

# Approximation of zero-sum continuous-time Markov games under the discounted payoff criterion

Tomás Prieto-Rumeau · José María Lorenzo

Received: 1 April 2014 / Accepted: 20 November 2014 / Published online: 10 December 2014  
© Sociedad de Estadística e Investigación Operativa 2014

**Abstract** We deal with a two-person zero-sum continuous-time Markov game  $\mathcal{G}$  with denumerable state space, general action spaces, and unbounded payoff and transition rates. We consider noncooperative equilibria for the discounted payoff criterion. We are interested in approximating numerically the value and the optimal strategies of  $\mathcal{G}$ . To this end, we propose a definition of a sequence of game models  $\mathcal{G}_n$  converging to  $\mathcal{G}$ , which ensures that the value and the optimal strategies of  $\mathcal{G}_n$  converge to those of  $\mathcal{G}$ . For numerical purposes, we construct finite state and actions game models  $\mathcal{G}_n$  that can be explicitly solved, and we study the convergence rate of the value of the games. A game model based on a population system illustrates our results.

**Keywords** Continuous-time zero-sum Markov games · Discounted payoff criterion · Approximation of game models

**Mathematics Subject Classification** 91A15 · 91A25 · 60J27

## 1 Introduction

This paper deals with a two-person zero-sum continuous-time Markov game with denumerable state space, general action spaces, and possibly unbounded payoff and transition rates. The optimality criterion consists in finding a Nash equilibrium for the

---

T. Prieto-Rumeau (✉)  
Department of Statistics and Operations Research, UNED, Madrid, Spain  
e-mail: tprieto@ccia.uned.es

J. M. Lorenzo  
Department of Statistics and Operations Research II, Universidad Complutense de Madrid,  
Madrid, Spain  
e-mail: j.lorenzo@ccee.ucm.es

total expected discounted payoff of the players. The existence of such Nash equilibrium, as well as the existence of optimal strategies for the players, has been established in [Guo and Hernández-Lerma \(2005\)](#). In that reference, it is shown that the value of the game is the solution of an optimality equation (also referred to as the Shapley equation).

Now we explain, somehow loosely, the form of this Shapley equation. Let  $i \in S$  be the state of the system, and denote by  $a \in A$  and  $b \in B$  the actions of the players that take values in some Borel spaces  $A$  and  $B$ . There is some operator  $\mathbf{H}$  that maps, for each fixed  $a \in A$  and  $b \in B$ , a function  $\{u(i)\}_{i \in S}$  into the function  $\{\mathbf{H}u\}(i, a, b)\}_{i \in S}$  such that the value of the game  $\{V(i)\}_{i \in S}$  is the unique solution of the equations

$$V(i) = \sup_{\phi \in \mathcal{P}(A)} \inf_{\psi \in \mathcal{P}(B)} \int_A \int_B (\mathbf{H}V)(i, a, b) \psi(db) \phi(da) \quad (1.1)$$

$$= \inf_{\psi \in \mathcal{P}(B)} \sup_{\phi \in \mathcal{P}(A)} \int_A \int_B (\mathbf{H}V)(i, a, b) \psi(db) \phi(da) \quad (1.2)$$

for all  $i \in S$ , where  $\mathcal{P}(A)$  and  $\mathcal{P}(B)$  denote the family of probability measures on  $A$  and  $B$ , respectively.

It should be clear that one cannot expect to solve, in general, the Eqs. (1.1)–(1.2) explicitly. For computational purposes, therefore, one should use some kind of discretization technique to, at least, approximate the value of the game and the optimal strategies of the players. This is precisely the purpose of this paper.

Let  $\mathcal{G}$  be the “original” game model, with denumerable state space and Borel action spaces, and let  $\{\mathcal{G}_n\}_{n \geq 1}$  be a sequence of game models. In this paper, we propose a definition of the convergence  $\mathcal{G}_n \rightarrow \mathcal{G}$  which, under adequate conditions, implies that the value of the games  $\mathcal{G}_n$  and the corresponding optimal strategies converge to the value and the optimal strategies of the game  $\mathcal{G}$ . Then, for computational purposes, we show how we can construct, starting from the game model  $\mathcal{G}$ , a sequence of game models  $\{\mathcal{G}_n\}_{n \geq 1}$  with finite state and action spaces that converge to  $\mathcal{G}$ . Such finite models can be solved explicitly and, hence, we can provide computable approximations of the value of the game model  $\mathcal{G}$ . Our proofs make use of separability of spaces of probability measures in the Wasserstein metric.

As far as we know, this is the first attempt to provide such computable approximations for continuous-time Markov games with denumerable state space and general action spaces. The reader interested in related works can consult ([Jaśkiewicz and Nowak 2006](#); [Nowak and Altman 2002](#)), in which the idea of approximating a game model  $\mathcal{G}$  with “simpler” models has been studied. The reference ([Chang et al. 2010](#)) also considers computational issues for a continuous-time game with general state space and finite action spaces. Approximation techniques similar to those developed in the present paper, but for continuous-time controlled (say, with a single player) Markov chains, have been studied in [Prieto-Rumeau and Hernández-Lerma \(2012\)](#) for the discounted reward criterion and in [Prieto-Rumeau and Lorenzo \(2010\)](#) for the average cost criterion; see also [Prieto-Rumeau and Hernández-Lerma \(2012\)](#). The reference [Guo and Zhang \(2014\)](#) proposes approximation techniques for discounted cost Markov decision processes

with constraints (the previous references are concerned with unconstrained problems). Their setting is similar to ours, in the sense that they propose a definition of convergence for control models. The technique proofs in Guo and Zhang (2014) mainly rely on linear programming, while (Prieto-Rumeau and Hernández-Lerma 2012; Prieto-Rumeau and Lorenzo 2010) use dynamic programming arguments.

At this point, it is interesting to make a comparison between the approximation approaches for control and game models. Approximating a game model by means of finite state and actions game models is, from a technical point of view, more complicated than such approximations for control models. The analogous to (1.1)–(1.2) for a control model, in which the state space is  $S$  and the action space of the controller is  $A$ , is the optimality (or dynamic programming) equation

$$V(i) = \sup_{a \in A} \{(\mathbf{H}V)(i, a)\} \quad \text{for } i \in S. \quad (1.3)$$

When making a finite approximation, one roughly considers an optimality equation as in (1.3) with finite  $S$  and  $A$ . Then, one can use, for instance, the policy iteration algorithm that solves this optimality equation in a finite number of steps. For a game model, however, the Eqs. (1.1)–(1.2) are, even in the case of finite  $S$ ,  $A$ , and  $B$ , of a continuous nature because we are optimizing on a set of probability measures (say, a simplex). This makes the computational problems less straightforward. Here, we combine linear programming with a “value iteration” algorithm to solve such problems. Moreover, from a computational perspective, the maximum of a function (as in (1.3)) is easier to approximate than the saddle point of a function (as in (1.1)–(1.2)).

The rest of the paper is organized as follows. In Sect. 2, we define the game models we will be dealing with, state our main assumptions, and recall some previously known results. The main convergence results are presented in Sect. 3. In Sect. 4, these results are specialized to the case of finite state and actions approximations. Moreover, under some additional conditions, we can obtain the convergence rate of the value of the games. Finally, in Sect. 5, we make an application to a population system for which we provide some explicit approximations.

## 2 Definitions and assumptions

The definition of the game model in this section and the corresponding results are mainly based on Guo et al. (2003); Guo and Hernández-Lerma (2005) and (Prieto-Rumeau and Hernández-Lerma 2012, Chapter 10).

### 2.1 The game model $\mathcal{G}$

**Definition of the game model.** We consider a two-player zero-sum continuous-time Markov game model  $\mathcal{G} = \{S, A, B, \mathbb{K}, Q, r\}$ , where the elements of  $\mathcal{G}$  are the following.

- $S = \{0, 1, 2, \dots\}$  is the state space.
- $A$  and  $B$  are the action spaces for players 1 and 2, respectively. We assume that  $A$  and  $B$  are Borel spaces (i.e., measurable subsets of complete and separable metric spaces). The corresponding metrics are denoted by  $d_A$  and  $d_B$ , respectively.
- For each  $i \in S$ , the measurable sets  $A(i) \subseteq A$  and  $B(i) \subseteq B$  stand for the actions available for players 1 and 2 in state  $i \in S$ , respectively. (Throughout this paper, measurability is always understood with respect to the Borel  $\sigma$ -algebra). The family of feasible triplets is defined as:

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

- The transition rate matrix of the system is  $Q = [q_{ij}(a, b)]$ . It is assumed that:
  1. The function  $(a, b) \mapsto q_{ij}(a, b)$  is measurable on  $A(i) \times B(i)$  for all  $i, j \in S$ ;
  2. The transition rates are conservative, that is,

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

with  $q_{ij}(a, b) \geq 0$  whenever  $i \neq j$ ;

3. The transition rates are stable, i.e.,  $q(i) := \sup_{a \in A(i), b \in B(i)} \{-q_{ii}(a, b)\} < \infty$ .
- Finally, the measurable function  $r : \mathbb{K} \rightarrow \mathbb{R}$  is interpreted as the reward rate function for player 1 and the cost rate function for player 2.

The game  $\mathcal{G}$  is played as follows. At each time  $t \geq 0$ , both players observe the state of the system  $x(t) = i \in S$  and then, independently and simultaneously, they choose actions  $a(t) = a \in A(i)$  and  $b(t) = b \in B(i)$ . In a small time interval  $[t, t + dt]$ :

- player 1 receives a reward  $r(i, a, b)dt$ ,
- player 2 incurs a cost  $r(i, a, b)dt$ ,
- the system remains in state  $i \in S$  with probability  $1 + q_{ii}(a, b)dt$  or makes a transition to the state  $j \neq i$  with probability  $q_{ij}(a, b)dt$ .

This procedure is carried on over all the time horizon  $t \in [0, \infty)$ .

We suppose that the reward/cost of the players is depreciated at a constant discount rate  $\alpha > 0$ , and so the reward/cost  $r(i, a, b)$  at time  $t \geq 0$  is brought to its present value, that is,  $e^{-\alpha t}r(i, a, b)$ . The goal of player 1 is to maximize his total expected discounted reward, loosely,

$$E \left[ \int_0^\infty e^{-\alpha t} r(x(t), a(t), b(t)) dt \right],$$

while player 2 wants to minimize his total expected discounted cost. A formal definition will be given below.

**Assumptions on the game model.** Next, we state all the assumptions we will make on the game model  $\mathcal{G}$ . We will use the following terminology.

**Definition 2.1** We say that  $w : S \rightarrow [1, \infty)$  is a Lyapunov function on  $S$  when  $w$  is monotone nondecreasing and, in addition,  $\lim_{i \rightarrow \infty} w(i) = +\infty$ .

**Assumption 2.2** The game model  $\mathcal{G}$  satisfies the following conditions.

- (i) There exists a Lyapunov function  $w$  on  $S$ , and constants  $c_1 < \alpha$  and  $d_1 \geq 0$  with

$$\sum_{j \in S} q_{ij}(a, b)w(j) \leq c_1w(i) + d_1 \quad \text{for all } (i, a, b) \in \mathbb{K}.$$

For each  $i \in S$ , we have  $q(i) \leq w(i)$ .

- (ii) There exists a constant  $M > 0$  such that  $|r(i, a, b)| \leq Mw(i)$  for all  $(i, a, b) \in \mathbb{K}$ .
- (iii) For each  $i \in S$ , the sets  $A(i)$  and  $B(i)$  are compact. For all  $i, j \in S$ , the functions  $r(i, a, b)$  and  $q_{ij}(a, b)$  are continuous on  $A(i) \times B(i)$ .
- (iv) There exist constants  $c_2 \in \mathbb{R}$  and  $d_2 \geq 0$  with

$$\sum_{j \in S} q_{ij}(a, b)w^2(j) \leq c_2w^2(i) + d_2 \quad \text{for all } (i, a, b) \in \mathbb{K}.$$

In the sequel, we will explain the use of the above hypotheses.

**Strategies of the players.** We introduce some notation. Let  $\bar{A}(i)$  and  $\bar{B}(i)$  be the families of probability measures on  $A(i)$  and  $B(i)$ , when endowed with their Borel  $\sigma$ -algebras  $\mathbb{B}(A(i))$  and  $\mathbb{B}(B(i))$ , respectively. Let

$$\pi^1 \equiv \{\pi_t^1(C|i)\}_{t \geq 0, i \in S, C \in \mathbb{B}(A(i))}$$

be such that  $\pi_t^1(\cdot|i)$  is in  $\bar{A}(i)$  for all  $t \geq 0$  and  $i \in S$ , and such that  $t \mapsto \pi_t^1(C|i)$  is a measurable function on  $[0, \infty)$  for all  $C \in \mathbb{B}(A(i))$  and  $i \in S$ . We say that  $\pi^1$  is a randomized Markov strategy for player 1, and we denote by  $\Pi^1$  the set of all such strategies. The family  $\Pi^2$  of randomized Markov strategies

$$\pi^2 \equiv \{\pi_t^2(C|i)\}_{t \geq 0, i \in S, C \in \mathbb{B}(B(i))}$$

for player 2 is defined similarly.

We say that  $\pi^1 \in \Pi^1$  is a randomized stationary strategy for player 1 when  $\pi_t^1(C|i)$  does not depend on  $t \geq 0$ . Thus, the class  $\Pi^{1,s}$  of stationary strategies for player 1 can be identified with

$$\Pi^{1,s} = \prod_{i \in S} \bar{A}(i).$$

Similarly, the class of randomized stationary strategies for player 2 is  $\Pi^{2,s} = \prod_{i \in S} \bar{B}(i)$ .

Given a pair of strategies  $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$ , define

$$q_{ij}(t, \pi^1, \pi^2) = \int_{A(i)} \int_{B(i)} q_{ij}(a, b)\pi_t^2(db|i)\pi_t^1(da|i).$$

The above integral is well defined and finite because the system’s transition rates are conservative and stable. In particular, they satisfy

$$-q_{ii}(t, \pi^1, \pi^2) = \sum_{j \neq i} q_{ij}(t, \pi^1, \pi^2) \leq q(i) \quad \text{for each } t \geq 0 \text{ and } i \in S.$$

We will also use the following notation. Given  $i, j \in S, \phi \in \bar{A}(i)$ , and  $\psi \in \bar{B}(i)$ , let

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

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

and for stationary strategies  $(\pi^1, \pi^2) \in \Pi^{1,s} \times \Pi^{2,s}$ , we write

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

Our next result summarizes the main results on the existence of the state and actions processes. See, e.g., Proposition 3.1 in Guo et al. (2003), Assumption A in Guo and Hernández-Lerma (2005), or Proposition 10.3 in Prieto-Rumeau and Hernández-Lerma (2012).

**Theorem 2.3** *Suppose that Assumption 2.2(i) is satisfied.*

- (i) *For every  $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$ , there exists a regular (nonhomogeneous) transition function*

$$\{P_{ij}^{\pi^1, \pi^2}(s, t)\}_{i, j \in S, 0 \leq s \leq t}$$

*with transition rates  $q_{ij}(t, \pi^1, \pi^2)$ , that is,*

$$\lim_{h \downarrow 0} \frac{P_{ij}^{\pi^1, \pi^2}(t, t+h) - \delta_{ij}}{h} = q_{ij}(t, \pi^1, \pi^2) \quad \text{for all } i, j \in S \text{ and } t \geq 0.$$

*Let  $\Omega = \mathbb{K}^{[0, \infty)} = \{(x(t), a(t), b(t))\}_{t \geq 0}$  be endowed with the product  $\sigma$ -algebra  $\mathcal{F}$ .*

- (ii) *Given an initial state  $i \in S$  at time 0 and  $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$ , there exists a unique probability measure  $P^{i, \pi^1, \pi^2}$  on  $(\Omega, \mathcal{F})$  such that:*
- *For each  $A_0 \in \mathbb{B}(A(i))$  and  $B_0 \in \mathbb{B}(B(i))$ , we have*

$$P^{i, \pi^1, \pi^2}\{x(0) = i, a(0) \in A_0, b(0) \in B_0\} = \pi_0^1(A_0|i) \cdot \pi_0^2(B_0|i).$$

- Given arbitrary  $n \geq 1$  and  $0 \leq s_1 \leq s_2 \leq \dots \leq s_n$ , and, on the other hand, given  $i_k \in S$ ,  $A_k \in \mathbb{B}(A(i_k))$ , and  $B_k \in \mathbb{B}(B(i_k))$ , for  $k = 1, \dots, n$ , we have

$$\begin{aligned}
 P^{i, \pi^1, \pi^2} \{ & x(s_1) = i_1, a(s_1) \in A_1, b(s_1) \in B_1, \dots, \\
 & x(s_n) = i_n, a(s_n) \in A_n, b(s_n) \in B_n \} \\
 &= \prod_{k=1}^n P_{i_{k-1}i_k}^{\pi^1, \pi^2}(s_{k-1}, s_k) \pi_{s_k}^1(A_k | i_k) \pi_{s_k}^2(B_k | i_k),
 \end{aligned}$$

with the convention that  $i_0 = i$  and  $s_0 = 0$ .

We will refer to  $\{x(t)\}_{t \geq 0}$  as the state process, while  $\{a(t)\}_{t \geq 0}$  and  $\{b(t)\}_{t \geq 0}$  are the action processes for players 1 and 2. The expectation operator associated with  $P^{i, \pi^1, \pi^2}$  will be denoted by  $E^{i, \pi^1, \pi^2}$ .

Let  $\mathcal{B}_w(S)$  be the family of functions  $u : S \rightarrow \mathbb{R}$  such that

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

We have that  $\|\cdot\|_w$  is a norm on  $\mathcal{B}_w(S)$ , under which it is a Banach space.

**The discounted reward/cost optimality criterion.** For the discount rate  $\alpha > 0$ , given an initial state  $i \in S$  and a pair of strategies  $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$ , we define the total expected discounted payoff as

$$V^\alpha(i, \pi^1, \pi^2) = E^{i, \pi^1, \pi^2} \left[ \int_0^\infty e^{-\alpha t} r(x(t), a(t), b(t)) dt \right]. \tag{2.3}$$

Thus,  $V^\alpha(i, \pi^1, \pi^2)$  is the total expected discounted reward for player 1, and it is the total expected discounted cost for player 2. Using Guo and Hernández-Lerma 2003, Lemma 3.2 or Guo and Hernández-Lerma 2009, Lemma 6.3, under Assumption 2.2(i), we have that

$$E^{i, \pi^1, \pi^2} [w(x(t))] \leq e^{c_1 t} w(i) + \frac{d_1}{c_1} (e^{c_1 t} - 1). \tag{2.4}$$

(if  $c_1 = 0$ , then the righthand term is  $w(i) + d_1 t$ ; to see this, just let  $c_1 \downarrow 0$  in (2.4)). As a consequence of Assumption 2.2(ii) given  $i \in S$  and  $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$ , we have

$$|V^\alpha(i, \pi^1, \pi^2)| \leq M \int_0^\infty e^{-\alpha t} E^{i, \pi^1, \pi^2} [w(x(t))] dt \leq \frac{Mw(i)}{\alpha - c_1} + \frac{d_1 M}{\alpha(\alpha - c_1)}.$$

Hence, letting  $\mathfrak{M} = \frac{M(\alpha + d_1)}{\alpha(\alpha - c_1)}$ , it follows that

$$\|V^\alpha(\cdot, \pi^1, \pi^2)\|_w \leq \mathfrak{M} \text{ for all } (\pi^1, \pi^2) \in \Pi^1 \times \Pi^2. \tag{2.5}$$

*Remark 2.4* If  $(\pi^1, \pi^2)$  is a pair of stationary strategies, by Theorem 2.3(ii) we have (cf. (2.3))

$$V^\alpha(i, \pi^1, \pi^2) = E^{i, \pi^1, \pi^2} \left[ \int_0^\infty r(x(t), \pi^1, \pi^2) dt \right] \text{ for each } i \in S.$$

Given the initial state  $i \in S$ , the *lower value* and *upper value* of  $\mathcal{G}$  are defined as:

$$L^\alpha(i) = \sup_{\pi^1 \in \Pi^1} \inf_{\pi^2 \in \Pi^2} V^\alpha(i, \pi^1, \pi^2)$$

$$U^\alpha(i) = \inf_{\pi^2 \in \Pi^2} \sup_{\pi^1 \in \Pi^1} V^\alpha(i, \pi^1, \pi^2),$$

respectively. We note that, as a consequence of (2.5), we have

$$\|L^\alpha\|_w \leq \mathfrak{M} \text{ and } \|U^\alpha\|_w \leq \mathfrak{M}.$$

The lower value of the game is the maximal discounted reward for player 1 when using a “maximin” strategy. Indeed, for every fixed strategy  $\pi^1 \in \Pi^1$ , the worst scenario for player 1 is when player 2 chooses the strategy  $\pi^2 \in \Pi^2$  attaining the infimum

$$\inf_{\pi^2 \in \Pi^2} V^\alpha(i, \pi^1, \pi^2).$$

Then, player 1 chooses the strategy yielding the maximal reward, that is, the one achieving the supremum in the definition of  $L^\alpha(i)$ . Similarly, the upper value of the game corresponds to the optimal payoff of player 2 when using a “minimax” strategy. It is easy to see that  $L^\alpha(i) \leq U^\alpha(i)$  for every  $i \in S$ .

**Definition 2.5** The game  $\mathcal{G}$  has a value when  $L^\alpha(i) = U^\alpha(i)$  for all  $i \in S$ . The function  $V^\alpha(i) := L^\alpha(i) = U^\alpha(i)$  is called the value function of the game  $\mathcal{G}$ .

In this case, we say that  $(\pi^{*1}, \pi^{*2}) \in \Pi^1 \times \Pi^2$  is a pair of discount optimal strategies when

$$V^\alpha(i, \pi^1, \pi^{*2}) \leq V^\alpha(i, \pi^{*1}, \pi^{*2}) \leq V^\alpha(i, \pi^{*1}, \pi^2)$$

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

A direct calculation shows that if  $V^\alpha$  is the value function of the game and  $(\pi^{*1}, \pi^{*2}) \in \Pi^1 \times \Pi^2$  is a pair of discount optimal strategies, then  $V^\alpha(i) = V^\alpha(i, \pi^{*1}, \pi^{*2})$  for each  $i \in S$ . A pair of optimal strategies is usually referred to as a noncooperative or Nash equilibrium of the game.

**Existence of the value function of  $\mathcal{G}$  and the Shapley equations.** The following is a consequence of our assumptions that will be useful in the forthcoming.



**Corollary 2.6** *Suppose that Assumption 2.2 holds.*

(i) *Given  $i \in S$  and  $k > i$ , we have*

$$\sum_{j \geq k} q_{ij}(\phi, \psi)w(j) \leq \frac{1}{w(k)} \left( c_2 w^2(i) + d_2 + q(i)w^2(i) \right)$$

*for all  $\phi \in \bar{A}(i)$  and  $\psi \in \bar{B}(i)$ .*

(ii) *Given arbitrary  $u \in \mathcal{B}_w(S)$ , the function  $(a, b) \mapsto \sum_{j \in S} q_{ij}(a, b)u(j)$  is continuous on  $A(i) \times B(i)$  for every  $i \in S$ .*

*Proof* (i) Given  $i \in S, k > i$ , and  $(a, b) \in A(i) \times B(i)$  we have

$$\sum_{j \geq k} q_{ij}(a, b)w(j) \leq \frac{1}{w(k)} \sum_{j \geq k} q_{ij}(a, b)w^2(j) \tag{2.6}$$

$$\begin{aligned} &\leq \frac{1}{w(k)} \left( \sum_{j \in S} q_{ij}(a, b)w^2(j) - q_{ii}(a, b)w^2(i) \right) \\ &\leq \frac{1}{w(k)} \left( c_2 w^2(i) + d_2 + q(i)w^2(i) \right), \end{aligned} \tag{2.7}$$

where we have used monotonicity of  $w$  in (2.6) and Assumptions 2.2(i) and (iv) in (2.7). By integration with respect to  $\phi(da)$  and  $\psi(db)$ , statement (i) follows.

(ii) From (2.7) we have that

$$\lim_{k \rightarrow \infty} \sup_{(a,b) \in A(i) \times B(i)} \sum_{j \geq k} q_{ij}(a, b)w(j) = 0$$

and, in particular,

$$\lim_{k \rightarrow \infty} \sup_{(a,b) \in A(i) \times B(i)} \left| \sum_{j \geq k} q_{ij}(a, b)u(j) \right| = 0.$$

Consequently, the series  $\sum q_{ij}(a, b)u(j)$  of continuous functions converges uniformly on  $A(i) \times B(i)$ . It is, therefore, a continuous function. □

The continuity of  $\sum_{j \in S} q_{ij}(a, b)w(j)$  is a usual requirement in Markov game models; see Guo and Hernández-Lerma 2005, Assumption C.3 or Prieto-Rumeau and Hernández-Lerma 2012, Assumption 10.7.b. As seen in Corollary 2.6(ii) above, this condition is in fact implied by our hypotheses.

The main result on the discounted game  $\mathcal{G}$  is the following. It is borrowed from (Guo and Hernández-Lerma 2005; Prieto-Rumeau and Hernández-Lerma 2012).

**Theorem 2.7** *Suppose that the game model  $\mathcal{G}$  satisfies Assumption 2.2.*

(i) *The game has a value  $V^\alpha \in \mathcal{B}_w(S)$  with  $\|V^\alpha\|_w \leq \mathfrak{M}$ .*

(ii) *The value function  $V^\alpha$  is the unique solution  $u$  in  $\mathcal{B}_w(S)$  of the equations*

$$\alpha u(i) = \sup_{\phi \in \bar{A}(i)} \inf_{\psi \in \bar{B}(i)} \left\{ r(i, \phi, \psi) + \sum_{j \in S} q_{ij}(\phi, \psi) u(j) \right\} \tag{2.8}$$

$$= \inf_{\psi \in \bar{B}(i)} \sup_{\phi \in \bar{A}(i)} \left\{ r(i, \phi, \psi) + \sum_{j \in S} q_{ij}(\phi, \psi) u(j) \right\} \tag{2.9}$$

for all  $i \in S$ .

(iii) There exists a pair of optimal randomized stationary strategies.

Moreover,  $(\pi^1, \pi^2) \in \Pi^{1,s} \times \Pi^{2,s}$  is a pair of optimal stationary strategies if and only if  $\pi^1(\cdot|i)$  and  $\pi^2(\cdot|i)$  attain the supremum and the infimum in (2.8) and (2.9), respectively, for every  $i \in S$ .

**Remark 2.8** The fact that there exists a pair of optimal stationary strategies implies that the value of the game satisfies

$$V^\alpha(i) = \sup_{\pi^1 \in \Pi^{1,s}} \inf_{\pi^2 \in \Pi^{2,s}} V^\alpha(i, \pi^1, \pi^2) = \inf_{\pi^2 \in \Pi^{2,s}} \sup_{\pi^1 \in \Pi^{1,s}} V^\alpha(i, \pi^1, \pi^2)$$

for all  $i \in S$ . Note that we are taking the infimum and the supremum over the family of stationary strategies (cf. the definition of the lower and upper value of the game).

### 2.2 The game models $\{\mathcal{G}_n\}_{n \geq 1}$

In the forthcoming, we consider a sequence of game models

$$\mathcal{G}_n = \{S_n, A, B, \mathbb{K}_n, Q_n, r_n\} \text{ for } n \geq 1.$$

The elements of these game models satisfy the following conditions.

- The state space  $S_n$  is a (finite or infinite) subset of  $S$ .
- The action spaces are  $A$  and  $B$ , as for the game model  $\mathcal{G}$ .
- The set of available actions in state  $i \in S_n$  is the measurable sets  $A_n(i) \subseteq A(i)$  and  $B_n(i) \subseteq B(i)$  for players 1 and 2, respectively. Let

$$\mathbb{K}_n = \{(i, a, b) \in S_n \times A \times B : a \in A_n(i), b \in B_n(i)\} \subseteq \mathbb{K}.$$

- The transition rate matrix is given by  $Q_n = [q_{ij}^n(a, b)]$  for  $i, j \in S_n$  and  $(a, b) \in A_n(i) \times B_n(i)$ . We assume that  $(a, b) \mapsto q_{ij}^n(a, b)$  is measurable on  $A_n(i) \times B_n(i)$  for all  $i, j \in S_n$ . The transition rates are assumed to be conservative and stable, that is,

$$\sum_{j \in S_n} q_{ij}^n(a, b) = 0 \text{ and } \sup_{a \in A_n(i), b \in B_n(i)} \{-q_{ii}^n(a, b)\} =: q_n(i) < \infty$$

for  $(i, a, b) \in \mathbb{K}_n$ , with the condition that  $q_{ij}^n(a, b) \geq 0$  for  $i \neq j$ .

- The reward/cost rate function is  $r_n : \mathbb{K}_n \rightarrow \mathbb{R}$ , assumed to be measurable.

The discount rate  $\alpha > 0$  is the same for all the game models  $\mathcal{G}_n$  and  $\mathcal{G}$ . The dynamics of the game  $\mathcal{G}_n$  is similar to that of the game  $\mathcal{G}$  studied previously. Next, we state our hypotheses on the sequence  $\{\mathcal{G}_n\}_{n \geq 1}$ .

**Assumption 2.9** The following statements hold for every  $n \geq 1$ .

- (i) For all  $(i, a, b) \in \mathbb{K}_n$

$$\sum_{j \in S_n} q_{ij}^n(a, b)w(j) \leq c_1w(i) + d_1,$$

where the Lyapunov function  $w$  and the constants  $c_1 < \alpha$  and  $d_1 \geq 0$  come from Assumption 2.2(i). For each  $i \in S_n$ , we have  $q_n(i) \leq w(i)$ .

- (ii) For the constant  $M > 0$  in Assumption 2.2(ii),  $|r_n(i, a, b)| \leq Mw(i)$  for all  $(i, a, b) \in \mathbb{K}_n$ .
- (iii) For each  $i \in S_n$ , the sets  $A_n(i) \subseteq A(i)$  and  $B_n(i) \subseteq B(i)$  are compact, while for all  $i, j \in S_n$ , the functions  $r_n(i, a, b)$  and  $q_{ij}^n(a, b)$  are continuous on  $A_n(i) \times B_n(i)$ .
- (iv) With  $c_2 \in \mathbb{R}$  and  $d_2 \geq 0$  as in Assumption 2.2(iv), we have

$$\sum_{j \in S_n} q_{ij}^n(a, b)w^2(j) \leq c_2w^2(i) + d_2 \quad \text{for all } (i, a, b) \in \mathbb{K}_n.$$

Roughly, Assumption 2.9 consists in supposing that Assumption 2.2 holds “uniformly” in  $n \geq 1$ . Next, we introduce some notation. For the game model  $\mathcal{G}_n$ , the families of randomized Markov strategies for players 1 and 2 are denoted by  $\Pi_n^1$  and  $\Pi_n^2$ , respectively, while the corresponding stationary strategies are

$$\Pi_n^{1,s} = \prod_{i \in S_n} \bar{A}_n(i) \quad \text{and} \quad \Pi_n^{2,s} = \prod_{i \in S_n} \bar{B}_n(i),$$

with  $\bar{A}_n(i)$  and  $\bar{B}_n(i)$  the families of probability measures on  $A_n(i)$  and  $B_n(i)$ , respectively. Let  $\mathcal{B}_w(S_n)$  be the Banach space of functions  $u : S_n \rightarrow \mathbb{R}$  with finite  $w$ -norm

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

(We note that we use the same notation  $\|u\|_w$  for  $u : S \rightarrow \mathbb{R}$  and  $u : S_n \rightarrow \mathbb{R}$ ). Notations such as

$$q_{ij}^n(\phi, \psi) \quad \text{and} \quad r_n(i, \phi, \psi)$$

for  $i, j \in S_n$ ,  $\phi \in \bar{A}_n(i)$ , and  $\psi \in \bar{B}_n(i)$  are given the obvious definitions; see (2.1)–(2.2).

We can apply Theorem 2.3 to  $\mathcal{G}_n$  and, therefore, there indeed exists a stochastic process  $\{(x(t), a(t), b(t))\}_{t \geq 0}$  taking values in  $\mathbb{K}_n$  that models the state and actions processes for the game model  $\mathcal{G}_n$ . In particular, the corresponding expectation operator

will be denoted by  $E_n^{i,\pi^1,\pi^2}$ . Given  $i \in S_n$  and  $(\pi^1, \pi^2) \in \Pi_n^1 \times \Pi_n^2$ , define the total expected discounted payoff for the game model  $\mathcal{G}_n$  as

$$V_n^\alpha(i, \pi^1, \pi^2) = E_n^{i,\pi^1,\pi^2} \left[ \int_0^\infty e^{-\alpha t} r_n(x(t), a(t), b(t)) dt \right].$$

We also have (cf. 2.5),

$$\|V_n^\alpha(\cdot, \pi^1, \pi^2)\| \leq \mathfrak{M} \quad \text{for all } \pi^1 \in \Pi_n^1 \text{ and } \pi^2 \in \Pi_n^2. \tag{2.10}$$

The lower and upper value functions of the game  $L_n^\alpha, U_n^\alpha \in \mathcal{B}_w(S_n)$ , and the value function  $V_n^\alpha$  (provided it exists) are given the usual definitions. We have a result similar to Corollary 2.6, which is stated without proof.

**Corollary 2.10** *Suppose that Assumption 2.9 holds and fix  $n \geq 1$ .*

(i) *Given  $i \in S_n$  and  $k > i$ , we have*

$$\sum_{j \geq k, j \in S_n} q_{ij}^n(\phi, \psi) w(j) \leq \frac{1}{w(k)} (c_2 w^2(i) + d_2 + q(i) w^2(i))$$

*for all  $\phi \in \bar{A}_n(i)$  and  $\psi \in \bar{B}_n(i)$ .*

(ii) *Given arbitrary  $u \in \mathcal{B}_w(S_n)$ , the function  $(a, b) \mapsto \sum_{j \in S_n} q_{ij}^n(a, b) u(j)$  is continuous on  $A_n(i) \times B_n(i)$  for every  $i \in S_n$ .*

Each discounted game model  $\mathcal{G}_n$  has a value function that can be characterized by the corresponding Shapley equations; cf. Theorem 2.7.

**Theorem 2.11** *Suppose that Assumption 2.9 is satisfied. Then, the following statements hold for each  $n \geq 1$ :*

- (i) *The game  $\mathcal{G}_n$  has a value  $V_n^\alpha \in \mathcal{B}_w(S_n)$  with  $\|V_n^\alpha\|_w \leq \mathfrak{M}$ .*
- (ii) *The value function  $V_n^\alpha$  is the unique solution  $u$  in  $\mathcal{B}_w(S_n)$  of the equations*

$$\alpha u(i) = \sup_{\phi \in \bar{A}_n(i)} \inf_{\psi \in B_n(i)} \left\{ r_n(i, \phi, \psi) + \sum_{j \in S_n} q_{ij}^n(\phi, \psi) u(j) \right\} \tag{2.11}$$

$$= \inf_{\psi \in B_n(i)} \sup_{\phi \in \bar{A}_n(i)} \left\{ r_n(i, \phi, \psi) + \sum_{j \in S_n} q_{ij}^n(\phi, \psi) u(j) \right\} \tag{2.12}$$

*for each  $i \in S_n$ .*

(iii) *There exists a pair of optimal randomized stationary strategies for the game model  $\mathcal{G}_n$ .*

*Moreover,  $(\pi^1, \pi^2) \in \Pi_n^{1,s} \times \Pi_n^{2,s}$  is a pair of optimal randomized stationary strategies if and only if  $\pi^1(\cdot|i)$  and  $\pi^2(\cdot|i)$  attain the supremum and the infimum in (2.11) and (2.12), respectively, for every  $i \in S_n$ .*

### 3 Convergence results

In what follows we will suppose that the game models  $\mathcal{G}$  and  $\{\mathcal{G}_n\}_{n \geq 1}$  satisfy Assumptions 2.2 and 2.9.

**Convergence of game models.** We propose a definition of the game models  $\mathcal{G}_n$  converging to the original game model  $\mathcal{G}$ . In this definition, we make use of the Hausdorff metric. We recall that the Hausdorff distance between two closed subsets  $C$  and  $D$  of a metric space  $(X, d_X)$  is defined as:

$$d_H(C, D) = \sup_{y \in C} \inf_{x \in D} \{d_X(x, y)\} \vee \sup_{x \in D} \inf_{y \in C} \{d_X(x, y)\}$$

(here,  $\vee$  stands for ‘‘maximum’’). The Hausdorff metric is indeed a metric on the family of closed subsets of  $X$  except that it might not be finite. We say that  $\{C_n\}_{n \geq 1}$  converges to  $C$  in the Hausdorff metric when  $d_H(C_n, C) \rightarrow 0$  as  $n \rightarrow \infty$ .

**Definition 3.1** We say that  $\mathcal{G}_n \rightarrow \mathcal{G}$  as  $n \rightarrow \infty$  when the following conditions hold:

(a) The sequence of states  $S_n \uparrow S$ .

We define  $n(i) = \min\{n \geq 1 : i \in S_n\}$  for each  $i \in S$ , and so  $i \in S_n$  if and only if  $n \geq n(i)$ .

(b) For each  $i \in S$ , the sequences of action sets  $A_n(i)$  and  $B_n(i)$  converge to  $A(i)$  and  $B(i)$  as  $n \rightarrow \infty$ , respectively, in the Hausdorff metric.

For every  $i \in S$ , given sequences  $\{a_n\}_{n \geq n(i)}$  and  $\{b_n\}_{n \geq n(i)}$ , with  $a_n \in A_n(i)$  and  $b_n \in B_n(i)$ , such that  $a_n \rightarrow a$  and  $b_n \rightarrow b$  for some  $a \in A(i)$  and  $b \in B(i)$ , we have:

(c)  $\lim_{n \rightarrow \infty} q_{ij}^n(a_n, b_n) = q_{ij}(a, b)$  for all  $j \in S$ , and

(d)  $\lim_{n \rightarrow \infty} r_n(i, a_n, b_n) = r(i, a, b)$ .

Observe that expressions such as  $d_H(A(i), A_n(i))$  or  $q_{ij}^n(a_n, b_n)$  are defined only for large enough  $n$  (namely,  $n \geq n(i)$  in the former case, and  $n \geq n(i) \vee n(j)$  in the latter). This is not made explicit in the notation since we are dealing with the limit as  $n \rightarrow \infty$ .

Our next lemma gives some equivalent statements of Definition 3.1.

**Lemma 3.2** (i) *The condition in Definition 3.1(c) can be replaced with the following statement. Given  $i, j \in S$  and  $\epsilon > 0$  there exists  $n_0 \geq n(i) \vee n(j)$  such that for all  $n \geq n_0$*

$$\sup_{(a,b) \in A_n(i) \times B_n(i)} |q_{ij}^n(a, b) - q_{ij}(a, b)| \leq \epsilon.$$

(ii) *The condition in Definition 3.1(d) can be replaced with the following statement. Given  $i \in S$  and  $\epsilon > 0$  there exists  $n_0 \geq n(i)$  such that for all  $n \geq n_0$*

$$\sup_{(a,b) \in A_n(i) \times B_n(i)} |r_n(i, a, b) - r(i, a, b)| \leq \epsilon.$$

*Proof* (i) First we prove that if Definition 3.1 holds, then (i) also holds. We proceed by contradiction. If (i) does not hold then there is some  $i, j \in S$  and  $\epsilon > 0$  such that, for infinitely many  $n \geq n(i) \vee n(j)$ , there exist  $a_n \in A_n(i)$  and  $b_n \in B_n(i)$  with

$$|q_{ij}^n(a_n, b_n) - q_{ij}(a, b)| > \varepsilon. \tag{3.1}$$

For such  $n$ , since  $a_n \in A_n(i) \subseteq A(i)$ , there exists a subsequence  $\{n'\}$  and  $a \in A(i)$  such that  $a_{n'} \rightarrow a$ . Similarly, for some subsequence, still denoted by  $\{n'\}$ , we have  $b_{n'} \rightarrow b$  for some  $b \in B(i)$ . Next, define  $\tilde{a}_n \in A_n(i)$  and  $\tilde{b}_n \in B_n(i)$ , for  $n \geq n(i)$ , as follows.

- If  $n$  belongs to the subsequence  $\{n'\}$  then let  $\tilde{a}_n = a_n$  and  $\tilde{b}_n = b_n$ ;
- Otherwise, let  $\tilde{a}_n$  and  $\tilde{b}_n$  be such that

$$d_A(\tilde{a}_n, a) = \inf_{x \in A_n(i)} d_A(x, a) \leq d_H(A_n(i), A(i))$$

$$d_B(\tilde{b}_n, b) = \inf_{y \in B_n(i)} d_B(y, b) \leq d_H(B_n(i), B(i)).$$

We have thus constructed sequences  $\tilde{a}_n \in A_n(i)$  and  $\tilde{b}_n \in B_n(i)$ , for  $n \geq n(i)$ , such that  $\tilde{a}_n \rightarrow a$  and  $\tilde{b}_n \rightarrow b$ . Consequently, by (c), for  $n$  large enough we have

$$|q_{ij}^n(\tilde{a}_n, \tilde{b}_n) - q_{ij}(a, b)| \leq \frac{\varepsilon}{2}.$$

In particular, recalling (3.1), along the subsequence  $\{n'\}$  we have

$$|q_{ij}(a_{n'}, b_{n'}) - q_{ij}(a, b)| > \frac{\varepsilon}{2}.$$

This contradicts the continuity of the transition rate function.

Conversely, let us now prove that Definition 3.1(a), (b), and (d), together with (i), imply (c). Fix  $i, j \in S$ , and let  $a_n \in A_n(i)$  and  $b_n \in B_n(i)$  be such that  $a_n \rightarrow a \in A(i)$  and  $b_n \rightarrow b \in B(i)$ . By the condition (i), given  $\varepsilon > 0$ , for  $n$  large enough we have

$$|q_{ij}^n(a_n, b_n) - q_{ij}(a, b)| \leq \frac{\varepsilon}{2}.$$

But now continuity of the function  $(a, b) \mapsto q_{ij}(a, b)$  implies that for  $n$  large enough we also have

$$|q_{ij}(a_n, b_n) - q_{ij}(a, b)| \leq \frac{\varepsilon}{2}.$$

This yields

$$|q_{ij}^n(a_n, b_n) - q_{ij}(a, b)| \leq \varepsilon$$

and so  $\lim_n q_{ij}^n(a_n, b_n) = q_{ij}(a, b)$ . This completes the proof that (c)  $\Leftrightarrow$  (i).

To prove statement (ii) we can proceed similarly. □

Given a sequence of functions  $u_n : S_n \rightarrow \mathbb{R}$ , for  $n \geq 1$ , we say that  $\{u_n\}$  converges pointwise to some function  $u : S \rightarrow \mathbb{R}$  when

$$\lim_{n \rightarrow \infty} u_n(i) = u(i) \quad \text{for all } i \in S.$$

Note that, for fixed  $i \in S$ ,  $u_n(i)$  is well defined only when  $n \geq n(i)$ . Since the above definition is concerned with the limit as  $n \rightarrow \infty$ , the requirement  $n \geq n(i)$  will not be explicit in the notation.

**Convergence of probability measures.** Now we recall some facts on convergence of probability measures; see, e.g., (Billingsley 1968, Chapter 1) or (Bogachev 2007, Chapter 8). Given a metric space  $(X, d_X)$ , let  $\mathcal{P}$  be the family of Borel probability measures on  $X$ . We say that the sequence  $\{\mu_n\} \subseteq \mathcal{P}$  converges weakly to  $\mu \in \mathcal{P}$ , and we will write  $\mu_n \xrightarrow{d} \mu$ , if

$$\lim_{n \rightarrow \infty} \int_X f d\mu_n = \int_X f d\mu \tag{3.2}$$

for all bounded and continuous functions  $f : X \rightarrow \mathbb{R}$ . As a consequence of the Portmanteau theorem (see Theorems 1.2 and 2.1 in Billingsley 1968), to have weak convergence it suffices<sup>1</sup> that (3.2) holds for all bounded and Lipschitz continuous functions  $f : X \rightarrow \mathbb{R}$ . (We recall that  $f : X \rightarrow \mathbb{R}$  is Lipschitz continuous if there exists a constant  $L > 0$  such that  $|f(x) - f(y)| \leq Ld_X(x, y)$  for all  $x, y \in X$ . In this case, we say that  $f$  is  $L$ -Lipschitz continuous).

In case that  $X$  is a compact metric space, we have that weak convergence is metrizable with the Wasserstein distance

$$\sup_{f \in \text{Lip}_1(X)} \left\{ \int_X f d\mu - \int_X f d\nu \right\} = \inf_{\lambda} \int_{X \times X} d_X(x, x') \lambda(dx, dx') =: d_W(\mu, \nu), \tag{3.3}$$

for  $\mu, \nu \in \mathcal{P}$ , where the supremum ranges over the set  $\text{Lip}_1(X)$  of all 1-Lipschitz continuous functions on  $X$ , and where the infimum ranges over the set of all probability measures  $\lambda$  on  $X \times X$  with marginals  $\mu$  and  $\nu$  [see Theorems 8.3.2 and 8.10.45, and Section 8.10(viii) in Bogachev 2007]. With this metric, we have that  $(\mathcal{P}, d_W)$  is a compact metric space [Bogachev 2007, Theorem 8.9.3(i)]. In addition, if  $\{x_1, x_2, \dots\}$  is a countable dense subset of  $X$ , then the countable family of probability measures

$$\sum_{j=1}^k \beta_j \delta_{x_j}$$

for all  $k \geq 1$  and rational  $\beta_1, \dots, \beta_k \geq 0$  with  $\sum \beta_j = 1$  is dense in  $(\mathcal{P}, d_W)$ , where  $\delta_x$  denotes the Dirac probability measure supported on  $x$ ; see [Bogachev 2007, Theorem 8.9.4(ii)] or Bolley (2008).

<sup>1</sup> This is not the usual statement of the Portmanteau theorem. Observe, however, that the function constructed in Billingsley 1968, Theorem 1.2 is bounded and Lipschitz continuous, and then proceed as in the proof of Billingsley 1968, Theorem 2.1. Another reference for this result is Bogachev 2007, Remark 8.3.1.

**Main results.** We need one preliminary lemma before proving our main convergence result.

**Lemma 3.3** *Suppose that the game models  $\mathcal{G}$  and  $\{\mathcal{G}_n\}_{n \geq 1}$  satisfy Assumptions 2.2 and 2.9, and also that  $\mathcal{G}_n \rightarrow \mathcal{G}$ . Suppose that the sequence  $v_n \in \mathcal{B}_w(S_n)$ , for  $n \geq 1$ , converges pointwise to  $v \in \mathcal{B}_w(S)$ , and that*

$$\sup_{n \geq 1} \|v_n\|_w = m < \infty.$$

For fixed  $i \in S$ , assume also that  $\phi_n \in \overline{A}_n(i)$  and  $\psi_n \in \overline{B}_n(i)$ , for  $n \geq n(i)$ , are such that

$$\phi_n \xrightarrow{d} \phi \text{ and } \psi_n \xrightarrow{d} \psi \text{ as } n \rightarrow \infty$$

for some  $\phi \in \overline{A}(i)$  and  $\psi \in \overline{B}(i)$ . Under these conditions,

$$\lim_{n \rightarrow \infty} [r_n(i, \phi_n, \psi_n) + \sum_{j \in S_n} q_{ij}^n(\phi_n, \psi_n)v_n(j)] = r(i, \phi, \psi) + \sum_{j \in S} q_{ij}(\phi, \psi)v(j).$$

*Proof* Let us first analyze the term  $r_n(i, \phi_n, \psi_n)$ . By Lemma 3.2(ii), given  $\epsilon > 0$  there exists  $n_0 \geq n(i)$  such that  $n \geq n_0$  implies

$$|r_n(i, a, b) - r(i, a, b)| \leq \frac{\epsilon}{2} \text{ for all } (a, b) \in A_n(i) \times B_n(i).$$

In particular, we have

$$\begin{aligned} |r_n(i, \phi_n, \psi_n) - r(i, \phi_n, \psi_n)| &\leq \int_{A_n(i)} \int_{B_n(i)} |r_n(i, a, b) - r(i, a, b)| \psi_n(db) \phi_n(da) \\ &\leq \frac{\epsilon}{2} \end{aligned} \tag{3.4}$$

for  $n \geq n_0$ . On the other hand, since the function  $r(i, a, b)$  is bounded and continuous on  $A(i) \times B(i)$ , by Theorem 3.2 in Billingsley (1968) we have that  $r(i, \phi_n, \psi_n)$  converges to  $r(i, \phi, \psi)$  as  $n \rightarrow \infty$ . Consequently, there is some  $n_1 \geq n(i)$  such that  $n \geq n_1$  gives

$$|r(i, \phi_n, \psi_n) - r(i, \phi, \psi)| \leq \frac{\epsilon}{2}. \tag{3.5}$$

From (3.4) and (3.5), we have that  $|r_n(i, \phi_n, \psi_n) - r(i, \phi, \psi)| \leq \epsilon$  for  $n \geq n_0 \vee n_1$ . Therefore,

$$\lim_{n \rightarrow \infty} r_n(i, \phi_n, \psi_n) = r(i, \phi, \psi).$$

We proceed with the proof. As a consequence of Corollaries 2.6(i) and 2.10(i) we deduce that, given  $\epsilon > 0$ , there exists some  $k > i$  such that  $\sum_{j \geq k} q_{ij}(\phi, \psi)w(j) \leq \epsilon$  and such that, for all  $n \geq n(i)$ ,



$$\sum_{j \geq k, j \in S_n} q_{ij}^n(\phi_n, \psi_n)w(j) \leq \epsilon.$$

Therefore, since  $\|v_n\|_w \leq m$  implies  $\|v\|_w \leq m$ , we have

$$\left| \sum_{j \geq k} q_{ij}(\phi, \psi)v(j) \right| \leq m\epsilon$$

and, for all  $n \geq n(i)$ ,

$$\left| \sum_{j \geq k, j \in S_n} q_{ij}^n(\phi_n, \psi_n)v_n(j) \right| \leq m\epsilon.$$

Consequently, if  $n \geq n(i)$  is such that, in addition,  $\{0, 1, \dots, k - 1\} \subseteq S_n$ , we have

$$\begin{aligned} & \left| \sum_{j \in S_n} q_{ij}^n(\phi_n, \psi_n)v_n(j) - \sum_{j \in S} q_{ij}(\phi, \psi)v(j) \right| \\ & \leq \sum_{j=0}^{k-1} |q_{ij}^n(\phi_n, \psi_n)v_n(j) - q_{ij}(\phi, \psi)v(j)| + 2m\epsilon. \end{aligned}$$

The left-hand term of the last expression can be made arbitrarily small by choosing  $n$  large enough. Indeed, as made at the beginning of this proof, we can prove that

$$q_{ij}^n(\phi_n, \psi_n) \rightarrow q_{ij}(\phi, \psi)$$

which, together with the fact that  $v_n(j) \rightarrow v(j)$  for all  $j \in S$ , yields the stated result because, once  $i \in S$  and  $\epsilon > 0$  are given, the state  $k$  remains fixed and does not depend on  $n$ . The proof is now complete. □

Our next result needs to introduce some terminology. Given randomized stationary strategies  $(\pi_n^1, \pi_n^2) \in \Pi_n^{1,s} \times \Pi_n^{2,s}$  for the game model  $\mathcal{G}_n$ , for  $n \geq 1$ , we say that the randomized stationary strategies  $(\pi^1, \pi^2) \in \Pi_s^1 \times \Pi_s^2$  are a limit strategy of  $\{(\pi_n^1, \pi_n^2)\}_{n \geq 1}$  if there exists a subsequence  $\{n'\}$  such that

$$\pi_{n'}^1(\cdot|i) \xrightarrow{d} \pi^1(\cdot|i) \quad \text{and} \quad \pi_{n'}^2(\cdot|i) \xrightarrow{d} \pi^2(\cdot|i)$$

for all  $i \in S$ . Clearly, every such sequence  $\{(\pi_n^1, \pi_n^2)\}$  indeed has a limit strategy because  $\pi_n^1 \in \bar{A}(i)$  and  $\pi_n^2 \in \bar{B}(i)$ , which are compact metric spaces with the Wasserstein metric. Next, we state our main convergence result.

**Theorem 3.4** *Suppose that the game models  $\mathcal{G}$  and  $\{\mathcal{G}_n\}_{n \geq 1}$  satisfy Assumptions 2.2 and 2.9. If  $\mathcal{G}_n \rightarrow \mathcal{G}$  then the following statements are satisfied.*

- (i) For all  $i \in S$ ,  $\lim_{n \rightarrow \infty} V_n^\alpha(i) = V^\alpha(i)$ .

(ii) If  $(\pi_n^1, \pi_n^2)$  is a pair of optimal randomized stationary strategies for  $\mathcal{G}_n$ , then any limit strategy  $(\pi^1, \pi^2) \in \Pi_S^1 \times \Pi_S^2$  is a pair of optimal randomized stationary strategies for  $\mathcal{G}$ .

*Proof* (i) Recall that the sequence  $\{V_n^\alpha\}$  of the values of the games  $\mathcal{G}_n$  verifies

$$|V_n^\alpha(i)| \leq \mathfrak{M}w(i) \quad \text{for all } n \geq 1 \text{ and } i \in S_n;$$

see (2.10). Therefore, using a diagonal argument, we deduce the existence of  $u \in \mathcal{B}_w(S)$  and a subsequence  $\{k_n\}$  such that

$$\lim_{n \rightarrow \infty} V_{k_n}^\alpha(i) = u(i) \quad \text{for all } i \in S.$$

Fix  $i \in S$  and, for  $n$  such that  $k_n \geq n(i)$ , consider the function on  $\bar{A}_{k_n}(i) \times \bar{B}_{k_n}(i)$

$$(\phi, \psi) \mapsto r_{k_n}(i, \phi, \psi) + \sum_{j \in S_{k_n}} q_{ij}^{k_n}(\phi, \psi) V_{k_n}^\alpha(j).$$

This function is continuous as a consequence of Corollary 2.10 and Theorem 3.2 in Billingsley (1968). Therefore,

$$\phi \mapsto \inf_{\psi \in \bar{B}_{k_n}(i)} \left\{ r_{k_n}(i, \phi, \psi) + \sum_{j \in S_{k_n}} q_{ij}^{k_n}(\phi, \psi) V_{k_n}^\alpha(j) \right\}$$

is upper semi-continuous on the compact set  $\bar{A}_{k_n}(i)$  and, hence, it has a maximum which is reached at some  $\phi_n \in \bar{A}_{k_n}(i)$ . There exists a further subsequence  $\{k_{n'}\}$  such that  $\phi_{n'} \xrightarrow{d} \phi_0$  for some  $\phi_0 \in \bar{A}(i)$ . Without loss of generality, and to simplify the notation, we will suppose that the whole sequence  $\{\phi_n\}$  converges to  $\phi_0$ .

Fix now arbitrary  $\psi \in \bar{B}(i)$ . For each  $n$  there exist

$$x_1, \dots, x_t \in B(i) \quad \text{and} \quad \beta_1, \dots, \beta_t \in [0, 1]$$

with  $\sum \beta_j = 1$  such that  $d_W(\psi, \hat{\psi}_n) \leq 1/n$ , with  $\hat{\psi}_n = \sum \beta_j \delta_{x_j}$ . Let  $y_j \in B_{k_n}(i)$  be such that  $d_B(y_j, x_j) = \min_{y \in B_{k_n}(i)} d_B(y, x_j)$  for each  $j = 1, \dots, t$ , and define

$$\tilde{\psi}_n = \sum_{j=1}^t \beta_j \delta_{y_j} \in \bar{B}_{k_n}(i).$$

If  $f$  is a bounded  $L$ -Lipschitz continuous function on  $B(i)$  then we have

$$\left| \int f d\tilde{\psi}_n - \int f d\psi \right| \leq \left| \int f d\tilde{\psi}_n - \int f d\hat{\psi}_n \right| + \left| \int f d\hat{\psi}_n - \int f d\psi \right|.$$

We note that

$$\left| \int f d\tilde{\psi}_n - \int f d\hat{\psi}_n \right| = \left| \sum_{j=1}^t \beta_j [f(y_j) - f(x_j)] \right| \leq L \sum_{j=1}^t \beta_j d_B(y_j, x_j) \leq L d_H(B_{k_n}(i), B(i)),$$

which converges to 0 as  $n \rightarrow \infty$ . On the other hand, since  $\hat{\psi}_n \xrightarrow{d} \psi$  we have  $\int f d\hat{\psi}_n \rightarrow \int f d\psi$ . So, we have shown that

$$\lim_{n \rightarrow \infty} \int f d\tilde{\psi}_n = \int f d\psi$$

for all bounded and Lipschitz-continuous functions on  $B(i)$ . This implies that  $\tilde{\psi}_n \xrightarrow{d} \psi$ . Summarizing, given arbitrary  $\psi \in \bar{B}(i)$  we have constructed  $\tilde{\psi}_n \in \bar{B}_{k_n}(i)$  such that  $\{\tilde{\psi}_n\}$  converges weakly to  $\psi$ .

By Theorem 2.11(ii), the value  $V_{k_n}^\alpha$  of the game  $\mathcal{G}_{k_n}$  verifies

$$\begin{aligned} \alpha V_{k_n}^\alpha(i) &= \sup_{\phi \in \bar{A}_{k_n}(i)} \inf_{\psi \in \bar{B}_{k_n}(i)} \left\{ r_{k_n}(i, \phi, \psi) + \sum_{j \in S_{k_n}} q_{ij}^{k_n}(\phi, \psi) V_{k_n}^\alpha(j) \right\} \\ &= \inf_{\psi \in \bar{B}_{k_n}(i)} \left\{ r_{k_n}(i, \phi_n, \psi) + \sum_{j \in S_{k_n}} q_{ij}^{k_n}(\phi_n, \psi) V_{k_n}^\alpha(j) \right\} \\ &\leq r_{k_n}(i, \phi_n, \tilde{\psi}_n) + \sum_{j \in S_{k_n}} q_{ij}^{k_n}(\phi_n, \tilde{\psi}_n) V_{k_n}^\alpha(j). \end{aligned}$$

Taking the limit as  $n \rightarrow \infty$  and recalling Lemma 3.3, we obtain

$$\alpha u(i) \leq r(i, \phi_0, \psi) + \sum_{j \in S} q_{ij}(\phi_0, \psi) u(j).$$

But  $\psi \in \bar{B}(i)$  being arbitrary, we conclude that

$$\alpha u(i) \leq \inf_{\psi \in \bar{B}(i)} \left\{ r(i, \phi_0, \psi) + \sum_{j \in S} q_{ij}(\phi_0, \psi) u(j) \right\},$$

and so,

$$\alpha u(i) \leq \sup_{\phi \in \bar{A}(i)} \inf_{\psi \in \bar{B}(i)} \left\{ r(i, \phi, \psi) + \sum_{j \in S} q_{ij}(\phi, \psi) u(j) \right\}.$$

Arguing similarly, we can show that

$$\alpha u(i) \geq \inf_{\psi \in \bar{B}(i)} \sup_{\phi \in \bar{A}(i)} \left\{ r(i, \phi, \psi) + \sum_{j \in S} q_{ij}(\phi, \psi) u(j) \right\}.$$

Combining these two inequalities, we conclude that

$$\begin{aligned} \alpha u(i) &= \sup_{\phi \in \bar{A}(i)} \inf_{\psi \in \bar{B}(i)} \left\{ r(i, \phi, \psi) + \sum_{j \in S} q_{ij}(\phi, \psi)u(j) \right\} \\ &= \inf_{\psi \in \bar{B}(i)} \sup_{\phi \in \bar{A}(i)} \left\{ r(i, \phi, \psi) + \sum_{j \in S} q_{ij}(\phi, \psi)u(j) \right\} \end{aligned}$$

for each  $i \in S$ . By Theorem 2.7(ii), this implies that  $u$  equals  $V^\alpha$ , the value of the game  $\mathcal{G}$ .

So far, we have shown that if  $u$  is any limit point of  $\{V_n^\alpha\}$  then, necessarily,  $u = V^\alpha$ . But this implies that  $\lim_{n \rightarrow \infty} V_n^\alpha(i) = V^\alpha(i)$  for all  $i \in S$ . The proof of (i) is now complete.

- (ii) Suppose that  $(\pi^1, \pi^2) \in \Pi_S^1 \times \Pi_S^2$  is a limit strategy through the subsequence  $\{n'\}$ . Fix  $i \in S$  and write

$$\phi_n^* = \pi_n^1(\cdot|i), \quad \psi_n^* = \pi_n^2(\cdot|i), \quad \phi^* = \pi^1(\cdot|i), \quad \psi^* = \pi^2(\cdot|i)$$

for  $n \geq n(i)$ . Then, we have

$$\phi_{n'}^* \xrightarrow{d} \phi^* \quad \text{and} \quad \psi_{n'}^* \xrightarrow{d} \psi^*.$$

For  $n \geq n(i)$ , we know that  $\phi_{n'}^*$  and  $\psi_{n'}^*$  attain the supremum and the infimum in the Shapley equation for  $\mathcal{G}_{n'}$  for the state  $i$ ; recall Theorem 2.11(iii). Therefore,

$$\begin{aligned} \alpha V_{n'}^\alpha(i) &= \sup_{\phi \in \bar{A}_{n'}(i)} \inf_{\psi \in \bar{B}_{n'}(i)} \left\{ r_{n'}(i, \phi, \psi) + \sum_{j \in S_{n'}} q_{ij}^{n'}(\phi, \psi)V_{n'}^\alpha(j) \right\} \\ &= \inf_{\psi \in \bar{B}_{n'}(i)} \left\{ r_{n'}(i, \phi_{n'}^*, \psi) + \sum_{j \in S_{n'}} q_{ij}^{n'}(\phi_{n'}^*, \psi)V_{n'}^\alpha(j) \right\}. \end{aligned} \tag{3.6}$$

Proceeding as in the proof of part (i), we can show that for every  $\psi \in \bar{B}(i)$  there exists a sequence  $\psi_{n'} \in \bar{B}_{n'}(i)$  such that  $\psi_{n'} \xrightarrow{d} \psi$ , and so, by (3.6),

$$\alpha V_{n'}^\alpha(i) \leq r_{n'}(i, \phi_{n'}^*, \psi_{n'}) + \sum_{j \in S_{n'}} q_{ij}^{n'}(\phi_{n'}^*, \psi_{n'})V_{n'}^\alpha(j).$$

Taking the limit as  $n' \rightarrow \infty$  and recalling that  $V_n^\alpha$  converges pointwise to  $V^\alpha$  (part (i) of this theorem) give

$$\alpha V^\alpha(i) \leq r(i, \phi^*, \psi) + \sum_{j \in S} q_{ij}(\phi^*, \psi)V^\alpha(j).$$

Since  $\psi \in \overline{B}(i)$  is arbitrary, we have

$$\alpha V^\alpha(i) \leq \inf_{\psi \in \overline{B}(i)} \left\{ r(i, \phi^*, \psi) + \sum_{j \in S} q_{ij}(\phi^*, \psi) V^\alpha(j) \right\}.$$

But from the Shapley equation for  $\mathcal{G}$  we know that

$$\alpha V^\alpha(i) = \sup_{\phi \in \overline{A}(i)} \inf_{\psi \in \overline{B}(i)} \left\{ r(i, \phi, \psi) + \sum_{j \in S} q_{ij}(\phi, \psi) V^\alpha(j) \right\}.$$

Hence,  $\phi^*$  attains the supremum in the Shapley equation for  $i \in S$ .

Similarly,  $\psi^*$  attains the infimum in the Shapley equation for  $\mathcal{G}$  and  $i \in S$  and, by Theorem 2.7(iii), this implies that  $(\pi^1, \pi^2) \in \Pi_s^1 \times \Pi_s^2$  is indeed a pair of optimal strategies for  $\mathcal{G}$ . □

#### 4 Finite state and actions approximations

Given a game model  $\mathcal{G}$  satisfying Assumption 2.2, we show how to construct a sequence of game models  $\{\mathcal{G}_n\}_{n \geq 1}$  for which Assumption 2.9 holds. For each  $n \geq 1$ , the elements of the game model  $\mathcal{G}_n$  are the following.

- The state space is  $S_n = \{0, 1, \dots, n\}$ .
- For  $i \in S_n$ , let  $A_n(i)$  and  $B_n(i)$  be finite subsets of  $A(i)$  and  $B(i)$ , respectively, such that

$$d_H(A_n(i), A(i)) \rightarrow 0 \quad \text{and} \quad d_H(B_n(i), B(i)) \rightarrow 0$$

as  $n \rightarrow \infty$ .

- Given  $i \in S_n$  and  $0 \leq j < n$ , define  $q_{ij}^n(a, b) = q_{ij}(a, b)$ , and let

$$q_{in}^n(a, b) = \sum_{j \geq n} q_{ij}(a, b)$$

for  $(a, b) \in A_n(i) \times B_n(i)$ .

- The reward/cost rate is  $r_n(i, a, b) = r(i, a, b)$  for  $(a, b) \in A_n(i) \times B_n(i)$ .

We note that construction of  $A_n(i)$  and  $B_n(i)$  with the property of convergence in the Hausdorff metric is indeed possible. For instance, for each  $n \geq 1$ , consider the open cover of  $A(i)$  given by the open balls centered in  $a \in A(i)$  with radius  $1/n$ , and let  $A_n(i)$  be the centers of a finite subcover. Then,  $d_H(A(i), A_n(i)) \leq 1/n$ .

**Theorem 4.1** *If the game model  $\mathcal{G}$  satisfies Assumption 2.2, then the sequence  $\{\mathcal{G}_n\}_{n \geq 1}$  defined above satisfies Assumption 2.9 and, moreover,  $\mathcal{G}_n \rightarrow \mathcal{G}$ .*

*Proof* First of all, we observe that the transition rates of  $\mathcal{G}_n$  are conservative:

$$\sum_{j \in S_n} q_{ij}^n(a, b) = \sum_{j \in S} q_{ij}(a, b) = 0 \quad \text{for all } (i, a, b) \in \mathbb{K}_n,$$

and stable. Indeed,  $-q_{ii}^n(a, b) = -q_{ii}(a, b) \leq w(i) \quad \text{for } i < n$  and

$$-q_{nn}^n(a, b) = -\sum_{j \geq n} q_{nj}(a, b) \leq -q_{nn}(a, b) \leq w(n).$$

Concerning Assumption 2.9(i), observe that for all  $(i, a, b) \in \mathbb{K}_n$

$$\begin{aligned} \sum_{j \in S_n} q_{ij}^n(a, b)w(j) &= \sum_{j \in S_n} q_{ij}(a, b)w(j) + \sum_{j > n} q_{ij}(a, b)w(n) \\ &\leq \sum_{j \in S_n} q_{ij}(a, b)w(j) + \sum_{j > n} q_{ij}(a, b)w(j) \\ &= \sum_{j \in S} q_{ij}(a, b)w(j) \leq c_1w(i) + d_1, \end{aligned}$$

where we make use of the monotonicity of  $w$ . The fact that  $q_n(i) \leq w(i)$  has been established along with the stability of the transition rates of  $\mathcal{G}_n$ . So, Assumption 2.9(i) indeed holds.

Clearly, Assumptions 2.9(ii)–(iii) are also satisfied, while Assumption 2.9(iv) is proved similarly to Assumption 2.9(i).

It remains to check that  $\mathcal{G}_n \rightarrow \mathcal{G}$ . Items (a) and (b) in Definition 3.1 hold by construction of  $\mathcal{G}_n$ . Finally, given  $i, j \in S$ , if  $(a_n, b_n) \in A_n(i) \times B_n(i)$  are such that  $a_n \rightarrow a \in A(i)$  and  $b_n \rightarrow b \in B(i)$  then

$$r_n(i, a_n, b_n) = r(i, a_n, b_n) \quad \text{and} \quad q_{ij}^n(a_n, b_n) = q_{ij}(a_n, b_n)$$

for  $n > i \vee j$ , and so Definitions 3.1(c)–(d) hold by continuity of the transition and reward/cost rates of  $\mathcal{G}$ . □

As a consequence of Theorem 3.4, the value functions of the finite state and actions games  $\mathcal{G}_n$  converge to the value function of  $\mathcal{G}$ , and any limit strategy of optimal stationary strategies for  $\mathcal{G}_n$  is optimal for  $\mathcal{G}$ . Next, we address the issue of the rate of convergence of  $V_n^\alpha(i)$  to  $V^\alpha(i)$ . To establish such convergence rates, we need to strengthen our hypotheses.

**Assumption 4.2** (i) For each  $i, j \in S$ , the functions  $(a, b) \mapsto r(i, a, b)$  and  $(a, b) \mapsto q_{ij}(a, b)$  are  $L_{r_i}$ - and  $L_{q_{ij}}$ -Lipschitz continuous on  $A(i) \times B(i)$ , i.e.,

$$\begin{aligned} |r(i, a, b) - r(i, a', b')| &\leq L_{r_i} (d_A(a, a') + d_B(b, b')) \\ |q_{ij}(a, b) - q_{ij}(a', b')| &\leq L_{q_{ij}} (d_A(a, a') + d_B(b, b')) \end{aligned}$$

for all  $a, a' \in A(i)$  and  $b, b' \in B(i)$ , and some  $L_{r_i} > 0$  and  $L_{q_{ij}} > 0$ .

- (ii) With  $w$  the Lyapunov function in Assumption 2.2, there exist constants  $\delta > 2$ ,  $c_\delta < \alpha$ , and  $d_\delta \geq 0$  with

$$\sum_{j \in S} q_{ij}(a, b)w^\delta(j) \leq c_\delta w^\delta(i) + d_\delta \text{ for all } (i, a, b) \in \mathbb{K}. \tag{4.1}$$

Before establishing our main result, we need two preliminary lemmas.

**Lemma 4.3** *Suppose that the function  $h : S \rightarrow [0, \infty)$  satisfies  $q(i) \leq h(i)$  for all  $i \in S$ . If there exists a power  $\gamma > 0$  and a constant  $c_\gamma \geq 0$  such that*

$$\sum_{j \in S} q_{ij}(a, b)h^\gamma(j) \leq c_\gamma h^\gamma(i) \text{ for all } (i, a, b) \in \mathbb{K}, \tag{4.2}$$

then for every power  $0 < \gamma' < \gamma$

$$\sum_{j \in S} q_{ij}(a, b)h^{\gamma'}(j) \leq c_\gamma h^{\gamma'}(i) \text{ for all } (i, a, b) \in \mathbb{K}.$$

*Proof* Fix  $(i, a, b) \in \mathbb{K}$  and  $\eta > 0$ . Rewrite (4.2) as

$$\frac{1}{h(i) + \eta} \sum_{j \neq i} q_{ij}(a, b)h^\gamma(j) + \left(\frac{q_{ii}(a, b)}{h(i) + \eta} + 1\right)h^\gamma(i) \leq \left(\frac{c_\gamma}{h(i) + \eta} + 1\right)h^\gamma(i). \tag{4.3}$$

Let

$$p_i = \frac{q_{ii}(a, b)}{h(i) + \eta} + 1 \text{ and } p_j = \frac{q_{ij}(a, b)}{h(i) + \eta} \text{ for } j \neq i.$$

These coefficients are nonnegative and  $\sum_{j \in S} p_j = 1$ . Therefore, (4.3) is equivalent to

$$\sum_{j \in S} p_j h^\gamma(j) \leq \left(\frac{c_\gamma}{h(i) + \eta} + 1\right)h^\gamma(i).$$

Using Jensen’s inequality for the concave function  $x \mapsto x^{\gamma'/\gamma}$  yields

$$\sum_{j \in S} p_j h^{\gamma'}(j) \leq \left(\frac{c_\gamma}{h(i) + \eta} + 1\right)^{\gamma'/\gamma} h^{\gamma'}(i)$$

or, equivalently,

$$\sum_{j \in S} q_{ij}(a, b)h^{\gamma'}(j) \leq h^{\gamma'}(i) \left( \left(\frac{c_\gamma}{h(i) + \eta} + 1\right)^{\gamma'/\gamma} - 1 \right) (h(i) + \eta).$$

Since  $0 < \gamma'/\gamma < 1$ , we have

$$\left(\frac{c_\gamma}{h(i) + \eta} + 1\right)^{\gamma'/\gamma} - 1 \leq \frac{c_\gamma}{h(i) + \eta},$$

and so

$$\sum_{j \in S} q_{ij}(a, b)h^{\gamma'}(j) \leq c_\gamma h^{\gamma'}(i),$$

which completes the proof. □

Consequently, if there exists a power  $\gamma > 0$  such that the Lyapunov function  $w$  verifies

$$\sum_{j \in S} q_{ij}(a, b)w^\gamma(j) \leq c_\gamma w^\gamma(i) + d_\gamma \quad \text{for all } (i, a, b) \in \mathbb{K}$$

and some constants  $c_\gamma \in \mathbb{R}$  and  $d_\gamma \geq 0$ , then for every  $0 < \gamma' < \gamma$

$$\sum_{j \in S} q_{ij}(a, b)w^{\gamma'}(j) \leq (|c_\gamma| + d_\gamma)w^{\gamma'}(i) \quad \text{for all } (i, a, b) \in \mathbb{K}$$

(indeed, just note that  $\sum q_{ij}(a, b)w^\gamma(j) \leq (|c_\gamma| + d_\gamma)w^\gamma(i)$  and use Lemma 4.3). In particular, if Assumption 4.2(ii) holds then necessarily Assumption 2.2(iv) is satisfied and, moreover,

$$\sum_{j \in S} q_{ij}(a, b)w^{\delta-1}(j) \leq (|c_\delta| + d_\delta)w^{\delta-1}(i) \quad \text{for all } (i, a, b) \in \mathbb{K}. \tag{4.4}$$

**Lemma 4.4** *Consider a fixed  $n \geq 1$  and suppose that the game model  $\mathcal{G}_n$  satisfies Assumption 2.9. Suppose that there exists a function  $u \in \mathcal{B}_w(S_n)$  such that, for all  $i \in S_n$ ,*

$$\left| \alpha u(i) - \sup_{\phi \in \bar{A}_n(i)} \inf_{\psi \in \bar{B}_n(i)} \{r_n(i, \phi, \psi) + \sum_{j \in S_n} q_{ij}^n(\phi, \psi)u(j)\} \right| \leq h(i)$$

for some  $h(i) \geq 0$ . Assume, in addition, that there exist constants  $c_h < \alpha$  and  $d_h \geq 0$  such that

$$\sum_{j \in S_n} q_{ij}^n(a, b)h(j) \leq c_h h(i) + d_h \quad \text{for all } (i, a, b) \in \mathbb{K}_n.$$

Under these conditions,

$$|V_n^\alpha(i) - u(i)| \leq \frac{h(i)}{\alpha - c_h} + \frac{d_h}{\alpha(\alpha - c_h)} \quad \text{for each } i \in S_n.$$



*Proof* First of all, we note that for every  $(\pi^1, \pi^2) \in \Pi_n^1 \times \Pi_n^2, t \geq 0$ , and  $i \in S_n$  we have

$$E_n^{i, \pi^1, \pi^2} [h(x(t))] \leq e^{c_h t} h(i) + \frac{d_h}{c_h} (e^{c_h t} - 1) \quad \text{if } c_h \neq 0$$

or  $E_n^{i, \pi^1, \pi^2} [h(x(t))] \leq h(i) + d_h t$  when  $c_h = 0$  (the proof of these inequalities is similar to that of (2.4)). Therefore, in either case,

$$E_n^{i, \pi^1, \pi^2} \left[ \int_0^\infty e^{-\alpha t} h(x(t)) \right] \leq \frac{h(i)}{\alpha - c_h} + \frac{d_h}{\alpha(\alpha - c_h)} \quad \text{for all } i \in S_n. \tag{4.5}$$

For every  $i \in S_n$ , there exists some  $\phi \in \bar{A}_n(i)$  such that for all  $\psi \in \bar{B}_n(i)$

$$\alpha u(i) - r_n(i, \phi, \psi) - \sum_{j \in S_n} q_{ij}^n(\phi, \psi) u(j) \leq h(i).$$

Consequently, there exists a stationary policy  $\pi^1 \in \Pi_n^{1,s}$  such that for every stationary  $\pi^2 \in \Pi_n^{2,s}$

$$\alpha u(i) - r_n(i, \pi^1, \pi^2) - \sum_{j \in S_n} q_{ij}^n(\pi^1, \pi^2) u(j) \leq h(i) \quad \text{for all } i \in S_n.$$

Using Dynkin’s formula gives, for every  $i \in S_n$  and  $t \geq 0$ ,

$$\begin{aligned} & E_n^{i, \pi^1, \pi^2} [e^{-\alpha t} u(x(t))] - u(i) \\ &= E_n^{i, \pi^1, \pi^2} \left[ \int_0^t e^{-\alpha s} [-\alpha u(x(s)) + \sum_{j \in S_n} q_{x(s)j}^n(\pi^1, \pi^2) u(j)] ds \right] \\ &\geq -E_n^{i, \pi^1, \pi^2} \left[ \int_0^t e^{-\alpha s} [r_n(x(s), \pi^1, \pi^2) + h(x(s))] ds \right]. \end{aligned}$$

Now we let  $t \rightarrow \infty$  in this inequality. Recalling (2.4) and Remark 2.4 for the game model  $\mathcal{G}_n$ , and using dominated and monotone convergence, we obtain

$$\begin{aligned} u(i) &\leq V_n^\alpha(i, \pi^1, \pi^2) + E_n^{i, \pi^1, \pi^2} \left[ \int_0^\infty e^{-\alpha s} h(x(s)) ds \right] \\ &\leq V_n^\alpha(i, \pi^1, \pi^2) + \frac{h(i)}{\alpha - c_h} + \frac{d_h}{\alpha(\alpha - c_h)}, \end{aligned}$$

where we have used (4.5). Since this inequality holds for some  $\pi^1$  and all  $\pi^2$ , we obtain

$$\begin{aligned}
 u(i) &\leq \sup_{\pi^1 \in \Pi_n^{1,s}} \inf_{\pi^2 \in \Pi_n^{2,s}} \{V_n^\alpha(i, \pi^1, \pi^2)\} + \frac{h(i)}{\alpha - c_h} + \frac{d_h}{\alpha(\alpha - c_h)} \\
 &= V_n^\alpha(i) + \frac{h(i)}{\alpha - c_h} + \frac{d_h}{\alpha(\alpha - c_h)}
 \end{aligned}
 \tag{4.6}$$

for all  $i \in S_n$  (use Remark 2.8 for the game model  $\mathcal{G}_n$ ).

Observe now that, the sets  $A_n(i)$  and  $B_n(i)$  being finite, we have

$$\begin{aligned}
 &\inf_{\psi \in B_n(i)} \sup_{\phi \in \bar{A}_n(i)} \{r_n(i, \phi, \psi) + \sum_{j \in S_n} q_{ij}^n(\phi, \psi)u(j)\} \\
 &= \sup_{\phi \in \bar{A}_n(i)} \inf_{\psi \in B_n(i)} \{r_n(i, \phi, \psi) + \sum_{j \in S_n} q_{ij}^n(\phi, \psi)u(j)\};
 \end{aligned}$$

see Theorem 1 in [Frenk et al. \(2004\)](#). So, using a symmetric argument with the inequality

$$-h(i) \leq \alpha u(i) - \inf_{\psi \in B_n(i)} \sup_{\phi \in \bar{A}_n(i)} \{r_n(i, \phi, \psi) + \sum_{j \in S_n} q_{ij}^n(\phi, \psi)u(j)\}$$

gives the existence of  $\pi^2 \in \Pi_n^{2,s}$  such that for all  $\pi^1 \in \Pi_n^{1,s}$

$$V_n^\alpha(i, \pi^1, \pi^2) \leq u(i) + \frac{h(i)}{\alpha - c_h} + \frac{d_h}{\alpha(\alpha - c_h)} \text{ for all } i \in S_n,$$

and, therefore,

$$V_n^\alpha(i) \leq u(i) + \frac{h(i)}{\alpha - c_h} + \frac{d_h}{\alpha(\alpha - c_h)} \text{ for all } i \in S_n.$$

Together with (4.6), this proves the stated result. □

Finally, we state our main result on the convergence rates to the value of the game.

**Theorem 4.5** *Suppose that the game model  $\mathcal{G}$  satisfies Assumptions 2.2 and 4.2.*

*Let  $\{\mathcal{G}_n\}_{n \geq 1}$  be the sequence of finite state and actions truncations of  $\mathcal{G}$  constructed at the beginning of Sect. 4, and suppose that the action sets for  $\mathcal{G}_n$  are chosen in such a way that, for all  $n \geq 1$  and  $i \in S_n$ , and for some constant  $D > 0$ ,*

$$d_H(A_n(i), A(i)) \vee d_H(B_n(i), B(i)) \leq \frac{Dw^\delta(i)}{w^{\delta-2}(n+1)(L_{r_i} + 2\mathfrak{M}w(n) \sum_{j=0}^{n-1} L_{q_{ij}})}.$$

*Under these conditions, there exists a constant  $\epsilon > 0$  such that, for every  $n \geq 1$  and  $i \in S_n$ ,*

$$|V_n^\alpha(i) - V^\alpha(i)| \leq \epsilon \frac{w^\delta(i)}{w^{\delta-2}(n+1)}.$$

*Proof* Fix  $n \geq 1$  and  $i \in S_n$ . We have

$$\begin{aligned} \alpha V^\alpha(i) &= \sup_{\phi \in \bar{A}(i)} \inf_{\psi \in \bar{B}(i)} \left\{ r(i, \phi, \psi) + \sum_{j \in S} q_{ij}(\phi, \psi) V^\alpha(j) \right\} \\ &\leq \sup_{\phi \in \bar{A}(i)} \inf_{\psi \in \bar{B}_n(i)} \left\{ r(i, \phi, \psi) + \sum_{j \in S} q_{ij}(\phi, \psi) V^\alpha(j) \right\}. \end{aligned} \tag{4.7}$$

Note that for every  $(\phi, \psi) \in \bar{A}(i) \times \bar{B}(i)$

$$\begin{aligned} \sum_{j \in S} q_{ij}(\phi, \psi) V^\alpha(j) &= \sum_{j=0}^{n-1} q_{ij}(\phi, \psi) (V^\alpha(j) - V^\alpha(n)) \\ &\quad + \sum_{j>n} q_{ij}(\phi, \psi) (V^\alpha(j) - V^\alpha(n)), \end{aligned}$$

and recalling that  $\|V^\alpha\|_w \leq \mathfrak{M}$ ,

$$\left| \sum_{j>n} q_{ij}(\phi, \psi) (V^\alpha(j) - V^\alpha(n)) \right| \leq 2\mathfrak{M} \sum_{j>n} q_{ij}(\phi, \psi) w(j).$$

Observe now that proceeding as in the proof of Corollary 2.6(i) and recalling (4.4), we can show that

$$\begin{aligned} \sum_{j>n} q_{ij}(\phi, \psi) w(j) &\leq \frac{1}{w^{\delta-2}(n+1)} \cdot \left( (|c_\delta| + d_\delta) w^{\delta-1}(i) + q(i) w^{\delta-1}(i) \right) \\ &\leq (|c_\delta| + d_\delta + 1) \frac{w^\delta(i)}{w^{\delta-2}(n+1)}. \end{aligned} \tag{4.8}$$

Therefore, combining (4.7) and (4.8), we obtain

$$\alpha V^\alpha(i) \leq \sup_{\phi \in \bar{A}(i)} \inf_{\psi \in \bar{B}_n(i)} \left\{ r(i, \phi, \psi) + \sum_{j=0}^{n-1} q_{ij}(\phi, \psi) (V^\alpha(j) - V^\alpha(n)) \right\} + \bar{C} \frac{w^\delta(i)}{w^{\delta-2}(n+1)},$$

with

$$\bar{C} = 2\mathfrak{M}(|c_\delta| + d_\delta + 1).$$

By upper semicontinuity, the above supremum is indeed attained. Consequently, there exists  $\phi \in \bar{A}(i)$  such that, for every  $\psi \in \bar{B}_n(i)$ , we have

$$\alpha V^\alpha(i) \leq r(i, \phi, \psi) + \sum_{j=0}^{n-1} q_{ij}(\phi, \psi) (V^\alpha(j) - V^\alpha(n)) + \bar{C} \frac{w^\delta(i)}{w^{\delta-2}(n+1)}. \tag{4.9}$$

Given arbitrary  $\epsilon > 0$ , there exists a finite set  $\{x_1, \dots, x_k\} \subseteq A(i)$  and  $\beta_1, \dots, \beta_k \geq 0$ , with  $\beta_1 + \dots + \beta_k = 1$ , such that

$$d_W(\phi, \sum_{j=1}^k \beta_j \delta_{x_j}) \leq \epsilon.$$

For every  $x_j$ , let  $\hat{x}_j \in A_n(i)$  be such that

$$d_A(x_j, \hat{x}_j) = \min_{y \in A_n(i)} d_A(x_j, y) \leq d_H(A(i), A_n(i)).$$

It is easy to see (recall (3.3)) that

$$d_W\left(\sum_{j=1}^k \beta_j \delta_{x_j}, \sum_{j=1}^k \beta_j \delta_{\hat{x}_j}\right) \leq \sum_{j=1}^k \beta_j d_A(x_j, \hat{x}_j) \leq d_H(A(i), A_n(i)),$$

and so, letting  $\hat{\phi} = \sum_{j=1}^k \beta_j \delta_{\hat{x}_j} \in \bar{A}_n(i)$ ,

$$d_W(\phi, \hat{\phi}) \leq \epsilon + d_H(A(i), A_n(i)).$$

Summarizing, for  $\phi \in \bar{A}(i)$  we have found a probability measure in  $\hat{\phi} \in \bar{A}_n(i)$  which is “close” to  $\phi$  in the Wasserstein metric. By the Lipschitz continuity Assumption 4.2, observe that the function on  $A(i) \times B(i)$  given by

$$(a, b) \mapsto r(i, a, b) + \sum_{j=0}^{n-1} q_{ij}(a, b)(V^\alpha(j) - V^\alpha(n))$$

is Lipschitz continuous, with Lipschitz constant  $L_{r_i} + 2\mathfrak{M}w(n) \sum_{j=0}^{n-1} L_{q_{ij}}$ . Consequently, the same applies to

$$a \mapsto \int_{B_n(i)} \left[ r(i, a, b) + \sum_{j=0}^{n-1} q_{ij}(a, b)(V^\alpha(j) - V^\alpha(n)) \right] \psi(db).$$

Use now (3.3) to derive that

$$\begin{aligned} & \left| r(i, \phi, \psi) + \sum_{j=0}^{n-1} q_{ij}(\phi, \psi)(V^\alpha(j) - V^\alpha(n)) - r(i, \hat{\phi}, \psi) \right. \\ & \quad \left. + \sum_{j=0}^{n-1} q_{ij}(\hat{\phi}, \psi)(V^\alpha(j) - V^\alpha(n)) \right| \end{aligned}$$

$$\begin{aligned} &\leq \left( L_{r_i} + 2\mathfrak{M}w(n) \sum_{j=0}^{n-1} L_{q_{ij}} \right) \cdot d_W(\phi, \hat{\phi}) \\ &\leq \left( L_{r_i} + 2\mathfrak{M}w(n) \sum_{j=0}^{n-1} L_{q_{ij}} \right) \cdot (\epsilon + d_H(A(i), A_n(i))). \end{aligned}$$

Therefore, recalling (4.9), this yields that  $\alpha V^\alpha(i)$  is less than or equal to

$$\begin{aligned} &r(i, \hat{\phi}, \psi) + \sum_{j=0}^{n-1} q_{ij}(\hat{\phi}, \psi)(V^\alpha(j) - V^\alpha(n)) \\ &+ \bar{C} \frac{w^\delta(i)}{w^{\delta-2}(n+1)} + \left( L_{r_i} + 2\mathfrak{M}w(n) \sum_{j=0}^{n-1} L_{q_{ij}} \right) \cdot (\epsilon + d_H(A(i), A_n(i))). \end{aligned}$$

Since this holds for all  $\psi \in \bar{B}_n(i)$  and the particular  $\hat{\phi} \in \bar{A}_n(i)$  constructed above, we deduce that

$$\begin{aligned} \alpha V^\alpha(i) &\leq \sup_{\phi \in \bar{A}_n(i)} \inf_{\psi \in \bar{B}_n(i)} \left\{ r(i, \phi, \psi) + \sum_{j=0}^{n-1} q_{ij}(\phi, \psi)(V^\alpha(j) - V^\alpha(n)) \right\} \\ &+ \bar{C} \frac{w^\delta(i)}{w^{\delta-2}(n+1)} \\ &+ \left( L_{r_i} + 2\mathfrak{M}w(n) \sum_{j=0}^{n-1} L_{q_{ij}} \right) \cdot (\epsilon + d_H(A(i), A_n(i))). \end{aligned}$$

But  $\epsilon > 0$  being arbitrary and recalling our hypothesis on  $d_H(A(i), A_n(i))$ , we derive that

$$\begin{aligned} \alpha V^\alpha(i) &\leq \sup_{\phi \in \bar{A}_n(i)} \inf_{\psi \in \bar{B}_n(i)} \left\{ r(i, \phi, \psi) + \sum_{j=0}^{n-1} q_{ij}(\phi, \psi)(V^\alpha(j) - V^\alpha(n)) \right\} \\ &+ \frac{(\bar{C} + D)w^\delta(i)}{w^{\delta-2}(n+1)}. \\ &= \sup_{\phi \in \bar{A}_n(i)} \inf_{\psi \in \bar{B}_n(i)} \left\{ r_n(i, \phi, \psi) + \sum_{j \in S_n} q_{ij}^n(\phi, \psi) V^\alpha(j) \right\} + \frac{(\bar{C} + D)w^\delta(i)}{w^{\delta-2}(n+1)}, \end{aligned}$$

where the last equality is derived from the definition of the reward and transition rates of  $\mathcal{G}_n$ .

Using a symmetric argument, we can show that

$$\alpha V^\alpha(i) \geq \inf_{\psi \in \bar{B}_n(i)} \sup_{\phi \in \bar{A}_n(i)} \left\{ r_n(i, \phi, \psi) + \sum_{j \in S_n} q_{ij}^n(\phi, \psi) V^\alpha(j) \right\} - \frac{(\bar{C} + D)w^\delta(i)}{w^{\delta-2}(n + 1)}.$$

As in the proof of Theorem 4.1, we can show that the inequality in Assumption 4.2(ii) is satisfied by the transition rates of  $\mathcal{G}_n$  with the same constants  $c_\delta$  and  $d_\delta$ . Therefore, by Lemma 4.4, we conclude that, for every  $i \in S_n$ ,

$$|V_n^\alpha(i) - V^\alpha(i)| \leq \frac{(\bar{C} + D)w^\delta(i)}{(\alpha - c_\delta)w^{\delta-2}(n + 1)} + \frac{(\bar{C} + D)d_\delta}{\alpha(\alpha - c_\delta)w^{\delta-2}(n + 1)}.$$

Recalling the definition of the constants  $\bar{C}$  and  $\mathfrak{M}$ , and letting

$$c = \frac{(2M(\alpha + d_1)(|c_\delta| + d_\delta + 1) + D\alpha(\alpha - c_1))(d_\delta + \alpha)}{\alpha^2(\alpha - c_1)(\alpha - c_\delta)}$$

we have

$$|V_n^\alpha(i) - V^\alpha(i)| \leq c \frac{w^\delta(i)}{w^{\delta-2}(n + 1)}$$

for all  $n \geq 1$  and  $i \in S_n$ . □

The above theorem shows that, if a Lyapunov condition holds for the function  $w^\delta$  (with  $\delta > 2$ ) then, by making a suitable choice of the finite sets of actions  $A_n(i)$  and  $B_n(i)$ , the error when approximating  $V^\alpha(i)$  with  $V_n^\alpha(i)$  is of order  $1/w^{\delta-2}(n + 1)$ . Moreover, we have an explicit expression for the multiplicative constant  $c$  that depends on the initial data (and related constants) of the game model  $\mathcal{G}$ .

**Solving a finite state and action game.** Consider the finite state and actions game  $\mathcal{G}_n$  defined in the beginning of Sect. 4. Let  $\mathbf{q}_n > 0$  be such that

$$\mathbf{q}_n > -q_{ii}^n(a, b) \quad \text{for all } (i, a, b) \in \mathbb{K}_n \tag{4.10}$$

(it suffices to let  $\mathbf{q}_n > w(n)$ ). For  $u = \{u(i)\}_{i \in S_n} \in \mathbb{R}^{n+1}$  define the operator  $T_n u \in \mathbb{R}^{n+1}$  as

$$\begin{aligned} T_n u(i) &= \max_{\phi \in \bar{A}_n(i)} \min_{\psi \in \bar{B}_n(i)} \left\{ r_n(i, \phi, \psi) + \sum_{j \in S_n} q_{ij}^n(\phi, \psi) u(j) \right\} \\ &= \min_{\psi \in \bar{B}_n(i)} \max_{\phi \in \bar{A}_n(i)} \left\{ r_n(i, \phi, \psi) + \sum_{j \in S_n} q_{ij}^n(\phi, \psi) u(j) \right\} \end{aligned}$$

for  $i \in S_n$ . Define also  $\tilde{T}_n u \in \mathbb{R}^{n+1}$  as

$$\tilde{T}_n u(i) = \max_{\phi \in \bar{A}_n(i)} \min_{\psi \in \bar{B}_n(i)} \left\{ \frac{r_n(i, \phi, \psi)}{\alpha + \mathbf{q}_n} + \frac{\mathbf{q}_n}{\alpha + \mathbf{q}_n} \sum_{j \in S_n} \left( \frac{q_{ij}^n(\phi, \psi)}{\mathbf{q}_n} + \delta_{ij} \right) u(j) \right\} \tag{4.11}$$

$$= \min_{\psi \in \bar{B}_n(i)} \max_{\phi \in \bar{A}_n(i)} \left\{ \frac{r_n(i, \phi, \psi)}{\alpha + \mathbf{q}_n} + \frac{\mathbf{q}_n}{\alpha + \mathbf{q}_n} \sum_{j \in S_n} \left( \frac{q_{ij}^n(\phi, \psi)}{\mathbf{q}_n} + \delta_{ij} \right) u(j) \right\} \tag{4.12}$$

for  $i \in S_n$  (cf. Section 8 in Guo and Hernández-Lerma 2005). It is easily seen that the equation  $\alpha u = T_n u$  is equivalent to the fixed point equation  $u = \tilde{T}_n u$ . Therefore, as a consequence of Theorem 2.11, the value  $V_n^\alpha$  of the game  $\mathcal{G}_n$  is the unique fixed point of the operator  $\tilde{T}_n$ . Moreover, by a standard calculation, it follows that  $\tilde{T}_n$  is a contraction operator on  $\mathbb{R}^{n+1}$  with modulus  $\mathbf{q}_n/(\alpha + \mathbf{q}_n) < 1$  when considering the supremum norm; that is,

$$\|\tilde{T}_n u - \tilde{T}_n v\| \leq \frac{\mathbf{q}_n}{\alpha + \mathbf{q}_n} \|u - v\| \quad \text{for all } u, v \in \mathbb{R}^{n+1}.$$

Hence, the iterative procedure (which is a sort of value iteration algorithm):

1. Fix arbitrary  $u_0 \in \mathbb{R}^{n+1}$ ,
2. For  $k \geq 1$ , let  $u_k = \tilde{T}_n u_{k-1}$ ,

converges geometrically to  $V_n^\alpha$  in the supremum norm. Concerning the computation of  $\tilde{T}_n u$  for a given  $u \in \mathbb{R}^{n+1}$ , we can apply our next lemma, which uses the following notation. Given a positive integer  $N$ , define  $\Delta_N$  as the set of nonnegative  $\lambda_1, \dots, \lambda_N$  such that  $\lambda_1 + \dots + \lambda_N = 1$ .

**Lemma 4.6** *Given the real-valued matrix  $C = \{C_{s,t}\}_{1 \leq s \leq I, 1 \leq t \leq J}$ , define*

$$V^* = \max_{\lambda \in \Delta_I} \min_{1 \leq t \leq J} \sum_{1 \leq s \leq I} \lambda_s C_{s,t} = \min_{\mu \in \Delta_J} \max_{1 \leq s \leq I} \sum_{1 \leq t \leq J} \mu_t C_{s,t}.$$

*Let  $\mathbf{c} \geq 0$  be such that all the elements of the matrix  $D$ , with  $D_{s,t} = C_{s,t} + \mathbf{c}$ , are strictly positive. Consider the linear programming problem*

$$\min \mathbf{1}'\mathbf{x} \quad \text{subject to} \quad D'\mathbf{x} \geq \mathbf{1}, \mathbf{x} \geq \mathbf{0},$$

*and let  $\mathbf{x}^* \in \mathbb{R}^I$  be an optimal solution. Then,  $V^* = \frac{1}{\mathbf{1}'\mathbf{x}^*} - \mathbf{c}$ .*

*Proof* We have

$$V^* + \mathbf{c} = \max_{\lambda \in \Delta_I} \min_{1 \leq t \leq J} \sum_{1 \leq s \leq I} \lambda_s D_{s,t} =: \tilde{V}$$

and suppose that  $D_{s,t} \geq \epsilon > 0$  for all  $s$  and  $t$ . Observe that  $\tilde{V}$  equals the optimum of the linear programming problem: maximize  $v$  subject to

$$v \leq \sum_{1 \leq s \leq I} \lambda_s D_{s,t} \quad \text{for all } 1 \leq t \leq J,$$

with  $v \geq \epsilon$  and  $\lambda \in \Delta_I$ . Letting  $x_s = \lambda_s/v$  for  $s = 1, \dots, I$ , it follows that  $1/\tilde{V}$  equals the optimum of:

$$\min \mathbf{1}'\mathbf{x} \quad \text{subject to} \quad D'\mathbf{x} \geq \mathbf{1}, \mathbf{x} \geq \mathbf{0}, \mathbf{1}'\mathbf{x} \leq 1/\epsilon,$$

where the last constraint is redundant. □

Therefore, once  $u_{k-1}$  is known, we can effectively compute  $u_k$  by solving the linear programming problem described in Lemma 4.6. Namely, given  $i \in S_n$  and for all  $a_s \in A_n(i)$  with  $1 \leq s \leq I$ , and all  $b_t \in B_n(i)$  with  $1 \leq t \leq J$ , define

$$C_{s,t} = \frac{r_n(i, a_s, b_t)}{\alpha + \mathbf{q}_n} + \frac{\mathbf{q}_n}{\alpha + \mathbf{q}_n} \sum_{j \in S_n} \left( \frac{q_{ij}^n(a_s, b_t)}{\mathbf{q}_n} + \delta_{ij} \right) u_{k-1}(j)$$

and then use Lemma 4.6 to determine  $u_k$ .

Regarding a stopping criterion for the above algorithm, we have the following result. In the next lemma, the norm  $\|\cdot\|$  refers to the supremum norm on  $\mathbb{R}^{n+1}$ .

**Lemma 4.7** *Given the finite state and actions game model  $\mathcal{G}_n$ , consider the sequence of iterates  $\{u_k\}_{k \geq 0}$ , where  $u_0 \in \mathbb{R}^{n+1}$  is arbitrary and, for  $k \geq 1$ ,  $u_k = \tilde{T}_n u_{k-1}$ . Fix  $\epsilon > 0$  and let  $k \geq 1$  be such that  $\|u_{k-1} - u_k\| \leq \epsilon\alpha/\mathbf{q}_n$ . The following statements hold.*

- (i)  $\|u_k - V_n^\alpha\| \leq \epsilon$ .
- (ii) *The strategy  $\pi_*^1 \in \Pi_n^{1,s}$  such that, for all  $i \in S_n$ ,  $\pi_*^1(\cdot|i)$  attains the maximum in (4.11) for the iteration  $u_{k+1} = \tilde{T}_n u_k$  is  $2\epsilon$ -optimal for player 1, meaning that*

$$V_n^\alpha(i) - 2\epsilon \leq \inf_{\pi^2 \in \Pi_n^2} V_n^\alpha(i, \pi_*^1, \pi^2) \quad \text{for all } i \in S_n.$$

- (iii) *The strategy  $\pi_*^2 \in \Pi_n^{2,s}$  such that, for all  $i \in S_n$ ,  $\pi_*^2(\cdot|i)$  attains the minimum in (4.12) for the iteration  $u_{k+1} = \tilde{T}_n u_k$  is  $2\epsilon$ -optimal for player 2, meaning that*

$$V_n^\alpha(i) + 2\epsilon \geq \sup_{\pi^1 \in \Pi_n^1} V_n^\alpha(i, \pi^1, \pi_*^2) \quad \text{for all } i \in S_n.$$

*Proof* (i) We have

$$\begin{aligned} \|u_k - V_n^\alpha\| &\leq \|u_k - u_{k+1}\| + \|u_{k+1} - V_n^\alpha\| \\ &\leq \frac{\mathbf{q}_n}{\alpha + \mathbf{q}_n} \left( \|u_{k-1} - u_k\| + \|u_k - V_n^\alpha\| \right) \end{aligned}$$



because  $V_n^\alpha$  is the fixed point of  $\tilde{T}_n$ , and so

$$\|u_k - V_n^\alpha\| \leq \frac{\mathbf{q}_n}{\alpha} \|u_{k-1} - u_k\| \leq \epsilon.$$

(ii) For  $u : S_n \rightarrow \mathbb{R}$ , consider the operator

$$\tilde{U}u(i) = \min_{b \in \bar{B}_n(i)} \left\{ \frac{r_n(i, \pi_*^1(\cdot|i), b)}{\alpha + \mathbf{q}_n} + \frac{\mathbf{q}_n}{\alpha + \mathbf{q}_n} \sum_{j \in S_n} \left( \frac{q_{ij}^n(\pi_*^1(\cdot|i), b)}{\mathbf{q}_n} + \delta_{ij} \right) u(j) \right\}$$

for  $i \in S_n$ ,

which is a contraction on  $\mathbb{R}^{n+1}$  with modulus  $\frac{\mathbf{q}_n}{\alpha + \mathbf{q}_n}$ , and let  $W$  be its unique fixed point. The fixed point equation

$$W(i) = \min_{b \in \bar{B}_n(i)} \left\{ \frac{r_n(i, \pi_*^1(\cdot|i), b)}{\alpha + \mathbf{q}_n} + \frac{\mathbf{q}_n}{\alpha + \mathbf{q}_n} \sum_{j \in S_n} \left( \frac{q_{ij}^n(\pi_*^1(\cdot|i), b)}{\mathbf{q}_n} + \delta_{ij} \right) W(j) \right\}$$

for  $i \in S_n$ ,

corresponds to the discounted cost optimality equation of a continuous-time control problem (for player 2) when the strategy of player 1 is  $\pi_*^1$ ; see (Prieto-Rumeau and Hernández-Lerma 2012, Section 3.3). Therefore,  $W(i) = \inf_{\pi^2 \in \Pi_n^2} V_n^\alpha(i, \pi_*^1, \pi^2)$  for all  $i \in S_n$ .

Observe now that

$$\|W - V_n^\alpha\| \leq \|W - u_k\| + \|u_k - V_n^\alpha\|. \tag{4.13}$$

Now, on the one hand,

$$\begin{aligned} \|W - u_k\| &\leq \|W - \tilde{T}_n u_k\| + \|\tilde{T}_n u_k - u_k\| \\ &= \|\tilde{U}W - \tilde{U}u_k\| + \|\tilde{T}_n u_k - u_k\| \\ &\leq \frac{\mathbf{q}_n}{\alpha + \mathbf{q}_n} (\|W - u_k\| + \|u_{k-1} - u_k\|) \end{aligned}$$

because  $\tilde{T}_n u_k = \tilde{U}u_k$ , and so

$$\|W - u_k\| \leq \frac{\mathbf{q}_n}{\alpha} \|u_{k-1} - u_k\|.$$

On the other hand, as established in part (i),  $\|u_k - V_n^\alpha\| \leq \frac{\mathbf{q}_n}{\alpha} \|u_{k-1} - u_k\|$ . From (4.13), we obtain

$$\|W - V_n^\alpha\| \leq \frac{2\mathbf{q}_n}{\alpha} \|u_{k-1} - u_k\| \leq 2\epsilon,$$

and the result follows. The proof of (iii) is similar. □

As a consequence of this lemma, we can explicitly obtain an approximation of the value and nearly optimal strategies for both players for the game model  $\mathcal{G}_n$ .

## 5 Numerical application

In this section, we describe a Markov game  $\mathcal{G}$  on a population system and make a numerical application of our approximation techniques.

A population system is managed by players 1 and 2. The natural birth and death rates per individual are  $\lambda > 0$  and  $\mu > 0$ , respectively. Player 1 is interested in the system having a large population and, to this end, player 1 can decrease the mortality rate (for instance, using a suitable medical policy). On the other hand, the goal of player 2 is to have a small number of individuals; player 2 can choose policies that decrease the birth rate of the system (e.g., discouraging immigration).

We consider the following game model.

- The state space, standing for the number of individuals in the population, is  $S = \{0, 1, 2, \dots\}$ .
- The action sets of the players are  $A = B = [-1, 1]$ , while  $A(i) = B(i) = [-1, 1]$  for all  $i \in S$ .
- The system's transition rates  $q_{ij}(a, b)$  satisfy  $q_{ij}(a, b) = 0$  when  $|i - j| > 1$ . When  $|i - j| \leq 1$  we let

$$q_{01}(a, b) = -q_{00}(a, b) = \lambda - C_b|b|,$$

and, for  $i \geq 1$ ,

$$q_{i,i-1}(a, b) = \mu i - C_a|a|\sqrt{i}, \quad q_{i,i+1}(a, b) = \lambda i - C_b|b|i,$$

with  $q_{ii}(a, b) = -(q_{i,i-1}(a, b) + q_{i,i+1}(a, b))$ , for some constants  $0 < C_a < \mu$  and  $0 < C_b < \lambda$ , and all  $(a, b) \in A \times B$ .

- The payoff rate (interpreted as a reward for player 1 and a cost for player 2) is given by

$$r(i, a, b) = p i + C_r|ab|\sqrt{i} \quad \text{for } i \in S \text{ and } -1 \leq a, b \leq 1,$$

for some constants  $p > 0$  and  $C_r > 0$ .

In the above definitions, the term  $\sqrt{i}$  models the fact that the payoff has a concave behavior with respect to the population size, while the term  $|ab|$  in the payoff rate captures the ‘‘saddle-point interplay’’ between the actions of the players. Note that when the players take the actions  $a = 0$  and  $b = 0$  then they do not act on the dynamic system. In this case, the corresponding Markov process (referred to as the natural population system) is recurrent when  $\lambda \leq \mu$  and transient when  $\lambda > \mu$ .

We consider the Lyapunov function  $w$  given by  $w(i) = (\lambda + \mu + 1) \cdot (i + 1)$  for  $i \in S$ .

**Proposition 5.1** *Consider the population model  $\mathcal{G}$  defined above. If the discount rate  $\alpha > 0$  satisfies  $\lambda - \mu < \alpha$  then Assumption 2.2 holds. If, in addition, we have  $2(\lambda - \mu) < \alpha$  then Assumption 4.2 is satisfied.*

*Proof* Fix an integer  $k \geq 1$  and consider the Lyapunov function  $i \mapsto w^k(i)$ . Given a state  $i \geq 1$ , we have

$$\begin{aligned} \sum_{j \in S} q_{ij}(a, b)w^k(i) &= (w^k(i - 1) - w^k(i))(\mu i - C_a|a|\sqrt{i}) \\ &\quad + (w^k(i + 1) - w^k(i))(\lambda i - C_b|b|i). \end{aligned}$$

Noting that

$$\begin{aligned} (i + 2)^k - (i + 1)^k &= k(i + 1)^{k-1} + O(i^{k-2}) \quad \text{and} \\ i^k - (i + 1)^k &= -k(i + 1)^{k-1} + O(i^{k-2}), \end{aligned}$$

some elementary calculations give

$$\begin{aligned} \sum_{j \in S} q_{ij}(a, b)w^k(i) &= k(\lambda - \mu - C_b|b|)w^k(i) + O\left(i^{k-\frac{1}{2}}\right) \\ &\leq k(\lambda - \mu)w^k(i) + O\left(i^{k-\frac{1}{2}}\right). \end{aligned}$$

Therefore, given an integer  $k \geq 1$  and a constant  $c_k > k(\lambda - \mu)$ , there exists  $d_k \geq 0$  such that

$$\sum_{j \in S} q_{ij}(a, b)w^k(i) \leq c_k w^k(i) + d_k \quad \text{for all } (i, a, b) \in \mathbb{K}. \tag{5.1}$$

If  $\lambda - \mu < \alpha$  then choose  $\lambda - \mu < c_1 < \alpha$ , and so Assumption 2.2(i) holds. In particular, note also that  $-q_{ii}(a, b) \leq w(i)$  for all  $(i, a, b) \in \mathbb{K}$ . Regarding the other statements of Assumption 2.2, item (ii) holds by letting  $M = p + C_r$ , part (iii) is straightforward, and (iv) holds as a consequence of (5.1).

It should be clear that Assumption 4.2(i) is satisfied. If  $2(\lambda - \mu) < \alpha$ , then choose  $\delta > 2$  and  $c_\delta$  such that

$$\delta(\lambda - \mu) < c_\delta < \alpha,$$

and so Assumption 4.2(ii) holds. □

For each  $n \geq 1$ , consider now the finite state and actions game model  $\mathcal{G}_n$  as described in Sect. 4. As a consequence of Theorems 3.4, 4.1, and 4.5, we obtain the following results.

- (i) *Case  $\lambda \leq \mu$  (the natural population system is recurrent).* Given arbitrary discount rate  $\alpha > 0$  we have

$$\lim_{n \rightarrow \infty} V_n^\alpha(i) = V^\alpha(i) \quad \text{for all } i \in S.$$

Given arbitrary  $k > 0$ , by suitably choosing the action sets  $A_n(i)$  and  $B_n(i)$  we have

$$|V_n^\alpha(i) - V^\alpha(i)| = O(n^{-k}) \quad \text{for each } i \in S.$$

(ii) *Case  $\lambda > \mu$  (the natural population system is transient).* Given a discount rate  $\lambda - \mu < \alpha$  we have

$$\lim_{n \rightarrow \infty} V_n^\alpha(i) = V^\alpha(i) \quad \text{for all } i \in S.$$

If the discount rate is such that  $2(\lambda - \mu) < \alpha$  then for each  $0 < k < \frac{\alpha}{\lambda - \mu} - 2$  we can choose the finite sets  $A_n(i)$  and  $B_n(i)$  such that

$$|V_n^\alpha(i) - V^\alpha(i)| = O(n^{-k}) \quad \text{for each } i \in S.$$

**Numerical experimentation.** We choose the following values of the parameters:

$$\lambda = 2.6, \quad \mu = 2.5, \quad \alpha = 1.2, \quad C_a = C_b = C_r = 0.2, \quad \text{and} \quad p = 3.$$

For each  $n \geq 1$ , we consider the truncated game model  $\mathcal{G}_n$  with state space  $\{0, 1, \dots, n\}$ . The action sets  $A_n(i) \equiv A_n$  and  $B_n(i) \equiv B_n$  consist of the  $n + 1$  points  $\frac{2k}{n} - 1$  for  $k = 0, 1, \dots, n$ .

For  $n = 1, \dots, 30$ , we solve the finite game model  $\mathcal{G}_n$  using the value iteration procedure described in the previous section: we start from the initial value  $u_0 = \mathbf{0}$  and we let  $\mathbf{q}_n = \max_{(i,a,b) \in \mathbb{K}_n} \{-q_{ii}^n(a, b)\} + 0.1$ ; recall (4.10). As a stopping criterion for the value iteration algorithm, we let  $\varepsilon = 5 \times 10^{-5}$  and we stop at the iterate  $k$  when

$$\|u_k - u_{k-1}\| \leq \varepsilon \alpha / \mathbf{q}_n,$$

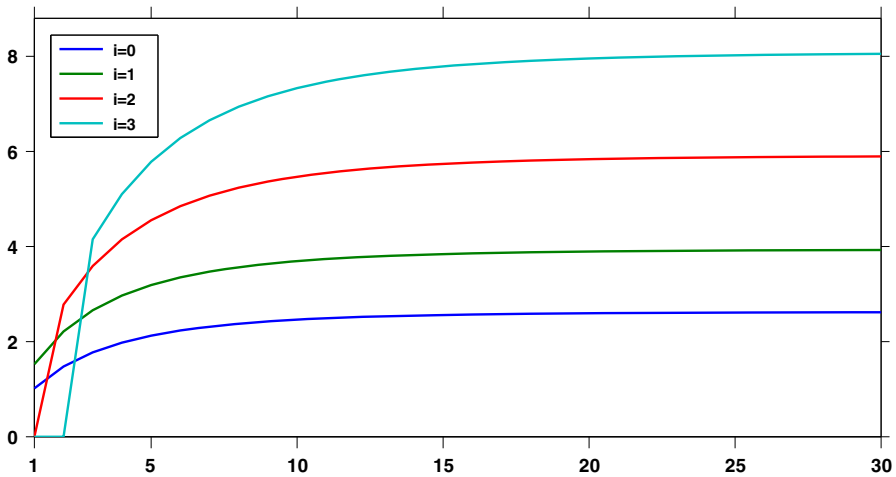
which ensures that  $\|u_k - V_n^\alpha\| \leq \varepsilon$  (this refers to the supremum norm in  $\mathbb{R}^{n+1}$ ).

In Fig. 1, we display the values  $V_n^\alpha(i)$  for  $i = 0, 1, 2, 3$  and  $1 \leq n \leq 30$ . We observe that the values of the games  $\mathcal{G}_n$  become stable for relatively small values of the truncation size  $n$ , say for  $n \geq 20$ . We obtain the approximations

$$V^\alpha(0) \simeq 2.6179, \quad V^\alpha(1) \simeq 3.9269, \quad V^\alpha(2) \simeq 5.8948, \quad V^\alpha(3) \simeq 8.0524.$$

By Lemma 4.7, the approximation error (with respect to the value  $V_{30}^\alpha$  of the game model  $\mathcal{G}_{30}$ ) is less than  $5 \times 10^{-5}$ . Empirically, we observe that convergence seems to occur faster than at the convergence rate given in Theorem 4.5. This is because the bounds used to derive the convergence rate are very conservative.

Concerning the approximation of optimal strategies, for  $n = 30$  we show in Table 1 the randomized strategies  $\pi_*^1(\cdot|i)$  and  $\pi_*^2(\cdot|i)$  for  $i = 0$  and  $i = 1$  as described in Lemma 4.7. Table 1 displays the corresponding probability distributions on the discretized sets of actions  $A_{30} = B_{30}$ . These are  $10^{-4}$ -optimal strategies. Empirically, this suggests that the optimal strategy for player 1 in the game model  $\mathcal{G}$  will be to



**Fig. 1** Value of the games  $V_n^\alpha(i)$  for  $n = 1, \dots, 30$

**Table 1** Optimal strategies in  $\bar{A}_{30}$  and  $\bar{B}_{30}$  for  $\mathcal{G}_{30}$

|          |         | Actions in $A_{30}$ and $B_{30}$ |            |            |     |            |       |       |
|----------|---------|----------------------------------|------------|------------|-----|------------|-------|-------|
|          |         | -1.000                           | -0.933     | -0.867     | ... | 0.867      | 0.933 | 1.000 |
| Player 1 | $i = 0$ | 0.033                            | 0.033      | 0.033      | ... | 0.033      | 0.033 | 0.000 |
|          | $i = 1$ | 0.499                            | $<10^{-4}$ | $<10^{-4}$ | ... | $<10^{-4}$ | 0.499 | 0.000 |
| Player 2 | $i = 0$ | 0.499                            | $<10^{-4}$ | $<10^{-4}$ | ... | $<10^{-4}$ | 0.499 | 0.000 |
|          | $i = 1$ | 0.499                            | $<10^{-4}$ | $<10^{-4}$ | ... | $<10^{-4}$ | 0.499 | 0.000 |

choose his actions uniformly on  $[-1, 1]$  in state  $i = 0$ , and to randomize between actions  $-1$  and  $1$ , with probabilities  $1/2$ , in state  $i = 1$ . For player 2, the estimation of an optimal strategy is to randomize between actions  $-1$  and  $1$ , with probabilities  $1/2$ , in both states  $i = 0$  and  $i = 1$ .

**Acknowledgments** Research supported by Grant MTM2012-31393 from the Spanish Ministerio de Economía y Competitividad.

**References**

Billingsley P (1968) Convergence of probability measures. Wiley, New York  
 Bogachev VI (2007) Measure theory, vol II. Springer, New York  
 Bolley F (2008) Separability and completeness for the Wasserstein distance. In: Séminaire de probabilités XLI. Lecture Notes in Math. 1934, Springer, Berlin, pp 371–377  
 Chang HS, Hu JQ, Fu MC, Marcus SI (2010) Adaptive adversarial multi-armed bandit approach to two-person zero-sum Markov games. IEEE Trans Automat Control 55:463–468  
 Frenk JBG, Kassay G, Kolumbán J (2004) On equivalent results in minimax theory Euro. J Oper Res 157:46–58  
 Guo XP, Hernández-Lerma O (2003) Zero-sum games for continuous-time Markov chains with unbounded transition and average payoff rates. J Appl Probab 40:327–345

- Guo XP, Hernández-Lerma O (2003) Drift and monotonicity conditions for continuous-time controlled Markov chains with an average criterion. *IEEE Trans Automat Control* 48:236–244
- Guo XP, Hernández-Lerma O (2005) Zero-sum continuous-time Markov games with unbounded transition and discounted payoff rates. *Bernoulli* 11:1009–1029
- Guo XP, Hernández-Lerma O (2009) *Continuous-time Markov decision processes: theory and applications*. Springer, New York
- Guo XP, Zhang WZ (2014) Convergence of controlled models and finite-state approximation for discounted continuous-time Markov decision processes with constraints. *Eur J Oper Res* 238:486–496
- Jaśkiewicz A, Nowak AS (2006) Approximation of noncooperative semi-Markov games. *J Optim Theory Appl* 131:115–134
- Nowak AS, Altman E (2002)  $\epsilon$ -equilibria for stochastic games with uncountable state space and unbounded costs. *SIAM J Control Optim* 40:1821–1839
- Prieto-Rumeau T, Hernández-Lerma O (2012) Discounted continuous-time controlled Markov chains: convergence of control models. *J Appl Probab* 49:1072–1090
- Prieto-Rumeau T, Hernández-Lerma O (2012) *Selected topics on continuous-time controlled Markov chains and Markov games*. Imperial College Press, London
- Prieto-Rumeau T, Lorenzo JM (2010) Approximating ergodic average reward continuous-time controlled Markov chains. *IEEE Trans Automat Control* 55:201–207