Average Optimality in Markov Games with General State Space

Ulrich Rieder
Department of Mathematics VII
University of Ulm

D-89069 Ulm, Germany

Abstract

A Markov game with general state space and the average reward as optimality criterion is considered. Asymmetric conditions are presented under which the game has a value and both player have ϵ-optimal stationary policies. We also study the convergence of the value iteration and the linear programming formulation. It is shown that the associated linear programs are solvable and that there is no duality gap.

Keywords and phrases: (Discrete-time) zero-sum Markov-games, average reward criterion, value iteration, infinite linear programming, strong duality

1 Introduction

This paper considers two-person zero-sum Markov games with Borel state and action spaces and with the expected average reward as optimality criterion. For finite or denumerable state spaces optimality results and algorithms have been derived by several authors. (cf. e.g. Raghavan and Filar [9], Sennott [14]). Markov games with general state spaces and the long run average payoff function have recently been investigated by Nowak [8] (see also his list of references for earlier papers). In this paper we are interested in asymmetric sufficient conditions under which the Markov game has a value and both players have ϵ-optimal stationary policies. We also present a linear programming formulation (in general vector spaces) and show that the associated linear programs are solvable and that there is no duality gap. The analysis is essentially based on some results on ergodicity and on measurable selection theorems.

Let us now briefly indicate how this paper is organized. In section 2 we formulate the mathematical game model. Section 3 is concerned with ϵ-optimal stationary policies. Sufficient conditions are given for a pair of stationary policies. In section 4 we introduce two 'measures of ergodicity' which are straightforward generalizations of the corresponding ergodicity factors in stochastic dynamic programming (or one-person Markov games) (cf. [11], Hernández-Lerma [4] and their references). The existence results are given in section 5. We first study the convergence of the value iteration. It is shown that the rate of convergence is geometric (with a well-defined best rate). Then we introduce the conditions (W) and (S) which are closely related to the continuity and compactness conditions of a previous paper (cf. [12]). Under each of the conditions the zero-sum Markov game has a value, for any $\epsilon > 0$ there exists an ϵ-optimal stationary policy for player I and an optimal stationary policy for player II. Another approach to the average reward problem via discounted Markov games is indicated in Theorem 5.10. Section 6 is concerned with the linear programming formulation of Markov games with Borel state and action spaces and with the average reward criterion. The main results are stated in Theorems 6.2 and 6.3.

2 The mathematical model

The mathematical model for a stationary two-person zero-sum **Markov game** is a tupel (S, X, Y, A, B, P, r) of the following meaning:

(i) S is the **state space**. S is assumed to be a Borel space, i.e. S is a non-empty Borel subset of some Polish (complete, separable metric) space and is endowed with the σ-algebra of Borel subsets S.

(ii) X and Y are the **action spaces** for player I and II, resp.. X and Y are assumed to be Borel spaces.

(iii) A and B are measurable subsets of $S \times X$ and $S \times Y$, resp.. It is assumed that A contains the graph of a measurable map from S into X. For any $s \in S$ the non-empty and measurable s-section $A(s)$ of A is called the **set of all admissible actions** for player I, if the system is in state s. Analogously for B. We assume that

$$D = \{(s, a, b) : a \in A(s),\ b \in B(s)\}$$

is a measurable subset of $S \times X \times Y$.

(iv) P is a transition probabilitiy (or kernel) from D to S. P specifies the **transition law** during a single stage.

(v) r is a bounded nonnegative measurable function on D, the so-called **reward function** for player I. Then $-r$ is the reward function for player II.

We use $P(X)$ to denote the set of all probability measures on X. Let $P(X)$ be endowed with the weak* topology based on bounded continuous functions. Then the σ-algebra of Borel subsets of $P(X)$ coincides with the smallest σ-algebra on $P(X)$ such that for each Borel subset C of X the map $p \to p(C)$ is measurable. Further, $P(X)$ is a Borel space. For $s \in S$ let $P(A(s))$ be the set of all probability measures on X with support $A(s)$. Then

$$K = \{(s, \mu, \nu) : \mu \in P(A(s)), \nu \in P(B(s))\}$$

is a measurable subset of $S \times P(X) \times P(Y)$.
We write $H_1 = S$, $H_{n+1} = D \times H_{n+1}$, $n \in \mathbb{N}$. As usual, a (randomized) **policy** for player I is defined as a sequence $\pi = (\pi_n)$ of measurable maps $\pi_n : H_n \to P(X)$ such that $\pi_n(h) \in P(A(s_n))$ for any $h = (s_1, a_1, b_1, \ldots, s_n)$, $n \in \mathbb{N}$. Let Δ denote the set of all policies for player I. We write F for the set of all measurable maps $f : S \to P(X)$ such that $f(s) \in P(A(s))$ for $s \in S$. A (randomized) **stationary policy** is a policy $(f_n) \in F^\infty$ where $f_n = f$ is independent of n. For such a policy we also write shortly f. Similarly, the set Σ of all policies and the set G are defined for player II.
In this paper we are concerned with the expected average reward criterion. For any $\pi \in \Delta$ and $\sigma \in \Sigma$ define

$$\phi_{\pi\sigma}(s) = \limsup_{n\to\infty} \frac{1}{n} E_{\pi\sigma s}\left[\sum_{k=1}^{n} r(\zeta_k, \alpha_k, \beta_k)\right], s \in S$$

where (ζ_k), (α_k) and (β_k) are the sequences of states and actions and $E_{\pi\sigma s}$ denotes the conditional expectation given that player I and II use the policies π and σ, resp., and the system starts in $s \in S$. Define

$$\phi = \inf_\sigma \sup_\pi \phi_{\pi\sigma}.$$

The function $s \to \phi(s)$ is called the **upper value function** of the Markov game. The zero-sum Markov game has a value if

$$\sup_\pi \inf_\sigma \phi_{\pi\sigma} = \inf_\sigma \sup_\pi \phi_{\pi\sigma}$$

and in this case ϕ is called the **value function** of the game.

Let $B(S)$ denote the set of all bounded measurable functions on S. For $f \in F, g \in G$ and $(s,\mu,\nu) \in K$ the isotone operators L, L_{fg} and U are well-defined on $B(S)$ by

$$\begin{aligned}
Lu(s,\mu,\nu) &= \int \mu(da) \int \nu(db) \left[r(s,a,b) + \int P(s,a,b,ds')u(s')\right] \\
&= r(s,\mu,\nu) + \int P(s,\mu,\nu,ds')u(s'), \\
L_{fg}u(s) &= Lu(s,f(s),g(s)), \\
&= r_{fg}(s) + P(f,g)u(s), \\
Uu(s) &= \inf_{\nu} \sup_{\mu} Lu(s,\mu,\nu),\ s \in S.
\end{aligned}$$

3 Stationary ϵ-optimal policies

Let $\epsilon \geq 0$. If the Markov game has a value, then $\pi^* \in \Delta$ will be called **ϵ-optimal for player I** if $\phi_{\pi^*\sigma} \geq \phi - \epsilon$ for all $\sigma \in \Sigma$, and $\sigma^* \in \Sigma$ will be called **ϵ-optimal for player II** if $\phi_{\pi\sigma^*} \leq \phi + \epsilon$ for all $\pi \in \Delta$. A policy is **optimal** if it is 0-optimal . In the sequel an important role will play the functional 'span' defined on the linear space $B(S)$ by

$$sp\ u = \sup\ u - \inf\ u.$$

It is easy to see that $sp(\cdot)$ is a semi-norm on $B(S)$, and $sp\ u = 0$ if and only if u is constant. The function $u \in B(S)$ is called a **span-fixed-point** of U if $sp(Uu - u) = 0$.

The following sufficient conditions for optimality may be regarded as extensions of the corresponding conditions in stochastic dynamic programming problems (cf. the survey [2]). Further conditions can be found in the recent paper [13].

Lemma 3.1 Let $u \in B(S)$. Let $\epsilon \geq 0$. If there exists $g \in G$ such that

$$\sup_f L_{fg}u \leq Uu + \epsilon$$

then

$$\phi_{\pi g} \leq \sup(Uu - u) + \epsilon \qquad \text{for all } \pi \in \Delta.$$

Proof. Let $\pi \in \Delta$. Making repeated use of the assumption we obtain

$$E_{\pi g s}\left[\sum_{k=1}^{n} r(\zeta_k, \alpha_k, \beta_k)\right] \le E_{\pi g s}\left[\sum_{k=1}^{n} r(\zeta_k, \alpha_k, \beta_k) + u(\zeta_{n+1})\right] - \inf\ u$$
$$\le u(s) + n \sup(Uu - u) + n\epsilon - \inf\ u$$

and

$$\frac{1}{n} E_{\pi g s}\left[\sum_{k=1}^{n} r(\zeta_k, \alpha_k, \beta_k)\right] \le \frac{u(s)}{n} - \frac{\inf\ u}{n} + \sup(Uu - u) + \epsilon.$$

Letting $n \to \infty$ we get immediately the assertion. ∎

Theorem 3.2 Let $u \in B(S)$ be a span-fixed-point of U. If for any $\epsilon > 0$ there exists $f_\epsilon \in F$ and $g^* \in G$ such that

$$\sup_f L_{f g^*} u = Uu \le \inf_g L_{f_\epsilon g} u + \epsilon$$

then the Markov game has a value, $\phi(s) = \sup(Uu - u) = \inf(Uu - u)$ for all $s \in S$, the stationary policy $(f_\epsilon, f_\epsilon, \ldots)$ is ϵ-optimal for player I and the stationary policy $(g^*, g^*, \ldots)$ is optimal for player II.

Proof. At first it follows from the assumption that

$$Uu = \sup_f \inf_g L_{fg} u.$$

By interchanging the role of the two players we conclude from Lemma 3.1 that

$$\phi_{f_\epsilon \sigma} \ge \inf(Uu - u) - \epsilon \quad \text{for all } \sigma \in \Sigma,$$

and

$$\phi_{\pi g*} \le \sup(Uu - u) \quad \text{for all } \pi \in \Delta.$$

Since $sp(Uu - u) = 0$, it is easy to see that the game has a value. The other assertions are obvious. ∎

From the proof above it follows that Theorem 3.2 (and hence the results of this paper) remains true if we would have taken the 'liminf' criterion.

$$\liminf_{n \to \infty} \frac{1}{n} E_{\pi \sigma s}\left[\sum_{k=1}^{n} r(\zeta_k, \alpha_k, \beta_k)\right]$$

instead of $\phi_{\pi\sigma}(s)$. In particular, for the stationary policies both limits are equal.

Corollary 3.3 Let $u \in B(S)$ be a span-fixed-point of U. If there exists a pair $(f^*, g^*) \in F \times G$ such that for all $f \in F$ and $g \in G$

$$L_{fg^*}u \leq L_{f^*g^*}u \leq L_{f^*g}u$$

then the Markov game has a value, $\phi(s) = \sup(Uu - u) = \inf(Uu - u)$ for all $s \in S$ and the stationary policies $(f^*, f^*, \ldots)$ and $(g^*, g^*, \ldots)$ are optimal for player I and player II, resp.

For a *proof* of the optimality of $(f^*, f^*, \ldots)$ use Lemma 3.1.

4 Ergodicity factors

In this section we shall define ergodicity factors which are closely related to the ergodic coefficient in Markov chain theory (cf. e.g.[6]). We use the following abbreviations for $f \in F$, $g \in G$ and $u \in B(S)$

$$P_s(f,g)u := P(f,g)u(s)$$

$$P_s^n(\pi,\sigma) := P_s(f_1,g_1)P(f_2,g_2)\ldots P(f_n,g_n), n \in I\!N$$

where $\pi = (f_1, f_2, \ldots, f_n) \in F^n$ and $\sigma = (g_1, g_2, \ldots, g_n) \in G^n$. If $\pi = (f, f, \ldots, f) \in F^n$ and $\sigma = (g, g, \ldots, g) \in G^n$ we write $P_s^n(f,g)$ instead of $P_s^n(\pi,\sigma)$. Then $P_s^n(f,g)$ is the n-step transition probability of the Markov chain $(\zeta_1, \zeta_2, \ldots,)$ associated with the stationary policies $(f, f, \ldots)$ and $(g, g, \ldots)$. For any probability measures P and Q on the Borel σ- algebra in S we define

$$|P - Q| = \sup_C |P(C) - Q(C)|.$$

Then $2|P - Q|$ is the total variation of the signed measure $P - Q$.

Let the nonnegative numbers γ_n and δ_n be defined by

$$\gamma_n = \sup_{s \neq s'} \sup_{f,g} |P_s^n(f,g) - P_{s'}^n(f,g)|, \quad n \in I\!N$$

$$\delta_n = \sup_{s \neq s'} \sup_{\pi,\pi'} \sup_{\sigma,\sigma'} |P_s^n(\pi,\sigma) - P_{s'}^n(\pi',\sigma')|, \quad n \in I\!N$$

where the supremum is taken over all n-stage policies $\pi, \pi' \in F^n$ and $\sigma, \sigma' \in G^n$. It is not difficult to show that

$$\gamma_n \le \delta_n \le 1 - \mu(S) \le 1, \quad n \in \mathbb{N}$$

where μ is some measure on the Borel σ-algebra in S such that for all Borel subsets C of S

$$P(s, a, b, C) \ge \mu(C), \quad (s, a, b) \in D.$$

The proofs of the following results run along the same lines as the proofs of the corresponding results in stochastic dynamic programming (cf.[11], [4]) and are, therefore, not given here.

Theorem 4.1 The following statements are equivalent:

(i) $\gamma := \inf_n \gamma_n^{1/n} < 1$

(ii) $\gamma_n \to 0 \; (n \to \infty)$

(iii) There are nonnegative numbers $\rho < 1$ and $c > 0$ such that for any pair $(f, g) \in F \times G$ a probability measure $Q(f, g)$ exists for which

$$|P_s^n(f, g) - Q(f, g)| \le c\rho^n, \quad s \in S, n \in \mathbb{N}$$

(iv) There is an integer $m \in \mathbb{N}$ and a number $d > 0$ such that for any pair $(f, g) \in F \times G$ a probability measure $Q(f, g)$ exists for which for all Borel subsets C of S

$$P_s^m(f, g)(C) \ge dQ(f, g)(C).$$

Let $\gamma < 1$. Theorem 4.1 tells us that for all pairs $(f, g) \in F \times G$ the kernels $P(f, g)$ satisfy the Doeblin condition (which is equivalent to the condition that $P(f, g)$ is quasi- compact) and it follows that

$$\phi_{fg}(s) = Q(f, g) r_{fg} \quad \text{for all } s \in S.$$

If the state space S is denumerable, then the condition $\gamma < 1$ implies that for all pairs $(f, g) \in F \times G$ the stochastic matrix $P(f, g)$ has no two disjoint closed sets of states and $P(f, g)$ has only one positive recurrent and aperiodic class of states.

It is easy to see from the definition that Uu is bounded, but Uu need not be universally measurable . Therefore we use the following device and make the following assumption throughout this paper.

Assumption (A). There exists a subspace $M(S)$ of $B(S)$ such that $U : M(S) \to M(S)$.

Examples of such subspaces will be given in the next section. Moreover, we shall assume that for any $u \in M(S)$

$$U^n u = \inf_\sigma \sup_\pi L_{f_1 g_1} \dots L_{f_n g_n} u, \; n \in I\!N$$

where the infimum and supremum are taken over all n-stage policies $\sigma = (g_k) \in G^n$ and $\pi = (f_k) \in F^n$, respectively. This assumption implies that we may restrict our attention to Markov policies in finite-stage games and is satisfied under each of the conditions (W) and (S) of section 5. The assumption is essentially used in Theorem 4.2. Furthermore we make use of the well-known inequalities

$$\inf(v - w) \le Uv - Uw \le \sup(v - w)$$

Theorem 4.2. Let $m \in I\!N$. Let $v, w \in M(S)$. Then

$$sp(U^n v - U^n w) \le \delta_m^{\lfloor n/m \rfloor} sp(v - w), \; n \in I\!N$$

where $\lfloor x \rfloor$ is the largest integer not greater than x.

If $M(S)$ is a complete metric subspace of $B(S)$ (under the metric $(v, w) \to \|v - w\|$), then the operator U^m has a span-fixed-point provided that $\delta_m < 1$.

5 The main results

In this paper we are mainly interested in the construction of optimal policies by means of finite-stage Markov games. Another approach by discounted games is shortly mentioned at the end of this section.
For any given $V^0 \in M(S)$ define recursively the functions $V^n \in M(S)$ by

$$V^n = UV^{n-1}, \; n \in I\!N.$$

Then $s \to V^n(s)$ is the upper value function in the n-stage Markov game (including the terminal reward V^0) when the system starts in $s \in S$.

In the sequel an important role will play the number δ defined by

$$\delta = \inf_n \delta_n^{1/n}.$$

Then $\delta \le 1$ and $\delta < 1$ if and only if $\delta_m < 1$ for some $m \in I\!N$.
Concerning the convergence of the value iteration we get the following results.

Theorem 5.1. Let $\delta_m < 1$.

(a) Let $u \in M(S)$ be a span-fixed-point of U. There exists $c \in \mathbb{R}$ such that for the constant $d = Uu - u$

$$||V^n - nd - u - c|| \le \delta_m^{\lfloor n/m \rfloor} sp(V^0 - u),\ n \in \mathbb{N}.$$

(b) The sequences $(\inf(V^n - V^{n-1}))$ and $(\sup(V^n - V^{n-1}))$ are non-decreasing and non-increasing, respectively, and both sequences converge geometrically to the same finite limit.

Proof.

(a) Consider the bounded function $u_n = V^n - nd - u$, $n \in \mathbb{N}$. Since

$$u_{n+1} = UV^n - U(u + nd)$$

we conclude

$$\inf u_n \le \inf u_{n+1} \le \sup u_{n+1} \le \sup u_n,\ n \in \mathbb{N}.$$

Moreover we have by Theorem 4.2

$$sp(u_n) = sp(U^n V^0 - U^n u) \le \delta_m^{\lfloor n/m \rfloor} sp(V^0 - u),\ n \in \mathbb{N}.$$

Altogether there exists $c \in \mathbb{R}$ such that

$$||u_n - c|| \le sp(u_n),\ n \in \mathbb{N}.$$

(b) follows in the same way. $\square$

It follows from Theorem 5.1 that for some $c \in \mathbb{R}$

$$\lim_n (V^n - nd) = u + c$$

and that the convergence is geometric (with 'best' geometric rate δ).

From Lemma 3.1 we obtain immediately

Theorem 5.2. If there exists a pair $(f_n, g_n) \in F \times G$ such that for all $f \in F$ and $g \in G$

$$L_{fg_n} V^{n-1} \le L_{f_n g_n} V^{n-1} \le L_{f_n g} V^{n-1}$$

then

$$\phi \leq \sup_{\pi} \phi_{\pi g_n} \leq \sup(V^n - V^{n-1})$$

and

$$\phi \geq \inf_{\sigma} \phi_{f_n \sigma} \geq \inf(V^n - V^{n-1}).$$

Now we fix some state $s_0 \in S$ and define

$$v_n(s) = V^n(s) - V^n(s_0), \; s \in S, n \in I\!N_0.$$

The function v_n is called the **relative upper value funtion at stage** n. Theorem 5.3 and Corollary 5.4 are direct consequences of Theorem 4.2.

Theorem 5.3. If $\delta_m < 1$, then (v_n) is uniformly bounded, $v := \lim_n v_n$ exists and

$$\|v - v_n\| \leq \delta_m^{\lfloor n/m \rfloor} sp(V^1 - V^0) \sum_{k=0}^{\infty} \delta_m^{\lfloor k/m \rfloor}, \; n \in I\!N.$$

Corollary 5.4 Let $\delta < 1$. If $v = \lim_n v_n \in M(S)$ and v is a span-fixed-point of U, then v is the unique span-fixed-point u of U such that $u(s_0) = 0$.
Note that the sequence (v_n) converges uniformely to v at a geometric rate and that the 'best' geometric rate of convergence is given by δ.

The purpose of this paper is to give sufficient conditions for the following **statements**:

(I): $v = \lim_n v_n$ is a span-fixed-point of U.

(II): $\phi = \inf_{\sigma} \sup_{\pi} \phi_{\pi\sigma} = \sup_{\pi} \inf_{\sigma} \phi_{\pi\sigma}$.

(III): For any $\epsilon > 0$ there exists an ϵ-optimal stationary policy for player I.

(IV): $Ls_n \Gamma_n(s) \subset \Gamma(s), \; s \in S$, where
$\Gamma_n(s) = \{\nu \in P(B(s)) : \sup_\mu LV^{n-1}(s, \mu, \nu) = V^n(s)\}, \; s \in S$
$\Gamma(s) = \{\nu \in P(B(s)) : \sup_\mu Lv(s, \mu, \nu) = Uv(s)\}, \; s \in S.$

(V): There exists an optimal stationary policy $(g, g, \ldots)$ for player II such that $g(s) \in Ls_n \Gamma_n(s), \; s \in S$.

We recall that $Ls_n C_n$ denotes the topological limes superior of a sequence (C_n) of subsets C_n of some metric space M, i.e. $Ls_n C_n = \{x \in M : x$ is an accumulation point of some sequence (x_n) with $x_n \in C_n, n \in I\!N\}$.

Statement (II) says that the Markov game has a value. We shall prove (see e.g. Theorem 5.7) that

$$\phi(s) = \lim_n (V^n(s_0) - V^{n-1}(s_0)),\ s \in S.$$

The proof of the statements (II), (III) and (V) is essentially based on a minimax theorem (for one-stage games) and on measurable selection theorems (cf. Theorem 3.2). Hence the main task consists of finding conditions such that the following three results are valid for any $u \in M(S)$

(i) $\quad Uu = \sup_f \inf_g L_{fg}u.$

(ii) For any $\epsilon > 0$ there exists $f \in F$ such that $\inf_g L_{fg}u \geq Uu - \epsilon$.

(iii) There exists $g \in G$ such that $\sup_f L_{fg}u = Uu$.

Let the statements (i), (ii) and (iii) be satisfied. By assumption (A), it follows by induction that for any $u \in M(S)$

$$\begin{aligned} U^n u &= \inf_\sigma \sup_\pi L_{f_1 g_1} \dots L_{f_n g_n} u \\ &= \sup_\pi \inf_\sigma L_{f_1 g_1} \dots L_{f_n g_n} u,\ n \in \mathbb{N} \end{aligned}$$

where the infimum and supremum are taken over all n-stage policies $\sigma = (g_k) \in G^n$ and $\pi = (f_k) \in F^n$, respectively.
A key tool for proving the statements (I) and (IV) is the following result (cf. [12], Lemma 4.1).

Lemma 5.5. Let X and Y be metric spaces. Let (u_n) be a uniformly convergent sequence of bounded functions $u_n : X \times Y \to \mathbb{R}$ such that $u_n(x, \cdot)$ is lower semi-continuous on Y, $x \in X$. Let $u_\infty = \lim_n u_n$.

(a) $u_\infty(x, \cdot)$ is lower semi-continuous on Y, $x \in X$.

(b) $Ls_n \psi_n \subset \psi_\infty$, where

$$\psi_n = \{y \in Y : \sup_x u_n(x, y) = \inf_y \sup_x u_n(x, y)\},\ n \in \mathbb{N} + \{\infty\}$$

(c) If Y is compact, then

$$\lim_n \inf_y \sup_x u_n(x, y) = \inf_y \sup_x \lim_n u_n(x, y).$$

5.1 Condition (W)

The transition probability P is called w-**continuous** if $(s,a,b) \to \int P(s,a,b,ds')u(s')$ is continuous for all bounded continuous functions $u : S \to \mathbb{R}$. Let $C_0(S)$ denote the set of all bounded lower semi-continuous functions on S. Then it follows by the well-known Portmanteau-Theorem that P is w-continuous if and only if $(s,a,b) \to \int P(s,a,b,ds')u(s')$ is lower semi-continuous for all $u \in C_0(S)$.

Throughout this subsection we impose

Condition (W).

(1) $A(s)$ is closed for $s \in S$, $s \to A(s)$ is lower semi-continuous.

(2) $B(s)$ is compact for $s \in S$, $s \to B(s)$ is upper semi-continuous.

(3) P is w-continuous.

(4) r is lower semi-continuous.

Lemma 5.6. U maps $C_0(S)$ into itself and the statements (i), (ii) and (iii) are valid.

Proof. At first we observe that $P(A(s))$ is closed, $s \to P(A(s))$ is lower semi-continuous and $P(B(s))$ is compact and $s \to P(B(s))$ is upper semi-continuous. Now let $u \in C_0(S)$. Then Lu is lower semi-continuous on K and from the well-known maximum theorems we get that $Uu \in C_0(S)$. Statement (i) follows immediately from Theorem 2 in [3]. By the selection theorem of Dubins and Savage (cf.[10], Theorem 4.9) statement (iii) is valid. For a proof of (ii) define for $\epsilon > 0$

$$\varphi(s) := \{\mu \in P(A(s)) : \inf_{\nu} Lu(s,\mu,\nu) > Uu(s) - \epsilon\}.$$

Then $Gr\varphi = \{(s,\mu) : \mu \in \varphi(s)\}$ is a Borel subset of $S \times P(X)$ and $\varphi(s)$ is open for $s \in S$. Then there exists $f \in F$ such that $f(s) \in \varphi(s)$ for all $s \in S$ (cf.[10]), Example 2.10) and the proof is complete. ∎

Theorem 5.7. Let $\delta < 1$. If condition (W) is satisfied then $v \in C_0(S)$ and the statements (I), (II), (III), (IV) and (V) are valid.

Proof. By Lemma 5.5 and Lemma 5.6, $v_n \in C_0(S)$ and $v \in C_0(S)$. Let $d_n = V^n(s_0) - V^{n-1}(s_0)$, $n \in \mathbb{N}$. Then we have

$$d_{n+1} + v_{n+1} = Uv_n, \; n \in \mathbb{N}.$$

By Theorem 5.1(b), $d := \lim_n d_n$ exists. From Theorem 5.3 it follows that for any $s \in S$ the sequence of lower semi-continuous functions $(\mu, \nu) \to Lv_n(s, \mu, \nu)$ is uniformly convergent. From Lemma 5.5 we conclude the statements (I) and (IV) since

$$\Gamma_n(s) = \{\nu \in P(B(s)) : \sup_\mu Lv_{n-1}(s, \mu, \nu) = Uv_{n-1}\},\ n \in I\!N.$$

By Lemma 5.6 there exists $g_n \in G$ such that $g_n(s) \in \Gamma_n(s)$, $s \in S$ and $n \in I\!N$. Since $P(B(s))$ is compact for $s \in S$ there exists $g \in G$ with

$$g(s) \in Ls_n\{g_n(s)\} \subset Ls_n \Gamma_n(s) \subset \Gamma(s),\ s \in S.$$

Now statements (II), (III) and (V) are a consequence of Theorem 3.2 and Lemma 5.6. □

5.2 Condition (S)

The transition probability P will be called **s-continuous** if for any $s \in S$ the function $(a, b) \to \int P(s, a, b, ds')u(s')$ is continuous for all $u \in B(S)$. Throughout this subsection we presume the following

Condition (S).

(1) A is measurable in $S \times X$ and has a Castaing representation, i.e. there is a sequence (ψ_n) of measurable maps $\psi_n : S \to X$ such that $\{\psi_1(s), \psi_2(s), \ldots\}$ is dense in $A(s)$ for $s \in S$.

(2) B is measurable in $S \times Y$ and $B(s)$ is compact, $s \in S$.

(3) P is s-continuous.

(4) $(a, b) \to r(s, a, b)$ is lower semi-continuous, $s \in S$.

Lemma 5.8. *U maps $B(S)$ into itself and the statements (i), (ii) and (iii) are valid.*

Proof. By (1) there exists a sequence (ψ'_n) of measurable maps $\psi'_n : S \to P(X)$ such that $\{\psi'_1(s), \psi'_2(s), \ldots\}$ is dense in $P(A(s))$, $s \in S$. Let $u \in B(S)$. Since $\mu \to Lu(s, \mu, \nu)$ is lower semi-continuous we may write

$$\sup_\mu Lu(s, \mu, \nu) = \sup_n Lu(s, \psi'_n(s), \nu).$$

Hence the left hand side is measurable in s and lower semi-continuous in ν. From the selection theorem of Brown and Purves (cf.[10], Corollary 4.3) we conclude that $Uu \in B(S)$ and that (iii) is valid. Statements (i) and (ii) can be proven as in Lemma 5.6. □

Theorem 5.9. Let $\delta < 1$. If condition (S) is satisfied then the statements (I), (II), (III), (IV) and (V) are valid.

The *proof* follows in the same way as the proof of Theorem 5.7 (with Lemma 5.8 instead of Lemma 5.6).

It is well-known that the ergodicity condition can be weakened if the average payoff problem is considered as a limit (as the discount factor approaches unity) of discounted Markov games (cf. [13],[8], [14]). Here we use the asymmetric condition (S), but the proof of Theorem 5.10 is similar as in [8], [14] and other papers.

Theorem 5.10. Assume $\gamma < 1$ and condition (S). Then there exists a span-fixed-point of U in $B(S)$ and the Markov game has a value ϕ ($\phi(s)$ is independent of $s \in S$). For any $\epsilon > 0$ there exists an ϵ-optimal stationary policy for player I and an optimal stationary policy for player II.

Proof. Let $s_0 \in S$. For any $\beta \in (0, 1)$ define

$$w_\beta(s) = V_\beta(s) - V_\beta(s_0), \ s \in S$$

where $s \to V_\beta(s)$ is the value function of the β-discounted Markov game. In view of condition (S), the β-discounted game has a value (cf. [12]). From Theorem 4.1 we obtain that the family $\{w_\beta; \beta \in (0,1)\}$ is uniformly bounded. Hence there exists a sequence (β_n) with $\beta_n \uparrow 1$ such that $w := \lim_n w_{\beta_n}$ exists and $w \in B(S)$. By using Lemma 5.5 and Lemma 5.8 we conclude that w is a span-fixed-point of U in $B(S)$. Now the other statements are easily derived from Theorem 3.2. □

6 The linear programming formulation

In this section we consider the linear programming (LP) formulation of the average payoff Markov game introduced in section 2. For finite state and action spaces there is an extensive literature on the LP approach (cf. e.g. the surveys

[7], [9]). We shall use some basic facts on LP in general vector spaces (cf. [1], Chapter 3, and also [5]).
At first we obtain the following characterization of the value function ϕ of the Markov game.

Theorem 6.1. Assume (S) and $\gamma < 1$. Then ϕ is the smallest **superharmonic** $c \in \mathbb{R}$, i.e. there exists $u \in B(S)$ and $g \in G$ such that

$$\begin{aligned} c + u(s) &\geq L_{fg}u(s) \text{ for all } s \in S \text{ and } f \in F \\ c &\geq \phi. \end{aligned}$$

Theorem 6.1 implies that ϕ is the unique c-part in the optimal solution of the following **mathematical program**:

$$\begin{cases} c \longrightarrow \min \\ c + u(s) \geq Lu(s, a, g(s)), (s, a) \in A \\ c \in \mathbb{R}, u \in B(S), g \in G. \end{cases}$$

In view of the constraints the mathematical program is not linear. Therefore, for the remaining part of this section we **assume** that the transitions are controlled only by player I, i.e.

$$P(s, a, b, \cdot) \text{ is independent of } b, \text{ for all } (s, a) \in A.$$

Then the program becomes the following **linear program**

$$(P)\begin{cases} c \longrightarrow \min \\ c + u(s) - \int P(s, a, ds')u(s') \geq r(s, a, g(s)), (s, a) \in A \\ c \in \mathbb{R}, u \in B(S), g \in G. \end{cases}$$

with the **dual linear program** (cf. [1], pp. 38–39)

$$(D)\begin{cases} m(S) \longrightarrow \max & \\ \lambda_1(C) = \int P(s, a, C)\lambda(d(s, a)), & C \subset S \\ m(C) \leq \int 1_C(s)r(s, a, g(s))\lambda(d(s, a)), & g \in G \text{ and } C \subset S \\ \lambda \in P(A), m \text{ finite measure on } S & \end{cases}$$

where λ_1 denotes the marginal (projection) of λ on S and the sets C are Borel sets in S.

Let $v(P)$ and $v(D)$ denote the values of the linear programs (P) and (D), respectively. By means of Theorem 5.10 we have the following propositions.

Theorem 6.2. Assume (S) and $\gamma < 1$. Then the linear program (P) has an optimal solution and $v(P) = v(D)$. Moreover,

$$\begin{aligned} v(P) &= \inf_g \sup_f \phi_{fg}(s), s \in S \\ v(D) &= \sup_f \inf_g \phi_{fg}(s), s \in S. \end{aligned}$$

Theorem 6.3. Assume (S) and $\gamma < 1$.

(a) Let (c^*, u^*, g^*) be an optimal solution of (P). Then the stationary policy $(g^*, g^*, \ldots)$ is optimal for player II, $c^* = \phi$ and u^* is a span-fixed-point of U in $B(S)$.

(b) Let (λ^*, m^*) be an optimal solution of (D). Then there exists $f^* \in F$ such that for all Borel sets C in A

$$\lambda^*(C) = \int \lambda_1^*(ds) f^*(s)(C(s))$$

and the stationary policy $(f^*, f^*, \ldots)$ is optimal for player I.

Sufficient conditions for the existence of an optimal solution of (D) are given by the following **symmetric** compactness and continuity conditions (S^*).

Condition (S^*)**.**

(1) A is measurable in $S \times X$ and $A(s)$ is compact, $s \in S$.

(2) B is measurable in $S \times Y$ and $B(s)$ is compact, $s \in S$.

(3) P is s-continuous.

(4) $a \to r(s,a,b)$ is upper semi-continuous, $(s,b) \in B$;
$b \to r(s,a,b)$ is lower semi-continuous, $(s,a) \in A$.

For a *proof* of the existence of an optimal solution of (D) use the fact that player I has an optimal stationary policy in view of condition (S^*) and Corollary 3.3.

References

[1] E.J. Anderson and P. Nash: Linear programming in infinite-dimensional spaces. Wiley, Chichester, England 1987.

[2] A. Arapostatis, V.S. Borkar, E. Fernandez-Gaucherand, M.K. Ghosh and S.I. Marcus: Discrete-time controlled Markov processes with average cost criterion: a survey. SIAM J. Control Optim. 31 (1993), 282–344.

[3] K. Fan: Minimax theorems. Proc. Nat. Acad. Sci. 39 (1953), 42–47.

[4] O. Hernández-Lerma: Adaptive Markov control processes. Springer-Verlag, New York, 1989

[5] O. Hernández-Lerma and J.B. Lasserre: Linear programming and average optimality of Markov control processes on Borel spaces-unbounded costs. SIAM J. Control and Optim. 32 (1994), 480–550.

[6] D. Isaacson and R. Madsen: Markov Chains. Wiley, 1976.

[7] L.C.M. Kallenberg: Survey of linear programming for standard and nonstandard Markovian control problems. Part I: Theory. ZOR–Mathem. Meth. Oper. Res. 40 (1994), 1–42.

[8] A. Nowak: Zero-sum average payoff stochastic games with general state space. Games and Econ. Behav. 7 (1994), 221–232.

[9] T.E.S. Raghavan and J. Filar: Algorithms for stochastic games—a survey. ZOR–Mathem. Meth. Oper. Res. 35 (1991), 437–472.

[10] U. Rieder: Measurable selection theorems for optimization problems. manuscripta math. 24 (1978), 115–131.

[11] U. Rieder: On non-discounted dynamic programming with arbitrary state space. University of Ulm, Germany, 1979.

[12] U. Rieder: Non-cooperative dynamic games with general utility functions. in Stochastic Games and Related Topics, T.E.S. Raghavan et al. (eds.), Kluwer Academic Publishers (1991), 161–174.

[13] M. Schäl: Average optimality in dynamic programming with general state space. Mathem. Oper. Res. 18 (1993), 163–172.

[14] L. Sennott: Zero-sum stochastic games with unbounded costs: Discounted and average cost cases. ZOR–Mathem. Meth. Oper. Res. 39 (1994), 209–225.