International Game Theory Review, Vol. 1, No. 1 (1999) 9–31

MARKOV STRATEGIES ARE BETTER THAN STATIONARY STRATEGIES

J. FLESCH, F. THUIJSMAN and O. J. VRIEZE
Department of Mathematics, Maastricht University,
P.O. Box 616, NL-6200 MD Maastricht, The Netherlands

We examine the use of stationary and Markov strategies in zero-sum stochastic games with finite state and action spaces. It is natural to evaluate a strategy for the maximising player, player 1, by the highest reward guaranteed to him against any strategy of the opponent. The highest rewards guaranteed by stationary strategies or by Markov strategies are called the stationary utility or the Markov utility, respectively. Since all stationary strategies are Markov strategies, the Markov utility is always larger or equal to the stationary utility. However, in all presently known subclasses of stochastic games, these utilities turn out to be equal. In this paper, we provide a colourful example in which the Markov utility is strictly larger than the stationary utility and we present several conditions under which the utilities are equal. We also show that each stochastic game has at least one initial state for which the two utilities are equal. Several examples clarify these issues.

1. Introduction

A zero-sum stochastic game Γ can be described by a state space $S := \{1, \ldots, z\}$ and a corresponding collection $\{M_1, \ldots, M_z\}$ of matrices, where matrix M_s has size $m_s^1 \times m_s^2$ for $i_s \in I_s := \{1, \ldots, m_s^1\}$ and $j_s \in J_s := \{1, \ldots, m_s^2\}$, entry (i_s, j_s) of M_s consists of a payoff $r(s, i_s, j_s) \in \mathbb{R}$ and a probability vector $p(s, i_s, j_s) = (p(t|s, i_s, j_s))_{t \in S}$. The elements of S are called states and for each state $s \in S$ the elements of I_s and J_s are called (pure) actions of player 1 and player 2 in state s. The game is to be played at stages in $\mathbb{N} = \{1, 2, 3, \ldots\}$ in the following way. The play starts at stage 1 in an initial state, say in state $s^1 \in S$, where, simultaneously and independently, both players are to choose an action: player 1 chooses an $i_{s^1}^1 \in I_{s^1}$, while player 2 chooses a $j_{s^1}^1 \in J_{s^1}$. These choices induce an immediate payoff $r(s^1, i_{s^1}^1, j_{s^1}^1)$ from player 2 to player 1. Next, the play moves to a new state according to the probability vector $p(s^1, i_{s^1}^1, j_{s^1}^1)$, say to state s^2. At stage 2, new actions $i_{s^2}^2 \in I_{s^2}$ and $j_{s^2}^2 \in J_{s^2}$ are to be chosen by the players in state s^2. Then, player 1 receives payoff $r(s^2, i_{s^2}^2, j_{s^2}^2)$ from player 2 and the play moves to some state s^3 according to the probability vector $p(s^2, i_{s^2}^2, j_{s^2}^2)$ and so on.

The sequence $h^n = (s^1, i_{s^1}^1, j_{s^1}^1; \ldots; s^n, i_{s^n}^n, j_{s^n}^n, s^{n+1})$ is called the history up to stage n. The players are assumed to have complete information and perfect recall.

A mixed action for a player in state s is a probability distribution on the set of his actions in state s. Mixed actions in state s will be denoted by x_s for player 1 and

by y_s for player 2 and the sets of mixed actions in state s by X_s and Y_s respectively. A strategy is a decision rule that prescribes a mixed action for any past history of the play. Such general strategies, so-called behaviour strategies, will be denoted by π for player 1 and by σ for player 2. We use the notations Π and Σ for the respective behaviour strategy spaces of the players. A strategy is called pure if it specifies one pure action for each possible history. We denote the respective pure strategy spaces by Π^p and Σ^p. If for all past histories, the mixed actions prescribed by a strategy depend only on the current stage and state, then the strategy is called Markov, while if they only depend on the current state, then the strategy is called stationary. Thus, the stationary strategy spaces are $X := \times_{s\in S} X_s$ for player 1 and $Y := \times_{s\in S} Y_s$ for player 2; while the Markov strategy spaces are $F := \times_{n\in\mathbb{N}} X$ for player 1 and $G := \times_{n\in\mathbb{N}} Y$ for player 2. We will use the respective notations x and y for stationary strategies and f and g for Markov strategies for players 1 and 2.

We will distinguish absorbing and non-absorbing states in the state space. A state is called absorbing, if the probability of leaving the state is zero for all available pairs of actions, otherwise the state is called non-absorbing.

Let H_s denote the set of infinite histories with initial state s. If h is an infinite history then h^n will denote the head of history, h up to stage n, while h^0 is sometimes used for the initial state. Likewise, if U_s is a non-empty subset of H_s, then U_s^n is the set of histories h^n where $h \in U_s$. As a special case, H_s^n is the set of all histories up to stage n with initial state s. For a given pair of strategies (π, σ) and an initial state s, we use the notation $H_s(\pi, \sigma)$ for the set of histories $h \in H_s$ which are consistent with (π, σ), i.e., h^n has a positive probability with respect to (π, σ) for any $n \in \mathbb{N}$.

For an infinite history $h = (s^n, i_{s^n}^n, j_{s^n}^n)_{n\in\mathbb{N}} \in H_s$, we will evaluate the sequence of payoffs by the limiting average reward, defined by

$$\gamma(h) := \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} r(s^n, i_{s^n}^n, j_{s^n}^n)\,.$$

A pair of strategies (π, σ) together with an initial state $s \in S$, by Kolmogorov's existence theorem [Kolmogorov (1933)], determines a probability measure $\mathcal{P}_{s\pi\sigma}$ on the sigma-field of subsets of H_s. For (π, σ) and initial state s, the sequences of payoffs are evaluated by the expected limiting average reward

$$\gamma(s, \pi, \sigma) := E_{s\pi\sigma}\, \gamma(\theta)\,,$$

where θ denotes the random variable for the infinite history. We will also use the vector notation

$$\gamma(\pi, \sigma) := (\gamma(s, \pi, \sigma))_{s\in S}\,.$$

Mertens and Neyman (1981) showed that

$$\sup_{\pi\in\Pi} \inf_{\sigma\in\Sigma} \gamma(s, \pi, \sigma) = \inf_{\sigma\in\Sigma} \sup_{\pi\in\Pi} \gamma(s, \pi, \sigma) =: v_s \qquad \forall s \in S\,. \tag{1.1}$$

Here $v := (v_s)_{s \in S}$ is called the limiting average value. A strategy π of player 1 is called optimal for initial state $s \in S$ if

$$\gamma(s, \pi, \sigma) \geq v_s \quad \forall \sigma \in \Sigma,$$

and ε-optimal for initial state $s \in S$, $\varepsilon > 0$, if

$$\gamma(s, \pi, \sigma) \geq v_s - \varepsilon \quad \forall \sigma \in \Sigma .$$

If a strategy of player 1 is optimal or ε-optimal for all initial states in S, then the strategy is respectively called optimal or ε-optimal. Optimality for strategies of player 2 is analogously defined. Although for all $\varepsilon > 0$, by (1.1), there exist ε-optimal strategies for both players, the famous example of Gillette (1957), the Big Match, examined by Blackwell and Ferguson (1968), demonstrated that in general the players need not have optimal strategies and that in the achieving of ε-optimality behaviour strategies are indispensable.

Against a fixed strategy of player 2, for all $\varepsilon > 0$, there always exists pure ε-best replies for player 1, i.e.,

$$\forall \sigma \in \Sigma \quad \forall \varepsilon > 0 \quad \exists \pi^{\varepsilon} \in \Pi^{p} : \quad \gamma(s, \pi^{\varepsilon}, \sigma) \geq \gamma(s, \pi, \sigma) - \varepsilon \qquad \forall s \in S, \quad \forall \pi \in \Pi . \tag{1.2}$$

Obviously, a similar statement holds for the best replies of player 2 as well.

It is in the spirit of the game to evaluate a strategy π of player 1 by the highest reward $\phi_s(\pi)$ that π guarantees when starting in state $s \in S$, so let

$$\phi_s(\pi) := \inf_{\sigma \in \Sigma} \gamma(s, \pi, \sigma), \qquad \phi(\pi) = (\phi_s(\pi))_{s \in S}.$$

For an initial state $s \in S$, we will call the highest reward that can be guaranteed by stationary strategies, the stationary utility, denoted by α_s and the highest reward that can be guaranteed by Markov strategies, the Markov utility, denoted by β_s. Formally,

$$\alpha_s := \sup_{x \in X} \phi_s(x), \quad \beta_s := \sup_{f \in F} \phi_s(f), \quad \alpha := (\alpha_s)_{s \in S}, \quad \beta := (\beta_s)_{s \in S} .$$

The fact that all stationary strategies are Markov strategies and the definition of the value yield

$$\alpha \leq \beta \leq v . \tag{1.3}$$

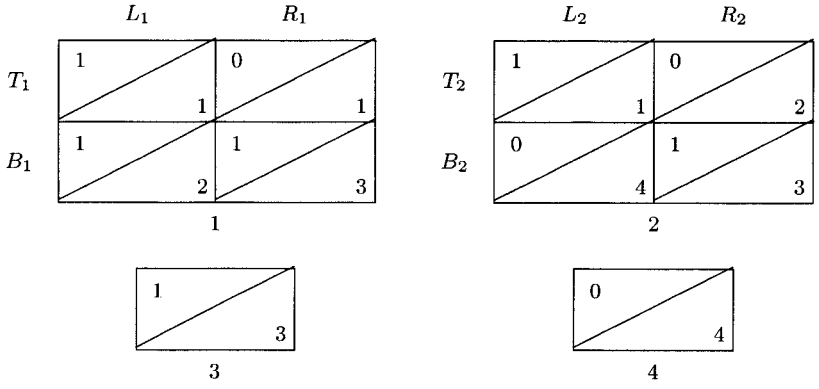
Although the class of Markov strategies is much richer than the class of stationary strategies, so far no substantial difference has been found in the use of stationary and Markov strategies in zero-sum stochastic games with finite state and action spaces. Most classes of stochastic games, examined so far, have the property that both players have stationary ε-optimal strategies for all $\varepsilon > 0$. Thus, in view of (1.3), in those classes Markov strategies do not yield higher rewards than stationary strategies. These classes are the following ones: irreducible or unichain stochastic

games [Rogers (1969) and Sobel (1971)], perfect information games [Liggett and Lippman (1969)], switching control games [Filar (1981)], ARAT games [Raghavan *et al.* (1985)], (SER-)SIT games [Parthasarathy *et al.* (1984) and Thuijsman (1992)] and recursive games [Everett (1957)]. The only class in which stationary ε-optimal strategies are not available is the class of repeated games with absorbing states. In Sec. 3, we show that the equality $\alpha = \beta$ holds for this class as well. Thus, the goal of this paper is to explore if Markov strategies can be more efficient than stationary strategies, as well as to find sufficient conditions under which these two classes of strategies are equally good.

The organisation of the paper is as follows. In the next section, a colourful example illustrates that β can be strictly larger than α for some initial states. Moreover, in this game there are ε-optimal strategies for player 1 in terms of Markov strategies, for all $\varepsilon > 0$, even though player 1 has no stationary ε-optimal strategies at all. However, in Sec. 3, we give a short proof that in each zero-sum stochastic game, there is at least one initial state s for which $\alpha_s = \beta_s$. Finally, we present several sufficient conditions for $\alpha = \beta$. After recalling some of the most important classes of games in which $\alpha = \beta$ is already known, we will show that $\alpha = \beta$ (i) for repeated games with absorbing states (stochastic games where there is only one non-absorbing state), (ii) for games where α or β is constant, (iii) for games where player 1 has an optimal or a so-called best-Markov strategy. Several examples will be given to clarify these issues.

2. An Example Where $\alpha < \beta$ for Some Initial States

This section is devoted to an example demonstrating that β may be strictly larger than α for some initial states. Consider the following game Γ:



Here player 1 chooses rows and player 2 chooses columns. In each entry, the corresponding payoff is placed in the top-left corner, while the transition is placed in the bottom-right corner. In this game, each transition is represented by the state to which transition should occur with probability 1. The interesting initial states are obviously states 1 and 2, because states 3 and 4 are absorbing.

The main result of this section is the next theorem, which will follow from Lemmas 2.1 and 2.9 below.

Theorem 2.1. *In the game* Γ, *we have* $0 = \alpha_t < \beta_t = 1 = v_t$ *for initial states* $t = 1, 2$.

This theorem states that, for initial states 1 and 2 in the game Γ, player 1 can get at most 0 by using stationary strategies, while he can get as close to 1 as he likes by using Markov strategies.

Since there are two actions for each player in both non-trivial states, we may represent each mixed action in state 1 and in state 2 by the probability assigned to the first action, which makes the stationary and Markov strategy spaces

$$X = Y = [0,1] \times [0,1], \qquad F = G = \times_{n \in \mathbb{N}} ([0,1] \times [0,1]).$$

First, we intuitively discuss the main steps of the proof. We will start with an easy proof that $\alpha_t = 0$ for initial states $t = 1, 2$ (Lemma 2.1). Since the largest payoff in the game is 1, in view of (1.3), it remains to show that $\beta_t = 1$ for initial states $t = 1, 2$. However, for this step we need to analyse the game in detail. We define a Markov strategy f^K for player 1 where $K \in \mathbb{N}$: let

$$u^K(n) := \sqrt[K]{\frac{n}{n+1}} \quad \text{for all } n \in \mathbb{N}, \quad f^K := (u^K(n), u^K(n))_{n \in \mathbb{N}} \in F.$$

Observe that the Markov strategy f^K is symmetric in the sense that the prescribed mixed actions in state 1 and state 2 are the same for any stage. Note that the sequence $u^K(n)$ converges to 1 as $n \to \infty$, so f^K assigns less and less probability to actions B_1, B_2.

We will show that, for all $\varepsilon > 0$, for initial states 1 and 2, player 1 can guarantee a reward of at least $1 - \varepsilon$ by playing the Markov strategy f^K with a large $K \in \mathbb{N}$. Now the question is how player 2 can reply to the strategy f^K. Intuitively, player 2 has two hopes to decrease player 1's reward. The first one is achieving absorption in state 4 with payoff 0, but apparently player 2 can only achieve absorption in state 4 with probability at most ε (Lemma 2.6). Player 2's best candidate would be playing actions L_1 and L_2 whenever the play is in state 1 or in state 2. But in fact, whenever the play is in state 2 a transition occurs to state 1 with a large probability and from state 1 it takes a long time (for large stages an even longer time), before the play returns to state 2 again. Thus, using that the strategy f^K assigns less and less probability to B_2, the probability of absorption in state 4 turns out to be at most ε. On the other hand, using that the payoffs in entries (T_1, R_1) and (T_2, R_2) equal 0, player 2 could try to play actions R_1 and R_2 "often enough" and hope that the play will not absorb. But in that case, it will appear that the play will eventually absorb with probability 1 (Lemma 2.7) and the zero payoffs in entries (T_1, R_1) and (T_2, R_2) will have no influence on the reward then.

First, we show that by playing stationary strategies, player 1 can get at most 0 for initial states 1 and 2.

Lemma 2.1. *$\alpha_t = 0$ for initial states $t = 1, 2$ in the game Γ.*

Proof. For each strategy $x = (x_1, x_2)$, we define a strategy $y^x = (y_1^x, y_2^x)$ for player 2. Let

$$y_1^x := \begin{cases} 1 & \text{if } x_1 < 1 \\ 0 & \text{if } x_1 = 1 \end{cases}, \qquad y_2^x := \begin{cases} 1 & \text{if } x_2 < 1 \\ 0 & \text{if } x_2 = 1 . \end{cases}$$

Notice that, for $t = 1, 2$, we have $\gamma(t, x, y^x) = 0$ for all $x \in X$, so

$$\alpha_t = \sup_{x \in X} \phi_t(x) = \sup_{x \in X} \inf_{\sigma \in \Sigma} \gamma(t, x, \sigma) \leq \sup_{x \in X} \gamma(t, x, y^x) = 0 \qquad \forall t = 1, 2.$$

Since the smallest payoff in the game is 0, the proof is complete. □

For the analysis of f^K defined above, we need two important properties about the speed of convergence when $u^K(n)$ tends to 1 as n goes to infinity. The first property says that the convergence is fast in the sense that, intuitively, for any $\varepsilon > 0$, if $K \in \mathbb{N}$ is sufficiently large then the probability of ever playing action B_1 or action B_2 at stages 2^{n-1}, $n \in \mathbb{N}$, is at most $\varepsilon/2$. However, on the other hand, the second property tells us that in a "dense" set of stages, one of bottom actions B_1 and B_2 will eventually be chosen, so the convergence of $u^K(n)$ is not too fast.

Lemma 2.2. *The sequences $(u^K(n))_{n \in \mathbb{N}}$, where $K \in \mathbb{N}$, have the following properties:*

(1) *For any $\varepsilon > 0$, if $K \in \mathbb{N}$ is sufficiently large then*

$$\prod_{n=1}^{\infty} u^K(2^{n-1}) \geq 1 - \frac{\varepsilon}{2} .$$

(2) *If $A \subset \mathbb{N}$ satisfies*

$$\omega(A) := \limsup_{N \to \infty} \frac{1}{N} \cdot \#[A \cap \{1, \ldots, N\}] > 0$$

then for any $K \in \mathbb{N}$

$$\prod_{n \in A} u^K(n) = 0 .$$

Proof.

Part 1. Let $\varepsilon > 0$. For any sequence $(w^n)_{n\in\mathbb{N}}$ in $[0,1]$, we have

$$\prod_{n\in\mathbb{N}} w^n = 1 - [(1-w^1) + w^1(1-w^2) + w^1w^2(1-w^3) + \cdots],$$

thus

$$\begin{aligned}\prod_{n=1}^{\infty} u^K(2^{n-1}) &= \prod_{n=1}^{\infty} \sqrt[K]{\frac{2^{n-1}}{2^{n-1}+1}} \\ &= \sqrt[K]{\prod_{n=1}^{\infty} \frac{2^{n-1}}{2^{n-1}+1}} \\ &= \sqrt[K]{1 - \left(\frac{1}{2} + \frac{1}{2}\frac{1}{3} + \frac{1}{2}\frac{2}{3}\frac{1}{5} + \frac{1}{2}\frac{2}{3}\frac{4}{5}\frac{1}{9} + \cdots\right)}.\end{aligned}$$

Notice that

$$d := 1 - \left(\frac{1}{2} + \frac{1}{2}\frac{1}{3} + \frac{1}{2}\frac{2}{3}\frac{1}{5} + \frac{1}{2}\frac{2}{3}\frac{4}{5}\frac{1}{9} + \cdots\right) > 1 - \left(\frac{1}{2} + \frac{1}{2}\frac{1}{2} + \frac{1}{2}\frac{1}{4} + \frac{1}{2}\frac{1}{8} + \cdots\right) = 0.$$

Since d is positive, there exists a $\bar{K} \in \mathbb{N}$ such that for all $K \geq \bar{K}$

$$\prod_{n=1}^{\infty} u^K(2^{n-1}) = \sqrt[K]{d} \geq 1 - \frac{\varepsilon}{2},$$

so the proof of the first part is complete.

Part 2. By the definition of $\omega(A)$, there exists an increasing sequence $(n_k)_{k\in\mathbb{N}}$ in A such that

$$\frac{1}{n_k} \cdot \#[A \cap \{1, \ldots, n_k\}] \geq \frac{1}{2}\omega(A) \qquad \forall k \in \mathbb{N}. \tag{2.1}$$

As $\omega(A) > 0$, by taking a subsequence, we may assume without loss of generality that

$$\frac{1}{8} n_{k+1} \cdot \omega(A) \geq n_k \qquad \forall k \in \mathbb{N}. \tag{2.2}$$

Then (2.1) and (2.2) imply

$$\begin{aligned}\#[A \cap \{n_k + 1, \ldots, n_{k+1}\}] &\geq \#[A \cap \{1, \ldots, n_{k+1}\}] - n_k \\ &\geq \frac{1}{2} n_{k+1} \cdot \omega(A) - n_k \\ &\geq \frac{1}{4} n_{k+1} \cdot \omega(A).\end{aligned}$$

Since the left-hand-side is a natural number, we obtain

$$\#\left[A \cap \{n_k+1,\ldots,n_{k+1}\}\right] \geq \left\lceil \frac{1}{4}\, n_{k+1} \cdot \omega(A) \right\rceil ,$$

where $\lceil r \rceil$ denotes $\min\{n \in \mathbb{N} | \, r \leq n\}$. Using that $u^K(n)$ is increasing in n and applying (2.2) yield

$$\begin{aligned}
\prod_{n \in A \cap \{n_k+1,\ldots,n_{k+1}\}} u^K(n) &\leq \prod_{n=n_{k+1}-\left\lceil \frac{1}{4}\, n_{k+1}\cdot\omega(A)\right\rceil+1}^{n_{k+1}} u^K(n) \\
&= \sqrt[K]{\frac{n_{k+1} - \left\lceil \frac{1}{4}\, n_{k+1} \cdot \omega(A)\right\rceil + 1}{n_{k+1}+1}} \\
&\leq \sqrt[K]{\frac{n_{k+1} - \frac{1}{4}\, n_{k+1} \cdot \omega(A) + 1}{n_{k+1}}} \\
&= \sqrt[K]{1 - \frac{1}{4}\,\omega(A) + \frac{1}{n_{k+1}}} \\
&\leq \sqrt[K]{1 - \frac{1}{8}\,\omega(A)}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\prod_{n \in A} u^K(n) &= \prod_{n \in A \cap \{1,\ldots,n_1\}} u^K(n) \cdot \prod_{k \in \mathbb{N}} \left[\prod_{n \in (A \cap \{n_k+1,\ldots,n_{k+1}\})} u^K(n) \right] \\
&\leq \prod_{k \in \mathbb{N}} \sqrt[K]{1 - \frac{1}{8}\,\omega(A)} = 0,
\end{aligned}$$

so the proof is complete. □

The next lemma says that for initial states 1 and 2, if player 2 chooses actions L_1 and L_2 whenever the play is in state 1 or in state 2, then the strategy f^K with a large $K \in \mathbb{N}$, guarantees that the frequency of visits to state 2 rapidly decreases during the play. At first sight, the reason seems to be absorption in state 4, but as it will turn out in Lemma 2.6, absorption in state 4 does not play an important role here. The reason is that the lengths of periods of stay in state 1 increase during the play, which is due to the gradually decreasing probabilities for playing B_1 in state 1.

Lemma 2.3. *Let $\varepsilon > 0$, $t \in \{1,2\}$ and let $y = (1,1) \in Y$. For a history $h \in H_t$, let $m(h)$ be the number of stages at which the play is in state 2 during h. Let*

$M(h) := \{n \in \mathbb{N} | \, n \leq m(h)\}$. *Let* $(a^n(h))_{n \in M(h)}$ *denote the sequence of stages when state* 2 *is visited. Then, for large* $K \in \mathbb{N}$,

$$\mathcal{P}_{tf^K y}(a^n(\theta) \geq 2^{n-1} \quad \forall n \in M(\theta)) \geq 1 - \frac{\varepsilon}{2},$$

where θ *denotes the random variable for the infinite history.*

Proof. We only show the statement for initial state 2, for initial state 1 a similar proof can be given. So suppose that the initial state is state 2. Then, notice that $a^1(h) = 1$, $m(h) \geq 1$ and $M(h) \neq \emptyset$ for all $h \in H_2$ (similarly, for initial state 1, if $M(h) \neq \emptyset$ then $a^1(h) \geq 2$, which would only slightly modify the rest of the proof). For all $h \in H_2$, whenever $m(h) < \infty$, we define inductively

$$a^n(h) := \max\{2^{n-1}, 8a^{n-1}(h)\} \quad \forall n = m(h)+1, m(h)+2, \ldots \tag{2.3}$$

In view of (2.3), we have to show that for large $K \in \mathbb{N}$,

$$\mathcal{P}_{2f^K y}(a^n(\theta) \geq 2^{n-1} \quad \forall n \in \mathbb{N}) \geq 1 - \frac{\varepsilon}{2}. \tag{2.4}$$

Let $\eta^0 := 0$ and for $n \in \mathbb{N}$ let

$$\eta^n(h) := \begin{cases} \eta^{n-1}(h) + 1 & \text{if } a^{n+1}(h) \geq 8a^n(h) \\ \eta^{n-1}(h) - 1 & \text{otherwise.} \end{cases}$$

Observe that if the play is in state 2 at stage w and $w \in M(h)$ for the history h, then the probability with respect to $(2, f^K, y)$ that the play does not return to state 2 before stage $8w$, is at least the probability that the play moves to state 1 and it stays there till stage $8w-1$; so at least

$$u^K(w) \cdot \prod_{n=w+1}^{8w-2} u^K(n) = \prod_{n=w}^{8w-2} u^K(n).$$

Hence, for any $w, k \in \mathbb{N}$, if $\mathcal{P}_{2f^K y}(a^k(\theta) = w, \; k \in M(\theta)) > 0$, then

$$\mathcal{P}_{2f^K y}(a^{k+1}(\theta) \geq 8a^k(\theta) | \, a^k = w, \; k \in M(\theta)) \geq \prod_{n=w}^{8w-2} u^K(n). \tag{2.5}$$

On the other hand, if $\mathcal{P}_{2f^K y}(a^k(\theta) = w, \; k \notin M(\theta)) > 0$, then by (2.3), we have

$$\mathcal{P}_{2f^K y}(a^{k+1}(\theta) \geq 8a^k(\theta) | \, a^k = w, \; k \notin M(\theta)) = 1. \tag{2.6}$$

Therefore, for all $w \in \mathbb{N}$ and for all $K \in \mathbb{N}$ satisfying $\mathcal{P}_{2f^K y}(a^k(\theta) = w) > 0$, by (2.5) and (2.6) we have

$$\begin{aligned}\mathcal{P}_{2f^K y}(a^{k+1}(\theta) \geq 8a^k(\theta) \,|\, a^k(\theta) = w) &\geq \prod_{n=w}^{8w-2} u^K(n) \\ &= \sqrt[K]{\frac{w}{(8w-2)+1}} \\ &= \sqrt[K]{\frac{w}{8w-1}} \\ &\geq \sqrt[K]{\frac{1}{8}}. \end{aligned} \tag{2.7}$$

We now show that for large $K \in \mathbb{N}$

$$\mathcal{P}_{2f^K y}(\eta^n(\theta) \geq 1 \quad \forall n \in \mathbb{N}) \geq 1 - \frac{\varepsilon}{2}. \tag{2.8}$$

For simplicity, let $\xi^K := \sqrt[K]{\frac{1}{8}}$. On the set of integers, for any $K \in \mathbb{N}$, we define a birth and death process $\bar{\eta}_K^n$, $n = 0, 1, 2, \ldots$, as follows. Let $\bar{\eta}_K^0 := 0$ and for $n \in \mathbb{N}$ let

$$\bar{\eta}_K^n := \begin{cases} \bar{\eta}_K^{n-1} + 1 & \text{with probability } \xi^K \\ \bar{\eta}_K^{n-1} - 1 & \text{with probability } 1 - \xi^K. \end{cases}$$

Since ξ^K converges to 1 as K tends to infinity, for the birth and death process $\bar{\eta}_K^n$, $n = 0, 1, 2, \ldots$, we clearly have that for large $K \in \mathbb{N}$

$$\mathcal{P}(\bar{\eta}_K^n \geq 1 \quad \forall n \in \mathbb{N}) \geq 1 - \frac{\varepsilon}{2}.$$

Hence, by the definitions of η^n and $\bar{\eta}_K^n$ for $n = 0, 1, 2, \ldots$ and by (2.7), we have for large $K \in \mathbb{N}$ that

$$\mathcal{P}_{2f^K y}(\eta^n(\theta) \geq 1 \ \forall n \in \mathbb{N}) \ \geq \ \mathcal{P}(\bar{\eta}_K^n \geq 1 \ \forall n \in \mathbb{N}) \ \geq \ 1 - \frac{\varepsilon}{2},$$

which completes the proof of (2.8).

Let $\nu^0 := 0$ and let $\nu^n(h)$ be the number of jumps with $+1$ in $\eta^0(h), \eta^1(h), \ldots, \eta^n(h)$. Since for all $n \in \mathbb{N}$,

$$\eta^n(h) = (+1) \cdot \nu^n(h) + (-1) \cdot (n - \nu^n(h)) = 2\nu^n(h) - n,$$

(2.8) implies

$$\mathcal{P}_{2f^K y}\left(\nu^n(\theta) \geq \frac{n+1}{2} \quad \forall n \in \mathbb{N}\right) \geq 1 - \frac{\varepsilon}{2}. \tag{2.9}$$

Recall that $a^1(h) = 1$ for all $h \in H_2$ and notice that if $\nu^n(h) \geq \frac{n+1}{2}$ for some $n \in \mathbb{N}$, $h \in H_2$, then

$$a^n(h) \geq 8^{\nu^n(h)-1} \geq 2^{3(\frac{n+1}{2})-3} \geq 2^{n-1},$$

hence (2.9) implies (2.4), which completes the proof. □

The next lemma, which is not specific for this game Γ at all, provides useful lower- and upper-bounds for the probability that the infinite history belongs to a set V_t of infinite histories, on the condition that it belongs to some other set U_t. For the rather technical proof of this lemma, we refer to Flesch *et al.* (1997b).

Lemma 2.4. *Let $t \in S$, $\pi \in \Pi$ and $\sigma \in \Sigma$. Let $V_t, U_t \subset H_t(\pi, \sigma)$ such that $\emptyset \neq V_t \subset U_t$. Assume that*

$$\mathcal{P}_{t\pi\sigma}(\theta \in U_t) > 0,$$

where θ denotes the random variable for the infinite history. Then

$$\inf_{h \in V_t} Z_{t\pi\sigma, V_t|U_t}(h) \leq \mathcal{P}_{t\pi\sigma}(\theta \in V_t | \, \theta \in U_t) \leq \sup_{h \in V_t} Z_{t\pi\sigma, V_t|U_t}(h),$$

where

$$Z_{t\pi\sigma, V_t|U_t}(h) := \prod_{k=0}^{\infty} \mathcal{P}_{t\pi\sigma}(\theta^{k+1} \in V_t^{k+1} | \, \theta^k = h^k, \, \theta \in U_t) \qquad \forall h \in V_t.$$

The next lemma, which will follow from the previous result, intuitively states that the set of infinite histories in which absorption should occur with probability 1, but in which no absorption does occur, has probability zero.

Lemma 2.5. *Let $t \in \{1, 2\}$, $K \in \mathbb{N}$, $\sigma \in \Sigma^p$. Let*

$$\tilde{H}_t := \{h \in H_t(f^K, \sigma) | \text{ no absorption occurs in } h\}$$

$$\bar{H}_t := \left\{ h \in \tilde{H}_t \middle| \prod_{n \in C(h)} u^K(n) = 0 \right\},$$

where $C(h)$ is the set of stages n, when player 2 plays actions R_1, R_2, or L_2 after history h^{n-1}, according to the pure strategy σ. Let θ denote the random variable for the infinite history. Then,

$$\mathcal{P}_{tf^K\sigma}(\theta \in \bar{H}_t) = 0.$$

Proof. Let $Z_{tf^K\sigma,\bar{H}_t|H_t(f^K,\sigma)}$ be defined as in Lemma 2.4. By the definition of $\bar{H}_t$, we have for all $h \in \bar{H}_t$

$$Z_{tf^K\sigma,\bar{H}_t|H_t(f^K,\sigma)}(h) = \prod_{k=0}^{\infty} \mathcal{P}_{tf^K\sigma}(\theta^{k+1} \in \bar{H}_t^{k+1} | \, \theta^k = h^k)$$

$$\leq \prod_{k=0}^{\infty} \mathcal{P}_{tf^K\sigma}(\theta^{k+1} \in \tilde{H}_s^{k+1} | \, \theta^k = h^k)$$

$$= \prod_{n \in C(h)} u^K(n) = 0,$$

hence, Lemma 2.4 yields

$$\mathcal{P}_{tf^K\sigma}(\theta \in \bar{H}_t) = \mathcal{P}_{tf^K\sigma}(\theta \in \bar{H}_t | \, \theta \in H_t(f^K,\sigma)) \leq \sup_{h \in \bar{H}_t} Z_{tf^K\sigma,\bar{H}_t|H_t(f^K,\sigma)}(h) \leq 0,$$

which completes the proof. □

It turns out that the strategy f^K, with a large $K \in \mathbb{N}$, keeps the probability of absorption in state 4 small. In fact, the absorption probability in state 4 is maximal when player 2 always chooses actions L_1 and L_2 whenever the play is in state 1 or in state 2, but even then, in view of Lemma 2.3, the play does not visit state 2 "frequently enough", so using that f^K assigns less and less probability to action B_2, the probability of absorption in state 4 turns out to be small indeed.

Lemma 2.6. *Let $\varepsilon > 0$. If $K \in \mathbb{N}$ is sufficiently large, then for initial states $t = 1, 2$ in the game Γ, the probability of absorption in state 4 is at most ε with respect to (t, f^K, σ) for any $\sigma \in \Sigma$.*

Proof. It is easy to see that the stationary strategy $y = (1, 1)$ maximises the probability of absorption in state 4 against f^K with any $K \in \mathbb{N}$. Therefore, it is sufficient to show the statement for y.

We only show the statement for initial state 2. Then for initial state 1, the statement becomes immediate, since from that stage on when the play moves to state 2, the strategy f^K assigns even less probabilities to actions B_1 and B_2 than when starting from stage 1. So, assume the initial state to be state 2.

Let

$$\tilde{H}_2 := \{h \in H_2(f^K, y) | \text{ no absorption occurs in } h\}$$

$$\hat{H}_2 := \{h \in H_2(f^K, y) | \; a^n(h) \geq 2^{n-1} \quad \forall n \in M(h)\},$$

where $a^n(h)$ and $M(h)$ are defined as in Lemma 2.3. Observe that for large $K \in \mathbb{N}$, by Lemma 2.3, we have

$$\mathcal{P}_{2f^K y}(\theta \in \hat{H}_2) \geq 1 - \frac{\varepsilon}{2}. \tag{2.10}$$

Now, for $h \in \tilde{H}_2 \cap \hat{H}_2$, let $Z_{2f^K y, \tilde{H}_2 \cap \hat{H}_2 | \hat{H}_2}(h)$ be defined as in Lemma 2.4. By, using Lemma 2.2, if $K \in \mathbb{N}$ is sufficiently large, then for any $h \in \tilde{H}_2 \cap \hat{H}_2$

$$\begin{aligned}
Z_{2f^K y, \tilde{H}_2 \cap \hat{H}_2 | \hat{H}_2}(h) &= \prod_{k=0}^{\infty} \mathcal{P}_{2f^K y}(\theta^{k+1} \in \tilde{H}_2^{k+1} \cap \hat{H}_2^{k+1} | \, \theta^k = h^k, \ \theta \in \hat{H}_2) \\
&= \prod_{k=0}^{\infty} \mathcal{P}_{2f^K y}(\theta^{k+1} \in \tilde{H}_2^{k+1} | \, \theta^k = h^k, \ \theta \in \hat{H}_2) \\
&= \prod_{n \in M(h)} u^K(a^n(h)) \\
&\geq \prod_{n \in M(h)} u^K(2^{n-1}) \\
&\geq \prod_{n=1}^{\infty} u^K(2^{n-1}) \\
&\geq 1 - \frac{\varepsilon}{2} .
\end{aligned}$$

Hence, by applying (2.10) and Lemma 2.4 for large $K \in \mathbb{N}$, we get

$$\begin{aligned}
\mathcal{P}_{2f^K y}(\theta \in \tilde{H}_2) &\geq \mathcal{P}_{2f^K y}(\theta \in \tilde{H}_2 \cap \hat{H}_2) \\
&= \mathcal{P}_{2f^K y}(\theta \in \tilde{H}_2 \cap \hat{H}_2 | \, \theta \in \hat{H}_2) \cdot \mathcal{P}_{2f^K y}(\theta \in \hat{H}_2) \\
&\geq \mathcal{P}_{2f^K y}(\theta \in \tilde{H}_2 \cap \hat{H}_2 | \, \theta \in \hat{H}_2) \cdot \left(1 - \frac{\varepsilon}{2}\right) \\
&\geq \mathcal{P}_{2f^K y}(\theta \in \tilde{H}_2 \cap \hat{H}_2 | \, \theta \in \hat{H}_2) - \frac{\varepsilon}{2} \\
&\geq \inf_{h \in \tilde{H}_2 \cap \hat{H}_2} Z_{2f^K y, \tilde{H}_2 \cap \hat{H}_2 | \hat{H}_2}(h) - \frac{\varepsilon}{2} \\
&\geq 1 - \frac{\varepsilon}{2} - \frac{\varepsilon}{2} \\
&= 1 - \varepsilon ,
\end{aligned}$$

which means that if $K \in \mathbb{N}$ is large, then with respect to $(2, f^K, y)$, the probability of absorption in state 4 is at most ε. □

Now, we show that when player 1 uses f^K with any $K \in \mathbb{N}$ for initial states 1 or 2 and if player 2 chooses actions R_1 and R_2 "too frequently", absorption occurs with probability 1.

Lemma 2.7. *Let* $t \in \{1, 2\}$, $K \in \mathbb{N}$, $\sigma \in \Sigma^p$. *Let*

$$\tilde{H}_t := \{h \in H_t(f^K, \sigma) | \text{ no absorption occurs in } h\} .$$

For $A \subset \mathbb{N}$, let

$$\omega(A) := \limsup_{N\to\infty} \frac{1}{N} \cdot \#\left[A \cap \{1,\ldots,N\}\right].$$

For a history $h \in H_t$, let $A(h)$ denote the set of stages n, when player 2 chooses actions R_1 or R_2 after history h^{n-1}, according to the pure strategy σ. If

$$\mathcal{P}_{tf^K\sigma}(\theta \in \tilde{H}_t) > 0\,,$$

then

$$\mathcal{P}_{tf^K\sigma}(\omega(A(\theta)) = 0 | \, \theta \in \tilde{H}_t) = 1\,,$$

where θ denotes the random variable for the infinite history.

Proof. Suppose that $\omega(A(h)) > 0$ for some history $h \in H_t$. Then clearly, no absorption occurs in h, thus $h \in \tilde{H}_t$. By Lemma 2.2, we have

$$\prod_{n\in A(h)} u^K(n) = 0\,,$$

therefore

$$\{h \in H_t(f^K,\sigma) | \, \omega(A(h)) > 0\} \subset \bar{H}_t := \left\{ h \in \tilde{H}_t \middle| \prod_{n\in A(h)\cup B(h)} u^K(n) = 0 \right\},$$

where $B(h)$ is the set of stages n when player 2 plays action L_2 after history h^{n-1}, according to the pure strategy σ. Now, Lemma 2.5 yields

$$\mathcal{P}_{tf^K\sigma}(\omega(A(\theta)) > 0) \leq \mathcal{P}_{tf^K\sigma}(\theta \in \bar{H}_t) = 0\,,$$

which implies the statement. □

The next result tells us that when player 1 uses f^K with any $K \in \mathbb{N}$ for initial states 1 and 2 and given that no absorption occurs (and this has a positive probability), the reward equals 1 almost surely.

Lemma 2.8. *Let $t \in \{1,2\}$, $K \in \mathbb{N}$ and $\sigma \in \Sigma^p$. Let*

$$\tilde{H}_t := \{h \in H_t(f^K,\sigma) | \text{ no absorption occurs in } h\}\,.$$

If

$$\mathcal{P}_{tf^K\sigma}(\theta \in \tilde{H}_t) > 0\,,$$

then

$$\mathcal{P}_{tf^K\sigma}(\gamma(\theta) = 1 | \, \theta \in \tilde{H}_t) = 1,$$

where θ denotes the random variable for the infinite history.

Proof. Let $\omega(A)$ for $A \subset \mathbb{N}$ and $A(h)$ be defined as in Lemma 2.7. Let $R_k(h)$ denote the payoff at stage k according to the history h. Then for any $h \in \tilde{H}_t$,

$$\begin{aligned}
\gamma(h) &= \lim_{n\to\infty} \inf_{N\geq n} \frac{\sum_{n=1}^{N} R_k(n)}{N} \\
&= \lim_{n\to\infty} \inf_{N\geq n} \frac{\#\{k \in \{1,\ldots,N\} | \, R_k(h) = 1\}}{N} \\
&= \lim_{n\to\infty} \inf_{N\geq n} \frac{N - \#\{k \in \{1,\ldots,N\} | \, R_k(h) = 0\}}{N} \\
&= \lim_{n\to\infty} \inf_{N\geq n} \frac{N - \#[A(h) \cap \{1,\ldots,N\}]}{N} \\
&= 1 + \lim_{n\to\infty} \inf_{N\geq n} \frac{-\#[A(h) \cap \{1,\ldots,N\}]}{N} \\
&= 1 - \lim_{n\to\infty} \sup_{N\geq n} \frac{\#[A(h) \cap \{1,\ldots,N\}]}{N} \\
&= 1 - \omega(A(h)),
\end{aligned}$$

hence, Lemma 2.7 implies the result. □

Now, we are ready to prove that $\beta_t = 1$ for initial states $t = 1, 2$ and also that the Markov strategy f^K is ε-optimal for large $K \in \mathbb{N}$. More specifically, K can be any number that satisfies Lemma 2.6.

Lemma 2.9. *For all $t = 1, 2$ in the game Γ, we have $\beta_t = v_t = 1$ and also for any $\varepsilon > 0$ if $K \in \mathbb{N}$ is sufficiently large, then $\phi_t(f^K) \geq 1 - \varepsilon$.*

Proof. Let $t \in \{1, 2\}$ and let $\varepsilon > 0$. We only need to show that $\phi_t(f^K) \geq 1 - \varepsilon$ for large $K \in \mathbb{N}$, because then $\beta_t = v_t = 1$ follows from (1.3) and from the fact that the largest payoff in the game is 1. Let θ denote the random variable for the infinite history. By Lemma 2.8, we have for any $K \in \mathbb{N}$ and for any $\sigma \in \Sigma^p$ that

$$\begin{aligned}
\gamma(t, f^K, \sigma) &= E_{t f^K \sigma}\, \gamma(\theta) \\
&= E_{t f^K \sigma} \begin{cases} 0 & \text{if absorption occurs in state 4 in } \theta \\ 1 & \text{otherwise.} \end{cases}
\end{aligned}$$

This intuitively means that player 2 can only try to maximise the probability of absorption in state 4. Take $K \in \mathbb{N}$ as in Lemma 2.6. Then, the probability of absorption in state 4 is at most ε with respect to (t, f^K, σ), for any $\sigma \in \Sigma^p$, hence

$$\gamma(t, f^K, \sigma) \geq 1 - \varepsilon \qquad \forall \sigma \in \Sigma^p .$$

Since, in view of (1.2), it suffices to consider pure replies from player 2, we obtain

$$\phi_t(f^K) \geq 1 - \varepsilon\,,$$

which completes the proof. □

3. Sufficient Conditions for $\alpha = \beta$

The example of the previous section demonstrated that β may be strictly larger than α for some initial states. However, this cannot hold for all initial states, as stated in the next theorem.

Theorem 3.1. *In every zero-sum stochastic game, $\alpha_s = \beta_s\,(= v_s)$ for all states $s \in S^{\min} := \{t \in S |\, v_t = \min_{w \in S}\, v_w\}$.*

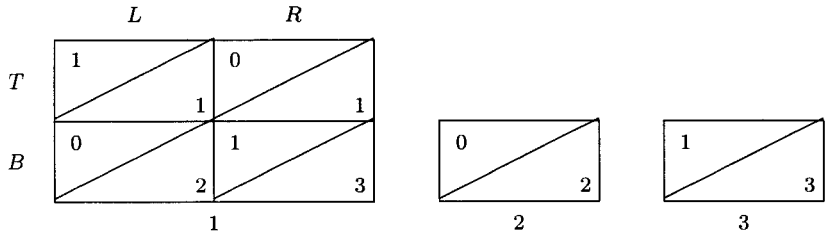
Proof. Let $s \in S^{\min}$. Then by the results of Thuijsman and Vrieze (1993), for any $\varepsilon > 0$, player 1 has a stationary ε-optimal strategy x^ε for initial state s. Hence, $\alpha_s = v_s$, thus (1.3) yields $\alpha_s = \beta_s\,(= v_s)$. □

We presented an example in Sec. 2 where α is smaller than β for some initial states. We also know from the previous section that α equals β for at least one initial state in every zero-sum game. Now, the question is what conditions would guarantee that α equals β for all initial states. We will present several sufficient conditions, however, first we would like to recall some of the more important classes of zero-sum games in which $\alpha = \beta$ is already known.

Clearly, we have $\alpha = \beta(= v)$ for any class of games where player 1 has stationary ε-optimal strategies, for all $\varepsilon > 0$. The existence of stationary (ε-)optimal strategies is known, for example in irreducible or unichain stochastic games [Rogers (1969) and Sobel (1971)], in perfect information games [Liggett and Lippman (1969)], in games with switching control [Filar (1981)], in games with additively decomposable reward and transition structure [ARAT, Raghavan *et al.* (1985)], in (SER-)SIT games [Parthasarathy *et al.* (1984) and Thuijsman (1992)], or in recursive games [Everett (1957)]. Moreover, the condition that the value is constant ($v_s = v_t$ for all $s, t \in S$) is also sufficient for the existence of stationary ε-optimal strategies [Thuijsman and Vrieze (1993), Theorem 3.1]. In such games, both players actually have Markov optimal strategies as well.

3.1. *Repeated games with absorbing states*

Repeated games with absorbing states are stochastic games where there is only one non-absorbing state. Kohlberg (1974) showed that these games have a value. However, to achieve this value, history dependent strategies are indispensable. We will show for these games that $\alpha = \beta$. First, consider the following example, known as the Big Match [Gillette (1957), Blackwell and Ferguson (1968)]:



We use the same notation as in the example in Sec. 2. Each mixed action in state 1 and each stationary strategy can be represented by the probability assigned to the first action in state 1 (T and L respectively). Thus, the stationary and Markov strategy spaces have the following form:

$$X = Y = [0,1], \qquad F = G = \times_{n\in\mathbb{N}} [0,1]\,.$$

We simply use I and J for the respective action spaces in state 1. Since state 2 and state 3 are absorbing, the only interesting initial state is state 1.

It is easy to check that $\alpha_1 = 0$. For $x \in X$, let

$$y^x := \begin{cases} 1 & \text{if } x < 1 \\ 0 & \text{if } x = 1\,. \end{cases}$$

We have $\gamma(1, x, y^x) = 0$ for all $x \in X$, thus, as the lowest payoff is 0, we obtain $\alpha_1 = 0$.

Now, we show that $\beta_1 = 0$ holds as well by showing that $\phi_1(f) = 0$ for all Markov strategies f of player 1. Take a Markov strategy $f \in F$. Let ξ denote the random variable for the stage when absorption occurs. If no absorption occurs at all, then let $\xi = 0$. Let g_1 be the strategy for player 2 which prescribes action L for each stage. For $N \in \mathbb{N}$, let

$$p^N := \mathcal{P}_{1fg_1}(\xi > N)\,,$$

so p^N is the probability that the play will absorb after stage N with respect to $(1, f, g_1)$. Take an arbitrary $\varepsilon > 0$. Since p^N converges to 0 as N tends to infinity, there exists a stage $\bar{N}$ with $p^{\bar{N}} \leq \varepsilon$. Let g_2 be the strategy that prescribes action L up to stage $\bar{N}$ and action R afterwards. Since

$$p^N = \mathcal{P}_{1fg_1}(\xi > N) = \mathcal{P}_{1fg_2}(\xi > N) \qquad \forall N \in \mathbb{N}\,,$$

the probability of absorption in state 3, with respect to $(1, f, g_2)$, is at most ε, thus $\gamma(1, f, g_2) \leq \varepsilon$. Because ε was arbitrary, we obtain $\phi_1(f) = 0$, hence $\beta_1 = 0$.

Therefore, we have shown that $\alpha = \beta$ in the Big Match. In fact, this argument can be generalised to all repeated games with absorbing states. We will not discuss all the technical details in the proof, but only give a brief sketch. In fact, the result also follows from Coulomb (1992).

Theorem 3.2. *In every zero-sum repeated game with absorbing states,* $\alpha = \beta$.

Proof. Take a zero-sum repeated game with absorbing states. We may suppose without loss of generality that, in each absorbing state, both players have only one action (otherwise we may replace the state by another absorbing state containing only the value of the corresponding one-shot game as payoff). Suppose that state 1 is the non-absorbing state. We assume that state 1 is the initial state; for the sake of simplicity we sometimes suppress state 1 in the notations. Any action of player 1 or player 2 in state 1 will also denote the stationary strategy which prescribes this action for each stage. It is well known that, against any stationary strategy $x \in X$ of player 1, there exists a best reply $j^x \in J$ [Hordijk *et al.* (1983)]. Hence, we have $\gamma(x, j^x) \leq \alpha$. This means that, for initial state 1, either (x, j^x) is absorbing and the expected absorption payoff is at most α, or (x, j^x) is non-absorbing and the expected one-shot payoff $r(1, x, j^x)$ is at most α (a stationary strategy pair (x, y) is called absorbing if $p(1|1, x, y) < 1$).

Let $\varepsilon > 0$. Take an arbitrary Markov strategy $f = (x^n)_{n\in\mathbb{N}} \in F$. It suffices to show that there exists a Markov strategy $g \in G$ such that $\gamma(f, g) \leq \alpha + \varepsilon$.

Step 1. Let $f_1 := f$ and $x_1^n := x^n$ for all $n \in \mathbb{N}$. Let $g_1 = (j^{x_1^n})_{n\in\mathbb{N}}$. Let ξ denote the random variable for the stage when absorption occurs. If no absorption occurs at all, then let $\xi = 0$. For $N \in \mathbb{N}$, let

$$p_1^N := \mathcal{P}_{f_1 g_1}(\xi > N)\,,$$

so p_1^N is the absorption probability after stage N with respect to (f_1, g_1). Since p_1^N converges to 0, there exists, for some small $\delta > 0$, a stage N_1 such that $p_1^{N_1} \leq p^* \cdot \delta$, where p^* is the smallest positive absorption probability in state 1:

$$p^* := \min_{i\in I,\, j\in J} \{p_{ij}^* | \, p_{ij}^* := 1 - p(1|\,1, i, j) \text{ and } p_{ij}^* > 0\}\,;$$

(we may assume that there exist an $i \in I$ and $j \in J$ such that $p(1|1, i, j) < 1$, otherwise the game is trivial).

If $p_1^{N_1} = 0$, then we have $\gamma(f, g_1) = \gamma(f_1, g_1) \leq \alpha$, because with respect to (f_1, g_1), the expected absorption payoff is at most α at each stage $n \leq N_1$; the probability of absorption after stage N_1 is zero; and the expected one-shot payoff is at most α at each stage $n > N_1$.

Assume now that $p_1^{N_1} > 0$. By the definition of N_1, the probability of absorption after stage N_1 for (f_1, g_1) is at most $p^* \cdot \delta$. Now, let $I_1^n := \{i \in I | \, (i, j^{x_1^n})$ be non-absorbing$\}$. Thus, the probability that, with respect to (f_1, g_1), player 1 will ever choose an action outside I_1^n at stages $n > N_1$ is at most δ.

Step 2. Let $x_2^n := x_1^n$ for $n \leq N_1$ and let x_2^n be the normalisation of x_1^n on I_1^n for $n > N_1$:

$$x_2^n(i) := \frac{x_1^n(i)}{\sum_{i\in I_1^n} x_1^n(i)} \quad \text{for all } i \in I_1^n, \qquad x_2^n(i) := 0 \quad \text{for all } i \in I \setminus I_1^n\,.$$

Let $f_2 := (x_2^n)_{n\in\mathbb{N}}$. Intuitively, f_2 coincides with f_1 up to stage N_1 and after stage N_1, the strategy f_2 equals the strategy f_1 on condition that no action outside I_1^n will ever be chosen at stages $n > N_1$. Let $g_2 := (j^{x_2^n})_{n\in\mathbb{N}}$, so by the definitions, g_1 and g_2 are the same for the first N_1 stages. One can show, using the properties of the construction, that with respect to (f, g_2), the probability of absorption outside I_1^n at stages $n > N_1$ is at most δ. Choose an $N_2 > N_1$ such that

$$p_2^{N_2} := \mathcal{P}_{f_2 g_2}(\xi > N_2) \leq \delta \cdot p^*.$$

Assume first that $p_2^{N_2} = 0$. Then we have $\gamma(f, g_2) \leq \alpha + \varepsilon$ for small δ, because with respect to (f, g_2), the expected absorption payoff at each stage in $n \leq N_1$ is at most α; the probability of absorption outside I_1^n at stages $n = N_1 + 1, \ldots, N_2$ is at most δ; the expected absorption payoff in I_1^n at each stage in $n = N_1 + 1, \ldots, N_2$ is at most α; the probability of absorption after stage N_2 is zero; and the expected one-shot payoff at each stage in $n > N_2$ is at most α.

Assume now that $p_2^{N_2} > 0$. Let $I_2^n := \{i \in I \,|\, (i, j^{x_2^n}) \text{ be non-absorbing}\}$ and repeat the above steps, in such a way that $N_{k+1} > N_k$ for all k, until at some step K we have $p_K^{N_K} = 0$. This results in a strategy g_K for player 2. Note that for $p_K^{N_K} = 0$ it is sufficient that $I_K^n = I_{K-1}^n$ holds for all $n > N_K$. Hence, we only need at most $K \leq \#I$ steps, because for any stage $n > N_k$, either I_{k+1}^n becomes smaller than I_k^n, or $I_k^n = I_{k+1}^n$ and then nothing changes at further steps for stage n. Using similar arguments as before, one can now show that $p_K^{N_K} = 0$ implies that $\gamma(f, g_K) \leq \alpha + \varepsilon$ if $\delta > 0$ is small enough. □

3.2. *Games with constant α or β*

In this section, we show that $\alpha = \beta$ in games where α or β is constant. We will need the following result.

Theorem 3.3. *In every zero-sum stochastic game,*

$$\min_{s\in S} \alpha_s = \min_{s\in S} \beta_s = \min_{s\in S} v_s, \qquad \max_{s\in S} \alpha_s = \max_{s\in S} \beta_s = \max_{s\in S} v_s\,.$$

Proof. It is known that, for any $\varepsilon > 0$, player 1 has a stationary strategy x^ε satisfying

$$\phi_t(x^\varepsilon) \geq \min_{s\in S} v_s - \varepsilon \qquad \forall t \in S;$$

[for example Flesch *et al.* (1996)]. Hence,

$$\min_{s\in S} \alpha_s \geq \min_{s\in S} v_s\,,$$

which in view of (1.3) implies the first part of the statement.

By the results of Thuijsman and Vrieze (1991), there is always a state t in $S^{\max} := \{s \in S | v_s = \max_{w \in S} v_w\}$ for which player 1 has a stationary optimal strategy x. Hence,

$$\max_{s \in S} \alpha_s \geq \alpha_t \geq \phi_t(x) \geq v_t = \max_{s \in S} v_s ,$$

thus (1.3) yields the second part of the statement. □

We have already discussed that if v is constant then $\alpha = \beta (= v)$ is also constant. Hence, the above theorem has the following corollary.

Corollary 3.1. *In every zero-sum stochastic game where either α or β is constant, $\alpha = \beta (= v)$ is constant.*

Notice that, in view of this corollary, α, β, or v is constant if and only if each of them is constant and they are equal. The following theorem provides a more relaxed view on constant values.

Theorem 3.4. *In every zero-sum stochastic game where for all $s, t \in S$, either $\alpha_s = \alpha_t$ or $\beta_s = \beta_t$, we have that $\alpha = \beta (= v)$ is constant.*

Proof. Using the inequality $\alpha \leq \beta$ and Theorem 3.3, it is clear that if state s has the property that $\beta_s = \min_{w \in S} \beta_w$, then $\alpha_s = \min_{w \in S} \alpha_w$. Similarly, if state t has the property that $\alpha_t = \max_{w \in S} \alpha_w$, then $\beta_t = \max_{w \in S} \beta_w$. By this condition, we have either $\alpha_s = \alpha_t$ or $\beta_s = \beta_t$. Therefore, by Theorem 3.3, either α or β is constant and Corollary 3.1 completes the proof. □

An interesting equivalent formulation of Theorem 3.4 is the following: if $\alpha \neq \beta$, then there must exist two states s and t such that $\alpha_s \neq \alpha_t$ and $\beta_s \neq \beta_t$.

3.3. *Games with optimal strategies or best-Markov strategies*

A best-Markov strategy means a Markov strategy f with the property that $\phi(f) \geq \phi(\bar{f})$ for all $\bar{f} \in F$, or equivalently $\phi(f) = \beta$. Optimal and best-Markov strategies do not necessarily exist, but if they do then their existence surprisingly implies $\alpha = \beta$, as stated in the next theorem.

Theorem 3.5. *In every zero-sum stochastic game, if player 1 has an optimal strategy or a best-Markov strategy, then $\alpha = \beta$.*

Proof. Suppose first that player 1 has an optimal strategy. Then by Flesch *et al.* (1997a), player 1 has stationary ε-optimal strategies for all $\varepsilon > 0$ as well. Hence, $\alpha = v$, so (1.3) yields the result.

Assume now that player 1 has a best-Markov strategy f, so $\phi(f) = \beta$. Using the results on so-called non-improving strategies in Flesch *et al.* (1997a), for all $\varepsilon > 0$, player 1 has stationary strategies guaranteeing $\phi_s(f) - \varepsilon = \beta_s - \varepsilon$ for all initial states s. Hence, $\alpha = \beta$ in this case as well. □

Note that in the game presented in Sec. 2 player 1 has neither optimal nor best-Markov strategies for initial states 1 and 2. We only show it for initial state 2. One can argue as follows. Since $\beta_2 = v_2 = 1$ in that game, it suffices to show that player 1 has no strategy guaranteeing 1 for initial state 2. Assume by way of contradiction that a strategy π guarantees 1 for initial state 2. As the largest payoff in the game is 1, π has to prescribe action T_2 with probability 1 whenever the play is in state 2 (otherwise the probability of absorption in state 4 with payoff 0 would be positive, if player 2 chooses action L_2). Thus, if player 2 always plays action R_2 in state 2, then the reward is 0, which is a contradiction. Therefore, player 1 has neither optimal nor best-Markov strategies for initial state 2.

4. Concluding Remarks

Alternative rewards. It is worthwhile to mention that the limiting average reward is sometimes defined as

$$E_{s\pi\sigma}\left(\limsup_{N\to\infty} R_N\right), \quad \liminf_{N\to\infty} E_{s\pi\sigma}\left(R_N\right), \quad \limsup_{N\to\infty} E_{s\pi\sigma}\left(R_N\right),$$

where R_N denotes the random variable for the average payoff up to stage $N \in \mathbb{N}$. All these reward functions are known to be equal for stationary strategy pairs (the limits exist). It is also known that the value is the same for these rewards [Mertens and Neyman (1981)]. As the reward function we used so far is always smaller or equal to any of the above rewards, with slight modifications in the proofs, all the results hold for these alternative rewards as well.

On alternative definitions of α and β. By the definition of α, for each $s \in S$ and for any $\delta > 0$, player 1 has a stationary strategy $x^{s\delta} \in X$ such that $\phi_s(x^{s\delta}) \geq \alpha_s - \delta$. In this finite state model, it can be shown however that for any $\delta > 0$, we can take $x^{s\delta}$ independent of the initial state. Thus, for all $\delta > 0$, there exists a $x^\delta \in X$ such that $\phi_s(x^\delta) \geq \alpha_s - \delta$ for all $s \in S$. This means that the following equality for stationary strategies makes sense:

$$\alpha = \sup_{x\in X} \phi(x)\,.$$

Therefore, we could have used this state independent equality as the definition of α as well. Note that for games with countable state space, this equivalence of definitions is not valid. Nowak and Raghavan (1991) presented a game with countable

state space, where even though player 1 has stationary ε-optimal strategies for each initial state, he has no stationary strategies that are ε-optimal for all initial states.

Finally, we wish to remark that it is not known to us whether β can be defined state independently or not.

References

Blackwell, D. and T. S. Ferguson (1968). "The Big Match". *Annals of Mathematical Statistics*, Vol. 33, 159–163.

Coulomb, J. M. (1992). "Repeated Games with Absorbing States and No Signals". *International Journal of Game Theory*, Vol. 21, 161–174.

Everett, H. (1957). "Recursive Games" in M. Dresher, A. W. Tucker and P. Wolfe (eds.), *Contributions to the Theory of Games*, Vol. III, *Annals of Mathematical Studies*, Vol. 39. Princeton: Princeton University Press, 47–78.

Filar, J. A. (1981). "Ordered Field Property for Stochastic Games when the Player who Controls Transitions Changes from State to State". *Journal of Optimization Theory and Applications*, Vol. 34, 503–515.

Flesch, J., F. Thuijsman and O. J. Vrieze (1997a). "Simplifying Optimal Strategies in Stochastic Games". *SIAM Journal on Control and Optimization*, Vol. 36, 1331–1347.

Flesch, J., F. Thuijsman and O. J. Vrieze (1997b). "Markov Strategies are Better Than Stationary Strategies". Report M97–09, Department of Mathematics, Maastricht University.

Gillette, D. (1957). "Stochastic Games with Zero Stop Probabilities" in M. Dresher, A. W. Tucker and P. Wolfe (eds.), *Contributions to the Theory of Games*, Vol. III, *Annals of Mathematical Studies*, Vol. 39. Princeton: Princeton University Press, 179–187.

Hordijk, A., O. J. Vrieze and G. L. Wanrooij (1983). "Semi-Markov Strategies in Stochastic Games". *International Journal of Game Theory*, Vol. 12, 81–89.

Kohlberg, E. (1974). "Repeated Games with Absorbing States". *Annals of Statistics*, Vol. 2, 724–738.

Kolmogorov, A. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung. Ergebnisse der Mathematik* 2, No. 3. Berlin: Springer Verlag.

Liggett, T. M. and S. A. Lippman (1969). "Stochastic Games with Perfect Information and Time Average Payoff". *SIAM Review*, Vol. 11, 604–607.

Mertens, J. F. and A. Neyman (1981). "Stochastic Games". *International Journal of Game Theory*, Vol. 10, 53–66.

Nowak, A. S. and T. E. S. Raghavan (1991). "Positive Stochastic Games and a Theorem of Ornstein" in T. E. S. Raghavan, T. S. Ferguson, O. J. Vrieze and T. Parthasarathy (eds.), *Stochastic Games and Related Topics.* Dordrecht: Kluwer Academic Publishers, 127–134.

Parthasarathy, T., S. H. Tijs and O. J. Vrieze (1984). "Stochastic Games with State Independent Transitions and Separable Rewards" in G. Hammer and D. Pallaschke (eds.), *Selected Topics in Operations Research and Mathematical Economics.* Berlin: Springer Verlag, 262–271.

Raghavan, T. E. S., S. H. Tijs and O. J. Vrieze (1985). "On Stochastic Games with Additive Reward and Transition Structure". *Journal of Optimization Theory and Applications*, Vol. 47, 451–464.

Rogers, P. D. (1969). *Non-Zerosum Stochastic Games.* Ph.D. thesis, Report ORC 69-8, Operations Research Center, University of California, Berkeley.

Sobel, M. J. (1971). "Noncooperative Stochastic Games". *Annals of Mathematical Statistics*, Vol. 42, 1930–1935.

Thuijsman, F. (1992). *Optimality and Equilibria in Stochastic Games.* CWI-Tract 82. Amsterdam: CWI.

Thuijsman, F. and O. J. Vrieze (1991). "Easy Initial States in Stochastic Games" in T. E. S. Raghavan, T. S. Ferguson, O. J. Vrieze and T. Parthasarathy (eds.), *Stochastic Games and Related Topics.* Dordrecht: Kluwer Academic Publishers, 85–100.

Thuijsman, F. and O. J. Vrieze (1993). "Stationary ε-Optimal Strategies in Stochastic Games". *OR Spektrum*, Vol. 15, 9–15.