Bernoulli **11**(6), 2005, 1009–1029

Zero-sum continuous-time Markov games with unbounded transition and discounted payoff rates

XIANPING GUO[1] and ONÉSIMO HERNÁNDEZ-LERMA[2]

[1]*School of Mathematics and Computational Science, Zhongshan University, Guangzhou, 510275, P.R. China. E-mail: mcsgxp@zsu.edu.cn*

[2]*Department of Mathematics, CINVESTAV-IPN, Apartado Postal 14-740, Mexico DF 07000, Mexico. E-mail: ohernand@math.cinvestav.mx*

This paper is concerned with two-person zero-sum games for continuous-time Markov chains, with possibly unbounded payoff and transition rate functions, under the discounted payoff criterion. We give conditions under which the existence of the value of the game and a pair of optimal stationary strategies is ensured by using the optimality (or Shapley) equation. We prove the convergence of the value iteration scheme to the game's value and to a pair of optimal stationary strategies. Moreover, when the transition rates are bounded we further show that the convergence of value iteration is exponential. Our results are illustrated with a controlled queueing system with unbounded transition and reward rates.

Keywords: controlled Q-process; discounted payoffs; value of the game; zero-sum Markov games

1. Introduction

Zero-sum stochastic dynamic games have been widely studied in the literature. Existing work can be roughly classified into four main groups. The first deals with *discrete-time games* (see, for instance, Basar and Olsder 1999; Filar and Vrieze 1997; Hernández-Lerma and Lasserre 2001; Sennott 1994; Shapley 1953); the second with *differential games* (e.g. Ardanuy and Alcalá 1992; Hamadène 1999; Ramachandran 1999); and the third with semi-Markov games in which the players can choose their actions only at *certain* (random) epochs, and which, therefore, can be reduced to discrete-time games (see Lal and Sinha 1992, for instance). In this paper, we study a fourth class of stochastic games, namely, games in which the state process evolves according to a continuous-time Markov chain, and the players can select their actions *continuously* in time. This fourth class has been studied by Lai and Tanaka (1984), Tanaka and Homma (1978) and Tanaka and Wakuta (1978). However, the latter references are all restricted to the case where the transition and payoff rates are *both bounded*, and, moreover, each player uses *only* stationary strategies. Here, we consider a much more general case.

More precisely, we consider zero-sum games for continuous-time Markov chains with a discounted payoff criterion. The transition and payoff rates are both allowed to be

1350–7265 © 2005 ISI/BS

unbounded, and each player may use randomized, time-varying, Markov strategies. We give conditions under which the optimality (or Shapley or dynamic programming) equation has a unique solution, which is used to show that the game has a value, as well as the existence of a pair of optimal stationary strategies. In addition, we prove the convergence of the value iteration scheme to the game's value and that it yields a pair of optimal stationary strategies. Moreover, when the transition rates are bounded we further show that the convergence of value iteration is exponential. Our results are illustrated with a controlled queueing system with unbounded transition and reward rates.

The rest of this paper is organized as follows. Sections 2 and 3 introduce the game model and the family of admissible strategies, respectively. The optimality criterion we are concerned with is presented in Section 4. Our main optimality results are stated in Section 5 and illustrated with examples in Section 6. Their proofs are postponed to Section 8 after some technical preliminaries in Section 7. We conclude in Section 9 with some general remarks.

2. The game model

In this section we introduce the (continuous-time, time-homogeneous) two-person zero-sum stochastic game model:

$$\{S, A, B, K_A, K_B, q, r\}, \tag{2.1}$$

where S is the *state space*, a denumerable set, and A and B are the *action spaces* for players 1 and 2, respectively, which are assumed to be Polish (i.e., complete and separable metric) spaces. The sets $K_A \subset S \times A$ and $K_B \subset S \times B$ are Borel sets that represent the constraint sets. That is, for each state $i \in S$, the i-section in K_A, namely

$$A(i) := \{a \in A | (i, a) \in K_A\},$$

represents the set of admissible actions for player 1 in state i; similarly, the i-section in K_B,

$$B(i) := \{b \in B | (i, b) \in K_B\},$$

stands for the family of admissible actions for player 2 in state i. Let

$$K := \{(i, a, b) | i \in S, a \in A(i), b \in B(i)\}, \tag{2.2}$$

which is a Borel subset of $S \times A \times B$.

The component q in (2.1) is the matrix of the game's *transition rates* $[q(j|i, a, b)]$ satisfying $q(j|i, a, b) \geqslant 0$ for all $(i, a, b) \in K$ and $i \neq j$, and which is assumed to be *conservative*, that is,

$$\sum_{j \in S} q(j|i, a, b) = 0, \qquad \forall\ (i, a, b) \in K, \tag{2.3}$$

and *stable*, that is,

$$q(i) := \sup_{a \in A(i), b \in B(i)} q_i(a, b) < \infty, \qquad \forall\ i \in S, \tag{2.4}$$

where $q_i(a, b) := -q(i|i, a, b)$ for all $a \in A(i)$ and $b \in B(i)$. Moreover, $q(j|i, a, b)$ is a measurable function on $A \times B$ for each fixed $i, j \in S$.

Finally, $r : K \to \mathbb{R} := (-\infty, +\infty)$ is the reward rate function for player 1 (or the cost rate function for player 2).

The game is played as follows. Players 1 and 2 observe *continuously* the current state of the system. Whenever the system is at state $i \in S$ at time $t \geqslant 0$, they independently choose actions $a_t \in A(i)$ and $b_t \in B(i)$ according to some admissible 'strategies' introduced in Definition 3.1 below. As a consequence of this, the following happens: (1) player 1 receives a reward rate $r(i, a_t, b_t)$; (2) player 2 incurs a cost rate $r(i, a_t, b_t)$; and (3) the system moves to a new state $j \neq i$ with a possibly non-homogeneous transition probability function determined by the transition rates $[q(j|i, a_t, b_t)]$. The goal of player 1 is to maximize his/her reward, whereas that of player 2 is to minimize his/her cost with respect to some performance criterion V_α, which in our present case is defined by (4.1) below.

3. Strategies

We begin this section with some notation. If X is a Polish space, we denote by $\mathcal{B}(X)$ its Borel σ-algebra, and by $P(X)$ the Borel space of probability measures on X, endowed with the topology of weak convergence.

A *randomized Markov strategy* for player 1, denoted by π^1, is a family $(\pi_t^1, t \geqslant 0)$ of stochastic kernels satisfying the following conditions:

(1) for each $t \geqslant 0$ and $i \in S$, $\pi_t^1(\cdot|i)$ is a probability measure on A such that $\pi_t^1(A(i)|i) = 1$;
(2) for every $E \in \mathcal{B}(A)$ and $i \in S$, $\pi_t^1(E|i)$ is a Borel measurable function in $t \geqslant 0$.

Without loss of generality, by (1) we may also regard $\pi_t^1(\cdot|i)$ as a probability measure on $A(i)$. We denote by $\prod_1^m$ the family of all randomized Markov strategies for player 1. Moreover, a strategy $\pi^1 = (\pi_t^1, t \geqslant 0) \in \prod_1^m$ is called *stationary* if, for each $i \in S$, there is a probability measure $\pi^1(\cdot|i) \in P(A(i))$ such that

$$\pi_t^1(\cdot|i) \equiv \pi^1(\cdot|i), \qquad \forall t \geqslant 0.$$

We denote this policy by $(\pi^1(\cdot|i), i \in S)$. The set of all stationary strategies for player 1 is denoted by $\prod_1^s$. The sets of all randomized Markov strategies $\prod_2^m$ and all stationary strategies $\prod_2^s$ for player 2 are defined similarly, with $P(B(i))$ in lieu of $P(A(i))$.

For each pair of strategies $(\pi^1, \pi^2) := ((\pi_t^1, \pi_t^2), t \geqslant 0) \in \prod_1^m \times \prod_2^m$, the associated transition and reward rates are defined, respectively, as follows: for each $i, j \in S$ and $t \geqslant 0$,

$$q(j|i, t, \pi^1, \pi^2) := \int_{B(i)} \int_{A(i)} q(j|i, a, b)\pi_t^1(\mathrm{d}a|i)\pi_t^2(\mathrm{d}b|i), \tag{3.1}$$

$$r(t, i, \pi^1, \pi^2) := \int_{B(i)} \int_{A(i)} r(i, a, b)\pi_t^1(\mathrm{d}a|i)\pi_t^2(\mathrm{d}b|i). \tag{3.2}$$

In particular, when π^1 and π^2 are both stationary, we write (3.1) and (3.2) as $q(j|i, \pi^1, \pi^2)$

and $r(i, \pi^1, \pi^2)$, respectively. In addition, the associated Q-matrix is $Q(t, \pi^1, \pi^2) := [q(j|i, t, \pi^1, \pi^2)]$. As is well known (see Anderson 1991; Chung 1960; Feller 1940; Hou and Guo 1998; or Hou 1994) any (probably substochastic) transition function $\bar{p}(s, i, t, j, \pi^1, \pi^2)$ for which $Q(t, \pi^1, \pi^2)$ is its transition rate matrix (i.e.

$$\frac{\partial \bar{p}(s, i, t, j, \pi^1, \pi^2)}{\partial t}|_{t=s} = q(j|i, s, \pi^1, \pi^2)$$

for all $i, j \in S$ and $s \geqslant 0$), is called a Q-*process*. A Q-process $\bar{p}(s, i, t, j, \pi^1, \pi^2)$ is said to be *honest* if $\sum_{j \in S} \bar{p}(s, i, t, j, \pi^1, \pi^2) = 1$ for all $i \in S$ and $t \geqslant s \geqslant 0$; see Anderson (1991), Chung (1960) or Feller (1940) for more details.

In the spirit of the conditions in Feller (1940) for the existence of such Q-processes, we will restrict ourselves to control strategies in the classes $\prod_1$ and $\prod_2$ defined as follows:

Definition 3.1. *For $k = 1, 2$, $\prod_k$ is any subset of randomized Markov strategies for player k such that $\prod_k$ contains $\prod_k^s$ and satisfies a continuity condition on the corresponding transition rates in $t \geqslant 0$ for each strategy in $\prod_\ell$ with $\ell \neq k$. Hence, $q(j|t, i, \pi^1, \pi^2)$ is continuous in $t \geqslant 0$ for each $i, j \in S$ and $(\pi^1, \pi^2) \in \prod_1 \times \prod_2$.*

Remark 3.1. Observe that $\prod_1 \times \prod_2$ is non-empty because it contains $\prod_1^s \times \prod_2^s$. Moreover, we provide an example in Section 6 showing that $\prod_1$ and $\prod_2$ can be chosen to be strictly larger than $\prod_1^s$ and $\prod_2^s$, respectively.

For each fixed pair $(\pi^1, \pi^2) \in \prod_1 \times \prod_2$, since the matrix $[q(j|i, a, b)]$ is conservative and stable (see (2.3), (2.4)), $Q(t, \pi^1, \pi^2)$ is also conservative and stable. Hence, for each $\pi^1 \in \prod_1$ and $\pi^2 \in \prod_2$ the existence of a Q-process, such as the *minimum* Q-process denoted by $p^{\min}(s, i, t, j, \pi)$ (i.e., $p^{\min}(s, i, t, j, \pi) \leqslant \bar{p}(s, i, t, j, \pi)$ for any Q-process $\bar{p}(s, i, t, j, \pi)$), is guaranteed; but it is not necessarily *regular* (i.e. unique and honest), that is, we might have $\sum_{j \in S} p^{\min}(s, i, t, j, \pi) < 1$ for some $i \in S$ and $t \geqslant s \geqslant 0$; see Feller (1940) and Chung (1960). Thus, for a Q-process to be regular and also for our payoff criterion (4.1) to be well defined, throughout this paper we make the following assumption:

Assumption A. *For each pair of strategies $(\pi^1, \pi^2) \in \prod_1 \times \prod_2$, there is a regular Q-process with transition rate matrices $\{Q(t, \pi^1, \pi^2),\ t \geqslant 0\}$.*

To ensure that Assumption A holds we may use, for instance, the following fact.

Proposition 3.1. *Each of the following conditions implies that Assumption A holds:*

(a) *The transition rates are bounded, that is, $\|q\| := \sup_{i \in S} q(i) < \infty$ with $q(i)$ as in* (2.4).

(b) *There exist N non-negative functions w_n on S and a positive constant c such that, for all $(i, a, b) \in K$,*

$$\sum_{j\in S} q(j|i, a, b)w_n(j) \leqslant w_{n+1}(i), \qquad \textit{for } n = 1, \ldots, N-1,$$

and, furthermore,

$$\sum_{j\in S} q(j|i, a, b)w_N(j) \leqslant 0$$

and

$$q(i) \leqslant c(w_1(i) + \ldots + w_N(i)), \qquad \text{for all } \ i \in S. \tag{3.3}$$

Proof. By (2.3) we see that condition (a) implies (b) with $w_1(i) := \|q\|$ for all $i \in S$ and $N = 1$. Thus, it suffices to verify that (b) implies Assumption A, which is done after Lemma 7.3 below. □

Remark 3.2. (a) The conditions in Feinberg (2004), Lai and Tanaka (1984), Puterman (1994), Sennott (1999), Tanaka and Homma (1978), Tanaka and Wakuta (1978) and Yushkevich and Fainberg (1979) imply Assumption A because all of these references require the transition rates to be bounded. On the other hand, in Section 6 we introduce a queueing system with *unbounded* transition rates for which Assumption A is true.

(b) For each $(\pi^1, \pi^2) \in \prod_1 \times \prod_2$, we denote by $P(s, t, \pi^1, \pi^2) := [p(s, i, t, j, \pi^1, \pi^2)]$ the regular Q-process, that is the transition probability function, and by $\{x(t, \pi^1, \pi^2)\}$ the associated right-continuous Markov chain. Further, for each initial state $i \in S$ at time $s \geqslant 0$, $P_{s,i}^{\pi^1,\pi^2}$ and $\mathrm{E}_{s,i}^{\pi^1,\pi^2}$ denote the corresponding probability measure and expectation operator determined by $P(s, t, \pi^1, \pi^2)$, respectively.

4. Discounted payoff criterion

For each pair of strategies $(\pi^1, \pi^2) \in \prod_1 \times \prod_2$, initial data $(s, i) \in \bar{S} := [0, \infty) \times S$ and a given discount factor $\alpha > 0$, the discounted payoff criterion $V_\alpha(s, i, \pi^1, \pi^2)$ is defined as

$$\begin{aligned} V_\alpha(s, i, \pi^1, \pi^2) &:= \int_s^\infty \mathrm{e}^{-\alpha(t-s)} \mathrm{E}_{s,i}^{\pi^1,\pi^2} r(t, x(t, \pi^1, \pi^2), \pi^1, \pi^2)\mathrm{d}t \\ &= \int_s^\infty \mathrm{e}^{-\alpha(t-s)} \left[\sum_{j\in S} p(s, i, t, j, \pi^1, \pi^2) r(t, j, \pi^1, \pi^2)\right] \mathrm{d}t. \end{aligned} \tag{4.1}$$

To introduce our optimality criterion we also need the following concepts. The functions on $\bar{S}$ defined as

$$L(s, i) := \sup_{\pi^1\in\Pi_1} \inf_{\pi^2\in\Pi_2} V_\alpha(s, i, \pi^1, \pi^2) \text{ and } U(s, i) := \inf_{\pi^2\in\Pi_2} \sup_{\pi^1\in\Pi_1} V_\alpha(s, i, \pi^1, \pi^2) \tag{4.2}$$

are called the *lower value* and the *upper value*, respectively, of the discounted payoff game. It is clear that

$$L(s, i) \leqslant U(s, i), \qquad \forall\, (s, i) \in \bar{S}. \tag{4.3}$$

Definition 4.1. *If $L(s, i) = U(s, i)$ for all $(s, i) \in \bar{S}$, then the common function is called the value of the game and is denoted by V.*

Definition 4.2. *Suppose that the game has a value V. Then a strategy π^{*1} in $\prod_1$ is said to be optimal for player 1 if*

$$\inf_{\pi^2 \in \Pi_2} V_\alpha(s, i, \pi^{*1}, \pi^2) = V(s, i), \qquad \forall\, (s, i) \in \bar{S}. \tag{4.4}$$

Similarly, $\pi^{*2} \in \prod_2$ is optimal for player 2 if

$$\sup_{\pi^1 \in \Pi_1} V_\alpha(s, i, \pi^1, \pi^{*2}) = V(s, i), \qquad \forall\, (s, i) \in \bar{S}. \tag{4.5}$$

If $\pi^{*k} \in \prod_k$ is optimal for player k $(k = 1, 2)$, then (π^{*1}, π^{*2}) is called a pair of optimal strategies (also known as a saddlepoint).

To ensure that the discounted payoff criterion $V_\alpha(s, i, \pi^1, \pi^2)$ is a finite-valued function we shall suppose the following.

Assumption B. *There exist N non-negative functions w_n, $n = 1, 2, \ldots, N$, such that, for all $(i, a, b) \in K$,*

$$\sum_{j \in S} q(j|i, a, b) w_n(j) \leqslant w_{n+1}(i), \qquad \textit{for } n = 1, \ldots, N-1, \tag{4.6}$$

and

$$\sum_{j \in S} q(j|i, a, b) w_N(j) \leqslant 0. \tag{4.7}$$

Observe that Assumption B is similar to – but not the same as – the condition in Proposition 3.1(b). In particular, Assumption B does not necessarily imply (3.3).

Let

$$W(i) := w_1(i) + \ldots + w_N(i), \qquad \text{for all } i \in S. \tag{4.8}$$

Since $[q(j|i, a, b)]$ is conservative, Assumption B still holds if we replace w_N with $w_N + 1$. Thus, from now on we suppose that $W \geqslant 1$.

Remark 4.1. (a) By (2.3), if the transition rates are bounded – that is, $\|q\| < \infty$ as in, for instance, Feinberg (2004), Lai and Tanaka (1984), Puterman (1994), Sennott (1999), Tanaka and Homma (1978), Tanaka and Wakuta (1978) and Yushkevich and Fainberg (1979) – then Assumption B trivially holds with $N = 1$ and $w_1 := \|q\|$. Moreover, if Assumption B holds and, in addition, $q(i) \leqslant cW(i)$ for all $i \in S$ and some constant $c > 0$, then, by Proposition 3.1, our Assumption A is also true. On the other hand, suppose that Assumption B holds and, furthermore, there exists a sequence $\{S_m, m \geqslant 1\}$ of subsets of S such that $S_m \uparrow S$, $\sup_{i \in S_m} q(i) < \infty$, and $\lim_{m \to \infty}[\inf_{j \notin S_m} W(j)] = +\infty$. Then from Guo and Hernández-Lerma (2003b) we see that for each pair of strategies $(\pi^1, \pi^2) \in \prod_1 \times \prod_2$ the associated Q-process with transition rate matrices $Q(t, \pi^1, \pi^2)$ is regular, and so Assumption A is not required.

(b) Our Assumption B was motivated by conditions in Lippman (1973, 1975) and Van Nunen and Wessels (1978) for discounted semi-Markov decision processes (e.g., the case where $B(i)$ is a singleton $\{b_i\}$ for each $i \in S$), which in a sense are stronger than Assumption B. To see this, let us suppose that $B(i) = \{b_i\}$ for each $i \in S$, and consider Lippman's (1973, 1975) conditions, which are also used in Van Nunen and Wessels (1978). As in Lippman (1973, 1975), let $p(\cdot|i, a) := p(\cdot|i, a, b_i)$ and $r(i, a) := r(i, a, b_i)$ be the transition probability and the one-period reward, respectively, for each $i \in S$ and $a \in A(i) \equiv A$. Further, $\alpha > 0$ is the discount factor, and $t(\cdot|i, a)$ is the probability distribution of the time until the next transition, given the current state $i \in S$ and the action $a \in A$. With this notation, we have what we will refer to as

Lippman's conditions: there exists a function $w(\cdot) \geqslant 1$ on S, an integer $N \geqslant 1$, a number β $(0 \leqslant \beta < 1)$ and positive numbers c and M such that, for all $i \in S$, $a \in A$,

(L1) $\beta(i, a) \equiv \int_0^\infty \mathrm{e}^{-\alpha\tau} t(\mathrm{d}\tau|i, a) \leqslant \beta$,
(L2) $|r(i, a, b_i)| w(i)^{-N} \leqslant M$,
(L3) $\sum_{j \in S} w^n(j) p(j|i, a, b_i) \leqslant [w(i) + c]^n$, for $n = 1, \ldots, N$.

Obviously, these conditions are different from our Assumption B. However, in the case $B(i) = \{b_i\}$ we get the following.

Proposition 4.1. *Under Lippman's conditions, for our game model we have:*

(a) *the transition rates are bounded, that is* $\|q\| = \sup_{i \in S} q(i) < \infty$;
(b) *Assumption B holds.*

Proof. (a) Since

$$\beta(i, a) \equiv \int_0^\infty \mathrm{e}^{-\alpha\tau} t(\mathrm{d}\tau|i, a) = \int_0^\infty \mathrm{e}^{-\alpha\tau}\, \mathrm{d}(1 - \mathrm{e}^{q(i|i,a,b_i)\tau}) = \frac{-q(i|i, a, b_i)}{\alpha - q(i|i, a, b_i)},$$

by (L1) we have $-q(i|i, a, b_i) \leqslant \alpha\beta/(1 - \beta)$ for all $i \in S$ and $a \in A$. This gives part (a).

(b) Since $p(j|i, a, b_i) = q(j|i, a, b_i)/(-q(i|i, a, b_i))$ when $i \neq j$ and $p(i|i, a, b_i) = 0$, by part (a) and (L3) with $n = 1$ we have

$$\sum_{j \in S} q(j|i, a, b_i) w(j) = (-q(i|i, a, b_i)) \sum_{j \in S} p(j|i, a) w(j) + q(i|i, a, b_i) w(i)$$

$$\leqslant -q(i|i, a, b_i)[w(i) + c] + q(i|i, a, b_i) w(i) \leqslant \frac{\alpha\beta c}{1 - \beta}.$$

Let $w_1 = w$, $w_2 = \alpha\beta c/(1 - \beta)$. Then Assumption B is true (with $N = 2$). □

Finally, to ensure the existence of a pair of optimal stationary strategies, in addition to

Assumptions A and B we impose the following continuity–compactness conditions, in which $W(\cdot)$ is the function in (4.8).

Assumption C. *For each $i \in S$:*

(1) *$A(i)$ and $B(i)$ are compact sets;*
(2) *$r(i, a, b)$ and $q(j|i, a, b)$ are continuous in $(a, b) \in A(i) \times B(i)$;*
(3) *the function $\sum_{j \in S} q(j|i, a, b)W(j)$ is continuous in $(a, b) \in A(i) \times B(i)$;*
(4) *there is a constant M such that*

$$|r(i, a, b)| \leqslant MW(i), \qquad \forall\, (i, a, b) \in K;$$

(5) *there exists a non-negative function $\tilde{W}$ on S and positive constants $\tilde{M}$, c and $\tilde{c}$ such that*

$$q(i)W(i) \leqslant \tilde{M}\tilde{W}(i), \qquad \sum_{j \in S} q(j|i, a, b)\tilde{W}(j) \leqslant c\tilde{W}(i) + \tilde{c}, \quad \forall (i, a, b) \in K.$$

Remark 4.2. (a) Assumptions B and C(1)–C(4) are a variant of hypotheses used in Guo and Hernández-Lerma (2003a), Guo and Liu (2001), and Guo and Zhu (2002) for continuous-time Markov control processes, and of hypotheses used in Hernández-Lerma and Lasserre (1999) for discrete-time Markov control processes. Assumption C(5) is for the existence of a pair of optimal stationary strategies.

(b) If the transition and reward rates are bounded (see Lai and Tanaka 1984; Tanaka and Homma 1978; Tanaka and Wakuta 1978), then Assumptions A, B, C(4) and C(5) hold. Moreover, an example in which the transition and reward rates are both unbounded and (all parts of) Assumptions A, B and C hold will be given in Section 6. On the other hand, if $r(i, a, b)$ is uniformly bounded on K then Assumption C(4) trivially holds, whereas Assumption C(5) and the continuity condition for $u = W$ in Assumption C(3) are not required.

To state our results, we use the weighted supremum norm $\|\cdot\|_W$ for real-valued functions u on S, defined as

$$\|u\|_W := \sup_{i \in S}[W(i)^{-1}|u(i)|], \tag{4.9}$$

and the Banach space

$$B(S) := \{u |\, \|u\|_W < \infty\}.$$

We will also use the following facts, which are essentially known already, but we state them here for completeness and ease of reference.

Proposition 4.2. *Suppose that Assumptions A, B and C hold, and let π^k be an arbitrary strategy in $\prod_k$ $(k = 1, 2)$.*

(a) *If there exists $u \in B(S)$ such that*

$$\alpha u(i) = r(t, i, \pi^1, \pi^2) + \sum_{j \in S} q(j|i, t, \pi^1, \pi^2)u(j), \qquad \forall\, (t, i) \in \bar{S},$$

then $u(i) = V_\alpha(s, i, \pi^1, \pi^2)$ *for all* $(s, i) \in \bar{S}$. *(Recall that* $\bar{S} := [0, \infty) \times S$.*)*

(b) *If there exists* $u \in B(S)$ *such that*

$$\alpha u(i) \leqslant r(t, i, \pi^1, \pi^2) + \sum_{j \in S} q(j|i, t, \pi^1, \pi^2)u(j), \qquad \forall\, (t, i) \in \bar{S},$$

then $u(i) \leqslant V_\alpha(s, i, \pi^1, \pi^2)$ *for all* $(s, i) \in \bar{S}$.

(c) *Similarly, if there exists* $u \in B(S)$ *such that*

$$\alpha u(i) \geqslant r(t, i, \pi^1, \pi^2) + \sum_{j \in S} q(j|i, t, \pi^1, \pi^2)u(j), \qquad \forall\, (t, i) \in \bar{S},$$

then $u(i) \geqslant V_\alpha(s, i, \pi^1, \pi^2)$ *for all* $(s, i) \in \bar{S}$.

(d) *For each* $(\pi^1, \pi^2) \in \prod_1^s \times \prod_2^s$, $V_\alpha(0, i, \pi^1, \pi^2)$ *is the unique solution in* $B(S)$ *of the equation*

$$\alpha u(i) = r(i, \pi^1, \pi^2) + \sum_{j \in S} q(j|i, \pi^1, \pi^2)u(j), \qquad \forall\, i \in S,$$

and, furthermore, $V_\alpha(0, i, \pi^1, \pi^2) = V_\alpha(s, i, \pi^1, \pi^2)$ *for all* $(s, i) \in \bar{S}$.

Proof. These results follow from Lemma 6.2 in Guo and Hernández-Lerma (2003a); see also Proposition 3.3 in Hernández-Lerma (1994) or Guo and Zhu (2002) and Lemma 7.3(a) below. □

5. Main results

We now state our main results. To do so, for any two states $i, j \in S$, any two probability measures $\phi \in P(A(i))$, $\psi \in P(B(i))$, and any pair $(\pi^1, \pi^2) \in \prod_1^s \times \prod_2^s$, let

$$q(j|i, \phi, \psi) := \int_{B(i)} \int_{A(i)} q(j|i, a, b)\phi(\mathrm{d}a)\psi(\mathrm{d}b). \tag{5.1}$$

$$r(i, \phi, \psi) := \int_{B(i)} \int_{A(i)} r(i, a, b)\phi(\mathrm{d}a)\psi(\mathrm{d}b), \tag{5.2}$$

$$q(j|i, \phi, \pi^2) := q(j|i, \phi, \pi_2(\cdot|i)),$$

$$q(j|i, \pi^1, \psi) := q(j|i, \pi_1(\cdot|i), \psi).$$

Under Assumption B, let

$$R_k := w_k + \ldots + w_N, \qquad \text{for } k = 1, 2, \ldots, N. \tag{5.3}$$

Theorem 5.1. *Suppose that Assumptions A, B, and C hold.*

(a) *Let* $u_0(i) := -M\sum_{k=1}^{N}\alpha^{-k}R_k(i)$ *and* $u_n(i) := Tu_{n-1}(i)$ *for each* $i \in S$ *and* $n \geqslant 1$, *where*

$$Tu(i) := \sup_{\phi\in P(A(i))}\inf_{\psi\in P(B(i))}\left\{\frac{r(i,\phi,\psi)}{1+\alpha+q(i)} + \frac{1+q(i)}{1+\alpha+q(i)}\sum_{j\in S}\left[\frac{q(j|i,\phi,\psi)}{1+q(i)}+\delta_{ij}\right]u(j)\right\},$$

and δ_{ij} *is the Kronecker delta. Then the limit* $\lim_{n\to\infty}u_n := u^*$ *exists and belongs to* $B(S)$.

(b) u^* *is a solution of the optimality equation*

$$\alpha u(i) = \sup_{\phi\in P(A(i))}\inf_{\psi\in P(B(i))}\{r(i,\phi,\psi)+\sum_{j\in S}q(j|i,\phi,\psi)u(j)\}, \qquad \forall\, i\in S. \tag{5.4}$$

(c) *There exists a pair of stationary strategies* $(\pi^{*1},\pi^{*2}) \in \prod_1^s \times \prod_2^s$ *such that, for all* $i\in S$,

$$\alpha u^*(i) = r(i,\pi^{*1},\pi^{*2}) + \sum_{j\in S}q(j|i,\pi^{*1},\pi^{*2})u^*(j) \tag{5.5}$$

$$= \sup_{\phi\in P(A(i))}\left\{r(i,\phi,\pi^{*2})+\sum_{j\in S}q(j|i,\phi,\pi^{*2})u^*(j)\right\} \tag{5.6}$$

$$= \inf_{\psi\in P(B(i))}\left\{r(i,\pi^{*1},\psi)+\sum_{j\in S}q(j|i,\pi^{*1},\psi)u^*(j)\right\} \tag{5.7}$$

$$= \sup_{\phi\in P(A(i))}\inf_{\psi\in P(B(i))}\left\{r(i,\phi,\psi)+\sum_{j\in S}q(j|i,\phi,\psi)u^*(j)\right\}. \tag{5.8}$$

(d) $u^*(i) = L(s,i) = U(s,i)$ *for all* $(s,i)\in\bar{S}$, *which means that the value V of the game exists and that the solution* u^* *of* (5.4) *is unique and equals V.*

(e) (π^{*1},π^{*2}) *in part* (c) *is a pair of optimal stationary strategies.*

(f) *For each* $n\geqslant 1$ *and* $i\in S$, *there exists* $(\phi_n^*(i),\psi_n^*(i)) \in P(A(i))\times P(B(i))$ *such that*

$$r(i,\phi_n^*(i),\psi_n^*(i)) + \sum_{j\in S}q(j|i,\phi_n^*(i),\psi_n^*(i))u_n(j)$$

$$= \sup_{\phi\in P(A(i))}\left\{r(i,\phi,\psi_n^*(i))+\sum_{j\in S}q(j|i,\phi,\psi_n^*(i))u_n(j)\right\} \tag{5.9}$$

$$= \inf_{\psi\in P(B(i))}\left\{r(i,\phi_n^*(i),\psi)+\sum_{j\in S}q(j|i,\phi_n^*(i),\psi)u_n(j)\right\}. \tag{5.10}$$

Moreover, any limit point (ϕ^*, ψ^*) of the sequence $\{\phi_n^*, \psi_n^*\}$ in $\prod_1^s \times \prod_2^s$ is a pair of optimal stationary strategies.

Theorem 5.1 (proved in Section 8) gives a very complete solution of the discounted game. Indeed, it gives (i) the *existence* of the value of the game and (ii) of a pair of optimal stationary strategies, as well as (iii) the convergence of the so-called *value iteration* functions u_n to the game's value and (iv) the 'convergence' (in the sense of part (f)) of the value iteration strategies $\{\phi_n^*, \psi_n^*\}$ to a pair of optimal stationary strategies. (For the one-player discrete-time case, value iteration is studied in Filar and Vrieze (1997), Hernández-Lerma and Lasserre (1999), Puterman (1994) and Tijms (1994).) Moreover, when the transition rates are *bounded*, we can further show (and in Section 8 prove) that the convergence is exponential, as follows.

Theorem 5.2. *Suppose that*

(i) *Assumptions A, B and C(1)–C(4) hold, and*
(ii) *the discount factor is $\alpha > 1$, and the transition rates are bounded, i.e. $\bar{q} := \|q\| < \infty$.*

Then

(a) *the operator T is a contraction on $B(S)$ with modulus $\gamma := (2 + \bar{q})/(1 + \alpha + \bar{q}) < 1$;*
(b) $\|T^n u - u^*\|_W \leqslant \gamma^n(\|u\|_W + M\sum_{k=1}^N \alpha^{-k})$ *for all $n \geqslant 1$ and $u \in B(S)$.*

6. Examples

There are many applications of game theory to queueing systems; see, for instance, Altman (2005) and the references therein. In this section we present two queueing games that illustrate our results.

Example 6.1. Consider a single-server queueing system in which the state variable denotes the total number of jobs (in service and waiting in the queue) at each time $t \geqslant 0$. There are 'natural' arrival and service rates, say λ and μ, respectively, in addition to service parameters $u(a)$ controlled by player 1, and arrival parameters $v(b)$ controlled by player 2. Thus, when the state of the system is $i \in S := \{0, 1, \ldots\}$, player 1 takes an action a from a given set $A(i) \subset A$, which may increase ($u(a) \geqslant 0$) or decrease ($u(a) \leqslant 0$) the service rate. These actions produce a cost (or reward) denoted by $c_1(a) \geqslant 0$ (or $c_1(a) \leqslant 0$) per unit time. Similarly, if the state is $i \in S$, player 2 takes an action b from a set $B(i) \subset B$ to reject ($v(b) \leqslant 0$) or to attract ($v(b) \geqslant 0$) customers, and these actions result at a cost (or reward) rate $c_2(b) \geqslant 0$ (or $c_2(b) \leqslant 0$). In addition, assuming that player 1 'owns' the system, he/she gets a reward $r(i) := pi$ for each unit of time during which the system remains in the state i, where $p > 0$ is a fixed fee per customer. We formulate this model as a continuous-time Markov game.

The corresponding transition rate $q(j|i, a, b)$ and reward rate $r(i, a, b)$ for player 1 are given as follows: For $i = 0$,

$$q(1|0, a, b) = -q(0|0, a, b) := u(a) + v(b),$$

and for $i \geqslant 1$,

$$q(j|i, a, b) = \begin{cases} \mu i + u(a), & \text{if } j = i - 1, \\ -(\mu + \lambda)i - u(a) - v(b), & \text{if } j = i, \\ \lambda i + v(b), & \text{if } j = i + 1, \\ 0, & \text{otherwise,} \end{cases} \tag{6.1}$$

$$r(i, a, b) = pi - c_1(a) + c_2(b), \qquad \text{for } (i, a, b) \in K, \tag{6.2}$$

with K as in (2.2).

The aim here is to find conditions under which there exists a pair of optimal stationary strategies achieving both the maximum discounted reward for player 1 and the minimum discounted cost for player 2. To do so, we use the following assumptions:

(E1) $u(a) + v(b) \geqslant 0$ for all $a \in A(0)$ and $b \in A(0)$; for each $i \geqslant 1$, $\mu i + u(a) \geqslant 0$ for all $a \in A(i)$ and $\lambda i + v(b) \geqslant 0$ for all $b \in B(i)$. Moreover, $0 \leqslant \lambda \leqslant \mu$.

(E2) The action sets A and B are metric spaces, and $A(i)$ and $B(i)$ are compact for each $i \in S$.

(E3) $c_1(a)$, $c_2(b)$, $u(a)$ and $v(b)$ are bounded in the supremum norm and continuous functions on their corresponding domains.

Under these conditions, we obtain the following.

Proposition 6.1. *Under* (E1)–(E3)*, the above queueing system satisfies the Assumptions A, B and C. Therefore (by Theorem 5.1), there exists an optimal pair of stationary strategies.*

Proof. We shall first verify Assumption B. Let $w_1(i) := p^{-1}i$ for all $i \in S$, and $\|u\| := \sup_{a\in A}|u(a)|$, $\|v\| := \sup_{b\in B}|v(b)|$, $\|c_1\| := \sup_{a\in A}|c_1(a)|$, $\|c_2\| := \sup_{b\in B}|c_2(b)|$. Then, for each $(i, a, b) \in K$, under (E1) we have:

$$\text{when } i \geqslant 1, \qquad \sum_{j\in S} q(j|i, a, b)w_1(j) = p^{-1}[(\lambda - \mu)i - u(a) + v(b)] \leqslant p^{-1}(\|u\| + \|v\|);$$

$$\text{when } i = 0, \qquad \sum_{j\in S} q(j|i, a, b)w_1(j) = p^{-1}[u(a) + v(b)] \leqslant p^{-1}(\|u\| + \|v\|).$$

Let $w_2(i) \equiv p^{-1}(\|c_1\| + \|c_2\| + \|u\| + \|v\|)$ for all $i \in S$. Then, for each $(i, a, b) \in K$, we have

$$\sum_{j\in S} q(j|i, a, b)w_1(j) \leqslant w_2(i), \tag{6.3}$$

$$\sum_{j\in S} q(j|i, a, b)w_2(j) \leqslant 0. \tag{6.4}$$

Hence, Assumption B holds. Now let $W := w_1 + w_2$. Then, by Lemma 7.3(a) below and (6.3)–(6.4), we obtain that, for all $t \geqslant s \geqslant 0$, $\pi^1 \in \prod_1$, $\pi^2 \in \prod_2$ and $i \in S$,

$$\int_s^t \sum_{j\in S} p(s, i, y, j, \pi^1, \pi^2)W(j)\mathrm{d}y < \infty.$$

Therefore, for all $t \geqslant s \geqslant 0$, $\pi^1 \in \prod_1$, $\pi^2 \in \prod_2$ and $i \in S$,

$$\int_s^t \sum_{j\in S} p(s, i, y, j, \pi^1, \pi^2)[(\lambda + \mu)j + \|u\| + \|v\|)]\mathrm{d}y < \infty,$$

which together with Proposition 2.1(c) in Guo and Hernández-Lerma (2003a) implies that Assumption A holds. Finally, since $|r(i, a, b)| \leqslant pi + \|c_1\| + \|c_2\| \leqslant pW(i)$ and $q(i) \leqslant (\lambda + \mu + \|u\| + \|v\|)(i + 1)$, letting $\tilde{M}(i) := \tilde{M}(i + 1)^2$ with $\tilde{M} := 9p^{-1}(\|c_1\| + \|c_2\| + \|u\| + \|v\| + \mu + \lambda + 1)^2$ and using (E2) and (E3) together with (6.1), we see that Assumption C holds. □

Example 6.2. In Example 6.1, we further suppose that $A(i) = \{a_1, a_2\}$, $B(i) = \{b_1, b_2\}$ for each $i \in S$, except $A(0) = \{a_1\}$, with $0 < a_1 < a_2 < b_1 < b_2$. Further, $u(a) = a$, $v(b) = b$. Then, by Proposition 6.1, Assumptions A, B and C are satisfied for these data. We now define non-stationary Markov policies $\tilde{\pi}^1 = (\tilde{\pi}^1_t, t \geqslant 0)$ and $\tilde{\pi}^2 = (\tilde{\pi}^2_t, t \geqslant 0)$ as

$$\tilde{\pi}^1_t(a|i) = \begin{cases} \frac{1}{2}\mathrm{e}^{-a_1 it}, & \text{if } a = a_1, \\ 1 - \frac{1}{2}\mathrm{e}^{-a_1 it}, & \text{if } a = a_2, \\ 1, & \text{if } i = 0,\ a = a_1, \end{cases} \tag{6.5}$$

and

$$\tilde{\pi}^2_t(b|i) = \begin{cases} 1 - \frac{1}{2}\mathrm{e}^{-b_2 it}, & \text{if } b = b_1, \\ \frac{1}{2}\mathrm{e}^{-b_2 it}, & \text{if } b = b_2, \end{cases} \tag{6.6}$$

respectively.

Moreover, let $\prod_1 := \prod_1^s \cup \{\tilde{\pi}^1\}$, $\prod_2 := \prod_2^s \cup \{\tilde{\pi}^2\}$. By (6.1), (6.5), (6.6) and (3.1) we see that $\prod_1$ and $\prod_2$ satisfy the requirements in Definition 3.1, and $\prod_1 \supset \prod_1^s$, $\prod_1 \neq \prod_1^s$; $\prod_2 \supset \prod_2^s$, $\prod_2 \neq \prod_2^s$.

Remark 6.1. It should be noted that in Examples 6.1 and 6.2 the reward and transition rates are both *unbounded*.

7. Technical preliminaries

In this section we present some results needed to prove Theorems 5.1 and 5.2. In the remainder of the paper, a real-valued function on S is regarded as a column vector, and operations on matrices and vectors are all componentwise.

Lemma 7.1. *If Assumption B holds, then for each $\pi^1 \in \prod_1$ and $\pi^2 \in \prod_2$, and $t \geqslant 0$:*

(a) $Q(t, \pi^1, \pi^2)w_n \leqslant w_{n+1}$, *for* $n = 1, \ldots, N-1$*;*
(b) $Q(t, \pi^1, \pi^2)w_N \leqslant 0$.

Proof. This follows from (3.1), (4.6) and (4.7). □

Lemma 7.2. *Under Assumptions B and C(1)–C(4), the functions $r(i, \phi, \psi)$ and $\sum_{j\in S} q(j|i, \phi, \psi)u(j)$ are continuous on $P(A(i)) \times P(B(i))$ for each fixed $u \in B(S)$ and $i \in S$.*

Proof. Under the stated assumptions, the two functions $\sum_{j\in S} q(j|i, a, b)W(j)$ and $r(i, a, b)$ are continuous and bounded on $A(i) \times B(i)$ for each $i \in S$. Hence, by the definition of weak convergence of probability measures, we obtain the continuity of $r(i, \phi, \psi)$. Similarly, replacing the probability measure $Q(\mathrm{d}y|x, a)$ in Hernández-Lerma and Lasserre (1999, p. 48) with $[q(j|i, a, b)/q(i) + \delta_{ij}]$, the 'extended Fatou Lemma' 8.3.7(a) in Hernández-Lerma and Lasserre (1999) gives the continuity of $\sum_{j\in S} q(j|i, \phi, \psi)u(j)$. □

Lemma 7.3. *If Assumptions A, B and C(4) hold, then for each pair of strategies $(\pi^1, \pi^2) \in \prod_1 \times \prod_2$, $u \in B(S)$ and $t \geqslant s \geqslant 0$,*

(a) $P(s, t, \pi^1, \pi^2)W \leqslant \sum_{k=1}^N ((k-1)!)^{-1}(t-s)^{(k-1)}W$*;*
(b) $\int_s^\infty \mathrm{e}^{-\alpha(t-s)} P(s, t, \pi^1, \pi^2)W\,\mathrm{d}t \leqslant (\sum_{k=1}^N \alpha^{-k})W$*;*
(c) $|U| \leqslant M(\sum_{k=1}^N \alpha^{-k})W$ *and* $|L| \leqslant M(\sum_{k=1}^N \alpha^{-k})W$*, with U and L as in* (4.2).

Proof. (a) It is well known (see Anderson 1991; Feller 1940; Hou 1994) that, given the Q-matrices $Q(t, \pi^1, \pi^2)$, the transition probability function $P(s, t, \pi^1, \pi^2) := [p(s, i, t, j, \pi^1, \pi^2), i, j \in S]$ can be constructed as

$$P(s, t, \pi^1, \pi^2) = \sum_{n=0}^{\infty} P_n(s, t, \pi^1, \pi^2), \tag{7.1}$$

where

$$P_0(s, t, \pi^1, \pi^2) := \mathrm{diag}(\mathrm{e}^{-\int_s^t q_i(u,\pi^1,\pi^2)\mathrm{d}u}, i \in S),$$

$$P_{n+1}(s, t, \pi^1, \pi^2) := \int_s^t P_0(s, u, \pi^1, \pi^2)(Q(u, \pi^1, \pi^2) + D(u, \pi^1, \pi^2))P_n(u, t, \pi^1, \pi^2)\mathrm{d}u, \tag{7.2}$$

for $n = 0, 1, \ldots,$ with

$$D(u, \pi^1, \pi^2) := \mathrm{diag}(q_i(u, \pi^1, \pi^2), i \in S), \qquad q_i(u, \pi^1, \pi^2) := -q(i|i, u, \pi^1, \pi^2),$$

for all $u \geqslant 0$ and $i \in S$. Hence, by (7.1), to prove (a) it suffices to show that

$$\sum_{n=0}^{m} P_n(s, t, \pi^1, \pi^2)W \leqslant \sum_{k=1}^{N} \frac{1}{(k-1)!}(t-s)^{k-1}R_k, \qquad \forall\, m \geqslant 0,\ t \geqslant s \geqslant 0, \tag{7.3}$$

with R_k as in (5.3). Obviously, $R_1 = W$ and, moreover, (7.3) is of course valid when $m = 0$. Now, by induction, suppose that (7.3) holds for some $m \geqslant 1$. Noting that $[Q(u, \pi^1, \pi^2) + D(u, \pi^1, \pi^2)] \geqslant 0$, by Fubini's theorem, (7.2) and the induction hypothesis, together with Lemma 7.1, we can obtain that

$$\begin{aligned}
&\sum_{n=1}^{m+1} P_n(s, t, \pi^1, \pi^2)W \\
&\quad = \int_s^t P_0(s, u, \pi^1, \pi^2)(Q(u, \pi^1, \pi^2) \\
&\qquad + D(u, \pi^1, \pi^2))\sum_{n=0}^{m} P_n(u, t, \pi^1, \pi^2)W \,\mathrm{d}u \\
&\quad \leqslant \sum_{k=1}^{N-1} \int_s^t P_0(s, u, \pi^1, \pi^2)\frac{(t-u)^{k-1}}{(k-1)!} R_{k+1}\,\mathrm{d}u \\
&\qquad + \sum_{k=1}^{N} \int_s^t P_0(s, u, \pi^1, \pi^2)D(u, \pi^1, \pi^2)\frac{(t-u)^{k-1}}{(k-1)!} R_k \mathrm{d}u \\
&\quad = \sum_{k=2}^{N} \frac{(t-s)^{k-1}}{(k-1)!} R_k + R_1 - P_0(s, t, \pi^1, \pi^2)R_1 \\
&\quad = \sum_{k=1}^{N} \frac{(t-s)^{k-1}}{(k-1)!} R_k - P_0(s, t, \pi^1, \pi^2)W,
\end{aligned} \tag{7.4}$$

which gives (7.3) for $m + 1$. Hence, (7.3) holds for all $m \geqslant 0$, and so part (a) follows.

(b) and (c) follow from (a). □

We next complete the proof of Proposition 3.1 using Lemma 7.3(a). Note that the Q-process $p(s, i, t, j, \pi^1, \pi^2)$ in (7.1) is *minimal* when the Q-process is not unique, that is, $p(s, i, t, j, \pi^1, \pi^2) \leqslant \bar{p}(s, i, t, j, \pi^1, \pi^2)$ for all $i, j \in S$, $t \geqslant s \geqslant 0$ and any Q-process $\bar{p}(s, i, t, j, \pi^1, \pi^2)$; see for instance Anderson (1991), Feller (1940) and Hou (1994). Moreover, by (2.4) and condition (b) in Proposition 3.1, we have $[-q(i|i, t, \pi^1, \pi^2)] \leqslant q(i) \leqslant cW(i)$ for all $i \in S$ and $t \geqslant 0$, and so Lemma 7.3(a) yields

$$\int_s^t \sum_{j\in S} p(i,\, s,\, u,\, j,\, \pi^1,\, \pi^2)[-q(\,j|j,\, u,\, \pi^1,\, \pi^2)]\mathrm{d}u < \infty.$$

Hence, by (7.1) we obtain $\lim_{n\to\infty}\int_s^t \sum_{j\in S} p_n(i,\, s,\, u,\, j,\, \pi^1,\, \pi^2)[-q(j|j,\, u,\, \pi^1,\, \pi^2)]\mathrm{d}u = 0$ which, together with the corollary in Feller (1940), yields Proposition 3.1.

8. Proof of Theorems 5.1 and 5.2

Proof of Theorem 5.1. (a) Let us express the operator T on $B(S)$ as

$$Tu(i) := \sup_{\phi\in P(A(i))}\inf_{\psi\in P(B(i))}\left\{\frac{r(i,\,\phi,\,\psi)}{\alpha+m(i)} + \frac{m(i)}{\alpha+m(i)}\sum_{j\in S}\left[\frac{q(\,j|i,\,\phi,\,\psi)}{m(i)} + \delta_{ij}\right]u(\,j)\right\}, \qquad \forall i\in S, \tag{8.1}$$

where $m(i) := 1 + q(i) > 0$ for each $i\in S$. Obviously, T is monotone. Moreover, $u_n := T\,u_{n-1} = T^n u_0$ for each $n \geqslant 1$, with $u_0 := -M\sum_{k=1}^{N}\alpha^{-k}R_k$. Thus, for each $i\in S$, $\phi\in P(A(i))$, and $\psi\in P(B(i))$, by Lemma 7.1 and (8.1) we obtain

$$\begin{aligned} u_1(i) &\geqslant -\frac{MW(i)}{\alpha+m(i)} - \frac{m(i)}{\alpha+m(i)}\left[\frac{M\sum_{k=1}^{N}\alpha^{-k}R_{k+1}(i)}{m(i)} + M\sum_{k=1}^{N}\alpha^{-k}R_k(i)\right] \\ &= -\left[\frac{MW(i)}{\alpha+m(i)} - M\frac{\alpha^{-1}m(i)}{\alpha+m(i)}R_1(i)\right] - M\sum_{k=2}^{N}\left[\frac{\alpha^{-k+1}}{\alpha+m(i)} + \frac{\alpha^{-k}m(i)}{\alpha+m(i)}\right]R_k(i) \\ &= -M\sum_{k=1}^{N}\alpha^{-k}R_k(i) = u_0(i). \end{aligned} \tag{8.2}$$

Therefore

$$-M\sum_{k=1}^{N}\alpha^{-k}R_k = u_0 \leqslant u_1 \leqslant \ldots \leqslant u_n \ldots,$$

and so $u_n \uparrow u^*$ for some function u^*. Hence, assuming for a moment that u^* is in $B(S)$, we have $Tu^* \geqslant Tu_n = u_{n+1}$ for all $n \geqslant 1$, which gives

$$Tu^* \geqslant u^*. \tag{8.3}$$

We shall now prove that $u^* \in B(S)$. To do so, it suffices to show that $u_n \leqslant -u_0$ for all $n \geqslant 0$, that is,

$$u_n(i) \leqslant M\sum_{k=1}^{N}\alpha^{-k}R_k(i), \qquad \forall n \geqslant 0 \text{ and } i\in S. \tag{8.4}$$

We prove (8.4) by induction.

When $n = 0$, (8.4) is obvious because $u_0 \leqslant 0 \leqslant -u_0$. Suppose now that (8.4) holds for some $n \geqslant 0$. Then, as in the proof of (8.2), by the induction hypothesis, Assumption C(4) and Lemma 7.1, we have

$$
\begin{aligned}
|u_{n+1}(i)| &\leqslant \frac{MW(i)}{\alpha+m(i)} + \sup_{\phi\in P(A(i))} \inf_{\psi\in P(B(i))} \left\{ \frac{m(i)}{\alpha+m(i)} \sum_{j\in S} \left[\frac{q(j|i,\phi,\psi)}{m(i)} + \delta_{ij} \right] \left[M \sum_{k=1}^{N} \alpha^{-k} R_k(j) \right] \right\} \\
&\leqslant \frac{MW(i)}{\alpha+m(i)} + \frac{m(i)}{\alpha+m(i)} \left[\frac{M\sum_{k=1}^{N-1} \alpha^{-k} R_{k+1}(i)}{m(i)} + M \sum_{k=1}^{N} \alpha^{-k} R_k(i) \right] \\
&= M \sum_{k=1}^{N} \alpha^{-k} R_k(i),
\end{aligned}
$$

which implies that (8.4) holds for $n+1$, and so it holds for all $n \geqslant 1$ and $i \in S$. Thus, the proof of (a) is completed.

(b) To prove this part, we need only show that (8.3) holds with equality because it is easily seen that (5.4) is equivalent to the fixed-point equation $u = Tu$. Now, for each fixed $n \geqslant 1$, $i \in S$ and $\phi \in P(A(i))$, we have proved that u_n is in $B(S)$, whereas, by Assumption C(1), $P(B(i))$ is compact. Thus, by Lemma 7.2 there exists $\psi_n^* \in P(B(i))$, which may depend on i and ϕ, such that

$$
\begin{aligned}
u_{n+1}(i) &\geqslant \inf_{\psi\in P(B(i))} \left\{ \frac{r(i,\phi,\psi)}{\alpha+m(i)} + \frac{m(i)}{\alpha+m(i)} \sum_{j\in S} \left[\frac{q(j|i,\phi,\psi)}{m(i)} + \delta_{ij} \right] u_n(j) \right\} \\
&= \frac{r(i,\phi,\psi_n^*)}{\alpha+m(i)} + \frac{m(i)}{\alpha+m(i)} \sum_{j\in S} \left[\frac{q(j|i,\phi,\psi_n^*)}{m(i)} + \delta_{ij} \right] u_n(j). \qquad (8.5)
\end{aligned}
$$

Since $P(B(i))$ is compact, without loss of generality we may suppose that $\psi_n^* \to \psi^* \in P(B(i))$. Therefore, as $-M\sum_{k=1}^{N}\alpha^{-k}W(i) \leqslant u_n \uparrow u^* \leqslant M\sum_{k=1}^{N}\alpha^{-k}W(i)$ for all $n \geqslant 1$, by the 'extended Fatou Lemma' 8.3.7(b) in Hernández-Lerma and Lasserre (1999) and Lemma 7.2 above, letting $n \to \infty$ in (8.5), we obtain

$$
\begin{aligned}
u^*(i) &\geqslant \frac{r(i,\phi,\psi^*)}{\alpha+m(i)} + \frac{m(i)}{\alpha+m(i)} \sum_{j\in S} \left[\frac{q(j|i,\phi,\psi^*)}{m(i)} + \delta_{ij} \right] u^*(j) \\
&\geqslant \inf_{\psi\in P(B(i))} \left\{ \frac{r(i,\phi,\psi)}{\alpha+m(i)} + \frac{m(i)}{\alpha+m(i)} \sum_{j\in S} \left[\frac{q(j|i,\phi,\psi)}{m(i)} + \delta_{ij} \right] u^*(j) \right\}. \qquad (8.6)
\end{aligned}
$$

As (8.6) holds for all $\phi \in P(A(i))$ and $i \in S$, we conclude that

$$u^* \geqslant Tu^*,$$

which, together with (8.3), gives $u^* = Tu^*$. Therefore, part (b) is proved.

(c) For each $i \in S$, $\phi \in P(A(i))$ and $\psi \in P(B(i))$, let

$$H(i,\phi,\psi) := \frac{r(i,\phi,\psi)}{\alpha+m(i)} + \frac{m(i)}{\alpha+m(i)} \sum_{j\in S}\left[\frac{q(j|i,\phi,\psi)}{m(i)} + \delta_{ij}\right] u^*(j).$$

Obviously, $H(i,\phi,\psi)$ is concave in ϕ and convex in ψ (as it is linear in both of them). Thus, Fan's (1953) minimax theorem gives part (c).

(d) Let $\pi^{*i} \in \prod_i^s$ $(i = 1, 2)$ be as in part (c). For any $\pi^1 = (\pi_t^1) \in \prod_1$, we have $\pi_t^1(\cdot|i) \in P(A(i))$ for all $t \geqslant 0$ and $i \in S$, and from (5.6), (3.1)–(3.2) and (5.1)–(5.2) we have

$$\alpha u^*(i) \geqslant r(t,i,\pi^1,\pi^{*2}) + \sum_{j\in S} q(j|t,i,\pi^1,\pi^{*2})u^*(j), \qquad \forall\,(t,i)\in\bar{S}. \tag{8.7}$$

By (8.7) and Proposition 4.2(c) we obtain that

$$u^*(i) \geqslant V_\alpha(s,i,\pi^1,\pi^{*2}), \qquad \forall\,\pi^1\in\Pi_1 \text{ and } (s,i)\in\bar{S},$$

which in turn implies that

$$u^*(i) \geqslant U(s,i), \qquad \forall\,(s,i)\in\bar{S}. \tag{8.8}$$

A similar argument gives

$$u^*(i) \leqslant V_\alpha(s,i,\pi^{*1},\pi^2), \qquad \forall\,\pi^2\in\Pi_2 \text{ and } (s,i)\in\bar{S},$$

so that

$$u^*(i) \leqslant L(s,i), \qquad \forall\,(s,i)\in\bar{S}. \tag{8.9}$$

By (8.9) and (8.8) we obtain

$$L(s,i) \geqslant u^*(i) \geqslant U(s,i), \qquad \forall\,(s,i)\in\bar{S},$$

which, together with (4.3), gives part (d).

(e) follows from (d), (5.5) and Proposition 4.2(a).

Finally, as in the proof of part (c), for each fixed $n \geqslant 1$ and $i \in S$, Fan's (1953) minimax theorem gives the existence of the sequence $\{(\phi_n^*(i), \psi_n^*(i))\}$ satisfying (5.9)–(5.10). By the 'extended Fatou Lemma' 8.3.7(b) in Hernández-Lerma and Lasserre (1999) and our Lemma 7.2 above, letting $n\to\infty$ in (5.9)–(5.10), we have that (5.5)–(5.7) hold with (π^{1^*}, π^{*2}) replaced by (ϕ^*, ψ^*). Then, as in the proof of Theorem 5.1(e), we conclude that part (f) is also true. □

Proof of Theorem 5.2. (a) A straightforward calculation using (4.6)–(4.8) shows that

$$\sum_{j\in S} q(j|i,a,b)W(j) \leqslant W(i), \qquad \forall\,(i,a,b)\in K. \tag{8.10}$$

Thus, if $\bar{q} := \|q\| < \infty$, then replacing $q(i)$ in the proof of Theorem 5.1 with $\bar{q}$, it follows from (8.1), (8.10) and (4.9) that

$$\begin{aligned}|Tu(i) - Tv(i)| &\leqslant \frac{1+\bar{q}}{1+\alpha+\bar{q}}\|u-v\|_W\left[\sum_{j\in S}\frac{q(j|i,\phi,\psi)}{1+\bar{q}}W(j) + W(i)\right]\\ &\leqslant \frac{1+\bar{q}}{1+\alpha+\bar{q}}\|u-v\|_W\left[\frac{W(i)}{1+\bar{q}} + W(i)\right]\\ &= \gamma\|u-v\|_W W(i),\end{aligned}$$

and so

$$\|Tu - Tv\|_W \leqslant \gamma\|u-v\|_W, \qquad \forall u, v \in B(S), \tag{8.11}$$

with $\gamma := (2+\bar{q})/(1+\alpha+\bar{q})$. Hence, as $\alpha > 1$, we conclude from (8.11) that T is a *contraction* on $B(S)$ with modulus $\gamma < 1$. This fact immediately yields part (a), as well as the existence of a unique solution (or fixed point) $u^* \in B(S)$ to the optimality equation $u = Tu$.

(b) By Theorem 5.1(a)–(d) and (8.4) we obtain

$$|u^*(i)| \leqslant M\sum_{k=1}^{N}\alpha^{-k}R_k(i) \leqslant M\sum_{k=1}^{N}\alpha^{-k}W(i).$$

Hence

$$\|u^*\|_W \leqslant M\sum_{k=1}^{N}\alpha^{-k}, \tag{8.12}$$

which, together with (8.11), gives part (b). □

9. Concluding remarks

In this paper we have studied zero-sum games for continuous-time Markov chains with respect to a discounted payoff criterion. Under reasonably mild assumptions we have shown that the game has a value, and also the existence of a unique solution to the optimality equation (also known as the Shapley or dynamic programming equation), and the existence of a pair of optimal stationary strategies. In addition, we have shown the convergence of the value iteration scheme. We believe that our formulation and approach are sufficiently general and, thus, provide a way to analyse other important problems, such as minimax control problems, which, as far as we know, have not been previously studied for continuous-time Markov chains. Research on these topics is in progress.

Other types of results are possible in the context of Theorem 5.1. For instance, we can easily obtain a 'martingale characterization' of optimal strategies, similar to that in Hernández-Lerma and Lasserre (2001) for discrete-time ergodic games.

We should also mention that our proof techniques can be simplified, of course, if we impose additional assumptions. For instance, if the payoff rate function $r(i, a, b)$ is *bounded*, then Assumptions B and C, as well as some of our arguments can be simplified in an obvious manner.

To conclude, it is worth noting that the *recursive* definition of the sequence $\{u_n\}$ in Theorem 5.1(a) may provide a useful way to compute, or at least to approximate, the value u^* of the game, as in the 'bounded' case of Theorem 5.2.

Acknowledgements

Research partially supported by CONACYT Grant 37355–E. The research of the first author has been partially supported by NCET and EYTP and the Natural Science Foundation of China.

References

Altman, E. (2005) Applications of dynamic games in queues. In A.S. Nowak and K. Szajowski (eds), *Advances in Dynamic Games*. Boston: Birkhauser.

Anderson, W.J. (1991) *Continuous-Time Markov Chains*. New York: Springer-Verlag.

Ardanuy, R. and Alcalá, A. (1992) Weak infinitesimal operators and stochastic differential games. *Stochastica*, **13**, 5–12.

Basar, T. and Olsder, G.J. (1999) *Dynamic Noncooperative Game Theory*, 2nd edition. Philadelphia: Society for Industrial and Applied Mathematics.

Chung, K.L. (1960) *Markov Chains with Stationary Transition Probabilities*. Berlin: Springer-Verlag.

Fan, K. (1953) Minimax theorems. *Proc. Natl. Acad. Sci. USA*, **39**, 42–47.

Fainberg, E.A. (2004) Continuous-time discounted jump Markov decision processes: a discrete-event approach. *Math. Oper. Res.*, **29**, 492–524.

Feller, W. (1940) On the integro-differential equations of purely discontinuous Markoff processes. *Trans. Amer. Math. Soc.*, **48**, 488–515.

Filar, J.A. and Vrieze, K. (1997) *Competitive Markov Decision Processes*. New York: Springer-Verlag.

Guo, X.P. and Hernández-Lerma, O. (2003a) Continuous-time controlled Markov chains. *Ann. Appl. Probab.*, **13**, 363–388.

Guo, X.P. and Hernández-Lerma, O. (2003b) Drift and monotonicity conditions for continuous-time Markov control processes with an average criterion. *IEEE Trans. Automat. Control*, **48**, 236–245.

Guo, X.P. and Liu, K. (2001) A note on optimality conditions for continuous-time Markov decision processes. *IEEE Trans. Automat. Control*, **146**, 1984–1989.

Guo, X.P. and Zhu, W.P. (2002) Denumerable state continuous time Markov decision processes with unbounded transition and reward rates under discounted criterion. *J. Appl. Probab.*, **39**, 233–250.

Hamadène, S. (1999) Nonzero sum linear-quadratic stochastic differential games and backward-forward equations. *Stochastic Anal. Appl.*, **17**, 117–130.

Hernández-Lerma, O. (1994) *Lectures on Continuous-Time Markov Control Processes*. Mexico City: Sociedad Matemática Mexicana.

Hernández-Lerma, O. and Lasserre, J.B. (1999) *Further Topics on Discrete-Time Markov Control Processes*. New York: Springer-Verlag.

Hernández-Lerma, O. and Lasserre, J.B. (2001) Zero-sum stochastic games in Borel spaces: average payoff criterion. *SIAM J. Control Optim.*, **39**, 1520–1539.

Hou, Z.T. (1994) *The Q-Matrix Problems on Markov Processes*. Changsha: Science and Technology Press of Hunan. (In Chinese.)

Hou, Z.T. and Guo, X.P. (1998) *Markov Decision Processes*. Changsha: Science and Technology Press of Hunan. (In Chinese.)

Lai, H.C. and Tanaka, K. (1984) On an *N*-person noncooperative Markov game with a metric state space. *J. Math. Anal. Appl.*, **101**, 78–96.

Lal, A.K. and Sinha, S. (1992) Zero-sum two-person semi-Markov games. *J. Appl. Probab.*, **29**, 56–72.

Lippman, S.A. (1973) Semi-Markov decision processes with unbounded rewards. *Management Sci.*, **19**, 717–731.

Lippman, S.A. (1975) On dynamic programming with unbounded rewards. *Management Sci.*, **21**, 1225–1233.

Puterman, M.L. (1994) *Markov Decision Processes*. New York: Wiley.

Ramachandran, K.M. (1999) A convergence method for stochastic differential games with a small parameter. *Stochastic Anal. Appl.*, **17**, 219–252.

Sennott, L.I. (1994) Zero-sum stochastic games with unbounded cost: discounted and average cost cases. *Z. Oper. Res.*, **39**, 209–225.

Sennott, L.I. (1999) *Stochastic Dynamic Programming and the Control of Queueing Systems*. New York: Wiley.

Shapley, L. (1953) Stochastic games. *Proc. Natl. Acad. Sci. USA*, **39**, 1095–1100.

Tanaka, K. and Homma, H. (1978) Continuous time non-cooperative *n*-person Markov games. *Bull. Math. Statist.*, **15**, 93–105.

Tanaka, K. and Wakuta, K. (1978) On continuous time Markov games with countable state space. *J. Oper. Res. Soc. Japan*, **21**, 17–27.

Tijms, H.C. (1994) *Stochastic Models: An Algorithmic Approach*. Chichester: Wiley.

Van Nunen, J.A.E.E. and Wessels, J. (1978) A note on dynamic programming with unbounded rewards. *Management Sci.*, **24**, 576–580.

Yushkevich, A.A. and Fainberg, E.A. (1979) On homogeneous Markov models with continuous time and finite or countable state space. *Theory Probab. Appl.*, **24**, 156–161.

Received June 2003 and revised March 2005