \frac{6xy - 8x - 4y + 12}{4 - xy}.$$

Now we determine the set of average optimal policies. For a fixed $x \in [0,1]$ we have

$$\inf_{0 \le y \le 1} J(x,y) = \begin{cases} J(x,1) = 2, & \text{for } 0 \le x < 1/2, \\ J(x,y) = 2, & \text{for } x = 1/2 \text{ and } 0 \le y \le 1, \\ J(x,0) = 3 - 2x, & \text{for } 1/2 < x \le 1, \end{cases}$$

and given $y \in [0,1]$

$$\sup_{0 \le x \le 1} J(x,y) = \begin{cases} J(0,y) = 3 - y, & \text{for } 0 \le y < 1, \\ J(x,y) = 2, & \text{for } y = 1 \text{ and } 0 \le x \le 1. \end{cases}$$

As a consequence, the value of the game is $V^* = 2$. It also follows that the family of optimal stationary strategies, which is given by the (x^*, y^*) such that

$$J(x^*, y^*) = \inf_{0 \le y \le 1} J(x^*, y) = \sup_{0 \le x \le 1} J(x, y^*) = V^* = 2,$$

is $(\pi^1_{x^*}, \pi^2_1)$ for $0 \le x^* \le 1/2$.

Consider now the stationary policy (π^1_x, π^2_1) for some $0 \le x \le 1$. Observe that the average reward/cost and the bias of this policy are

$$J(x,1) = 2 \quad \text{and} \quad h(\cdot, x, 1) := h(\cdot, \pi^1_x, \pi^2_1) = \begin{pmatrix} \frac{2}{4-x} \\ \frac{x-2}{4-x} \end{pmatrix}, \tag{5.4}$$

respectively. The unique bias optimal stationary policy is $(\pi^1_{1/2}, \pi^2_1)$. Its gain and bias are

$$J(1/2, 1) = 2 \quad \text{and} \quad h(\cdot, 1/2, 1) = \begin{pmatrix} 4/7 \\ -3/7 \end{pmatrix}. \tag{5.5}$$

It is worth noting that since player 2 has a unique average optimal strategy then the problem of finding bias optimal policies is reduced to a control (with one player) problem.

Suppose now that there exists a weakly overtaking optimal policy in the class of all stationary strategies for the above game model. By Theorem 5.3, such a policy is necessarily bias optimal and, therefore, $(\pi^1_{1/2}, \pi^2_1)$ would be weakly overtaking optimal. However, recalling (5.4), it follows that the gain and bias of (π^1_1, π^2_1) are

$$J(1,1) = 2 \quad \text{and} \quad h(\cdot, 1, 1) = \begin{pmatrix} 2/3 \\ -1/3 \end{pmatrix}, \tag{5.6}$$

and thus, by (5.5) and (5.6),

$$\liminf_{T \to \infty} \, [J_T(i, \pi^1_{1/2}, \pi^2_1) - J_T(i, \pi^1_1, \pi^2_1)] = -2/21 < 0 \quad \text{for } i \in S,$$

which contradicts (5.1).

As a conclusion, there does not exist any weakly overtaking optimal policy. The reason is that finding bias optimal policies for a game model cannot be reduced to finding bias optimal policies for a control problem. Indeed, when we look for bias optimal policies in the game model we restrict ourselves to the set of game average optimal policies, that is,

$$(\pi_x^1, \pi_1^2) \quad \text{for } 0 \le x \le 1/2. \tag{5.7}$$

However, even if the average optimal policy for player 2 is fixed, the game bias optimization problem is not equivalent to the control bias optimization problem when π_1^2 is fixed, for in this case the control average optimal policies are (recall (5.4))

$$(\pi_x^1, \pi_1^2) \quad \text{for every } x \in [0, 1];$$

cf. (5.7). This is precisely the (erroneous) argument invoked in the proof of Theorem 3 in [10] and Theorem 5 in [11] for discrete-time stochastic games. Note that a similar argument is used in the proof [10, Theorem 2] when dealing with strong 1-equilibria which, in the notation of Section 6 below, would be referred to as 0-strong equilibria.

6. Concluding remarks

In stochastic control, it is usual to consider the so-called sensitive discount optimality criteria as, for instance, n-discount optimality, for $n = -1, 0, 1, \ldots$, and Blackwell optimality (e.g. [13]). Roughly speaking, -1-discount optimality and 0-discount optimality are equivalent to average and bias optimality, respectively. The standard methodology to deal with these sensitive discount optimality criteria is the following.

(i) Solve the average optimality equations to determine -1-discount optimal strategies.

(ii) Find 0-discount optimal strategies in the class of -1-discount optimal policies and prove that they are 0-discount optimal in the class of *all* stationary strategies.

(iii) Find 1-discount optimal strategies in the class of 0-discount optimal policies and prove that they are 1-discount optimal in the class of *all* stationary policies, etc.

We thus obtain a sequence of "nested" control problems which in the limit, under suitable hypotheses, leads to the existence of Blackwell optimal policies.

The fact that the converse of Theorem 5.3 is not verified shows that the above methodology is not applicable to stochastic games. Indeed, Theorem 5.2 gives the existence of 0-discount optimal policies in the class of -1-discount optimal strategies, but not in the class of *all* stationary strategies.

Iteratively, we can find n-discount optimal policies in the class of $(n-1)$-discount optimal policies, but it seems that not in a larger class. As a conclusion, the analysis of sensitive discount optimality criteria appears to be of limited interest in stochastic games.

References

[1] Brock W (1970) On existence of weakly maximal programmes in a multisector economy. Rev. Econom. Studies 37:275–280

[2] Brown BW (1965) On the iterative method of dynamic programming on finite space discrete time Markov processes. Ann. Math. Statist. 36:1279–1285

[3] Carlson D, Haurie A, Leizarowitz A (1994) Overtaking equilibria for switching regulator and tracking games. In: Advances in Dynamic Games and Applications, edited by T. Basar

and A. Haurie, Annals of the International Society of Dynamic Games 1, Birkhauser, Boston, pp. 247–268

[4] Gale D (1967) On optimal development in a multi-sector economy. Rev. Econom. Studies 34:1–19

[5] Guo XP, Hernández-Lerma O (2003) Drift and monotonicity conditions for continuous-time controlled Markov chains. IEEE Trans. Autom. Control 48(2):236–244

[6] Guo XP, Hernández-Lerma O (2003) Nonzero-sum games for continuous-time Markov chains with unbounded discounted payoffs. Submitted

[7] Guo XP, Hernández-Lerma O (2003) Zero-sum games for continuous-time Markov chains with unbounded transition and average payoff rates. J. Appl. Prob. 40:327–345

[8] Hernández-Lerma O, Lasserre JB (1999) Further Topics on Discrete-Time Markov Control Processes. Springer, New York

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

[10] Nowak AS (1999) Sensitive equilibria for ergodic stochastic games with countable state spaces. Math. Methods Oper. Res. 50:65–76

[11] Nowak AS (1999) Optimal strategies in a class of zero-sum ergodic stochastic games. Math. Methods Oper. Res. 50:399–419

[12] Prieto-Rumeau T, Hernández-Lerma O (2003) Bias optimality for continuous-time controlled Markov chains. Technical Report No. 345, Mathematics Dept., CINVESTAV-IPN. (submitted)

[13] Prieto-Rumeau T, Hernández-Lerma O (2005) The Laurent series, sensitive discount and Blackwell optimality for continuous-time controlled Markov chains. Math. Methods Oper. Res. 61: 123–145

[14] Puterman ML (1994) Markov Decision Processes. Wiley, New York

[15] Ramsey FP (1928) A mathematical theory of savings. Econom. J. 38:543–559

[16] Rubinstein A (1979) Equilibria in supergames with the overtaking criterion. J. Economic Theory 21:1–9

[17] Tanaka K, Wakata K (1977) On continuous time Markov games with the expected average reward criterion. Sci. Rep. Niigata Univ. A 14:15–24

[18] von Weizsäcker CC (1965) Existence of optimal programs of accumulation for an infinite horizon. Rev. Econom. Stud. 32:85–104